Statistical Models and Methods for Lifetime Data

Second Edition

JERALD F. LAWLESS
University of Waterloo

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Statistical Models and Methods for Lifetime Data

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data

Lawless, J. F. (Jerald F.), 1944-
Statistical models and methods for lifetime data / Jerald F. Lawless.– 2nd ed.
p. cm. – (Wiley series in probability and statistics)
Includes bibliographical references and index.
ISBN 0-471-37215-3 (cloth : acid-free paper)
1. Failure time data analysis. 2. Survival analysis (Biometry). I. Title. II. Series.
QA276 .L328 2003
519.5—dc21 2002151805

Printed in the United States of America

10 9 8 7 6 5 4 3 2

To Liz

Contents

Preface to the Second Edition

The modeling and analysis of lifetimes is an important aspect of statistical work in a wide variety of scientific and technological fields. This book provides a comprehensive account of models and methods for lifetime data.

The field of lifetime data analysis has grown and matured since the first edition of the book was published. This second edition has accordingly been rewritten to reflect new developments in methodology, theory, and computing. The orientation and philosophy, however, remain close to the first edition. Lifetime data analysis is covered without concentrating exclusively on any specific field of application, though as before most of the examples are drawn from engineering and the biomedical sciences. There is a strong emphasis on parametric models, but non- and semiparametric methods are also given detailed treatments. Likelihood-based inference procedures are emphasized and serve to unify the methodology; implementation using both special lifetime data software, and general optimization software is discussed.

Extensive developments in software have made it possible to focus less on computational details and simplified methods of estimation than in the first edition, and to expand examples and illustrations. Graphical tools now feature more prominently. Many new references have been added, and some references and material from the first edition thus have been dropped. I have attempted, however, to retain enough early references to indicate the origins and evolution of different topics.

This edition of the book, like the first, is intended to serve as a reference for individuals interested in the analysis of lifetime data and as a text for graduate courses. Several appendices review aspects of statistical theory and computation that underlie the methodology presented in the book. A Problems and Supplements section concludes each chapter. There are many statistical software packages with capabilities in lifetime data analysis, and I have not attempted to provide an overview or comparison. Most of the examples were prepared using S-Plus, but other packages could have been used. Brief Computational Notes are provided at the ends of most chapters. Data sets discussed in the book are almost all either given or identified as being available electronically from Web locations mentioned in Appendix G.

There has been a small reorganization of topics from the first edition, consisting mainly of an expanded discussion of observation schemes and censoring (now in a new Chapter 2), and an expanded treatment of multiple failure modes (now given in

a separate Chapter 9). Several new topics have been introduced, including counting process-martingale tools, resampling and simulation methodology, estimating function methods, treatments of interval censored and truncated data, and discussions of multivariate lifetimes and event history models. In addition, material on many other topics has been updated and extensively revised, as have the Problems and Supplements sections.

To keep the book at a reasonable length I have had to omit or merely outline certain topics that might have been included. Some such topics are mentioned in the Bibliographic Notes that have been introduced at the ends of chapters, or in the Problems and Supplements sections. Statistical science encyclopedias (e.g. Kotz et al. 1988; Armitage and Colton 1998) are valuable sources of further information on a wide range of topics associated with lifetime data. Web-based tools for locating resource materials also make it relatively easy to research topics not covered in the book.

Chapter 1 contains introductory material on lifetime distributions and surveys important models. Chapter 2 deals with observation schemes for lifetime data and the formation of likelihood functions. Chapter 3 discusses graphical methods and nonparametric estimation of distribution characteristics based on different types of lifetime data. Chapter 4 introduces inference procedures for parametric models, including exponential, gamma, inverse Gaussian, and mixture models. Chapter 5 provides corresponding procedures for log-location-scale models and extensions to them; the Weibull, log-normal and log-logistic models are treated in detail. Chapter 6 discusses regression models, exploratory and diagnostic methods, and develops inference procedures for parametric models. Chapter 7 deals with semiparametric methodology for proportional or multiplicative hazards models. Chapter 8 presents rank-based and semiparametric procedures based on location-scale models. Chapter 9 gives a thorough treatment of multiple failure modes, or competing risks. Chapter 10 discusses goodness-of-fit tests and describes procedures for specific models in the book. Finally, Chapter 11 introduces several important topics that go beyond univariate survival analysis: multivariate lifetime models, sequences of lifetimes, event history processes, and joint models for lifetimes and coprocesses. It is shown how the methods of previous chapters can be applied to many problems in these areas.

I am indebted to various individuals for their contributions to this edition of the book. Ker-Ai Lee and Melanie Wigg assisted with computing and the preparation of examples. Some examples are based on joint work with Richard Cook, Jack Kalbfleisch, and graduate student Wenqing He. I have benefitted for many years from collaboration and conversations with Richard Cook, Jack Kalbfleisch, and Jock MacKay, and from my interactions with numerous fine graduate students at Waterloo.

I want to acknowledge and thank Lynda Clarke, who has labored long, hard, and expertly on the manuscript, as she did on the first edition of the book 20 years ago.

The University of Waterloo's Department of Statistics and Actuarial Science has provided a stimulating environment for research and teaching throughout my career. Part of the work for this edition was done during a sabbatical leave spent at University of Auckland (January–March 2000) and at University College London (January–March 2001); their hospitality is gratefully acknowledged. I also want to acknowledge support over many years from the research grants programs of the Natural Sciences and Engineering Research Council of Canada (NSERC), and to thank General Motors Canada for their cosponsorship, with NSERC, of a personal Industrial Research Chair.

Finally, I thank my family and especially my wife, Liz, for her patience and support during this project.

JERALD. F. LAWLESS

Waterloo, Ontario
April 2002

Preface to the First Edition

The statistical analysis of lifetime or response time data has become a topic of considerable interest to statisticians and workers in areas such as engineering, medicine, and the biological sciences. The field has expanded rapidly in recent years, and publications on the subject can be found in the literatures of several disciplines besides statistics. This book draws together material on the analysis of lifetime data and gives a comprehensive presentation of important models and methods.

My aim is to give a broad coverage of the area without unduly concentrating on any single field of application. Most of the examples in the book, however, come from engineering or the biomedical sciences, where these methods are widely used. The book contains what I feel are the most important topics in lifetime data methodology. These include various parametric models and their associated statistical methods, nonparametric and distribution-free methods, and graphical procedures. To keep the book at a reasonable length I have had to either sketch or entirely omit topics that could have usefully been treated in detail. Some of these topics are referenced or touched upon in the Problems and Supplements sections at the ends of chapters.

This book is intended as a reference for individuals interested in the analysis of lifetime data and can also be used as a text for graduate courses in this area. A basic knowledge of probability and statistical inference is assumed, but I have attempted to carefully lay out the models and assumptions upon which procedures are based and to show how the procedures are developed. In addition, several appendices review statistical theory that may be unfamiliar to some readers. Numerical illustrations are given for most procedures, and the book contains numerous examples involving real data. Each chapter concludes with a Problems and Supplements section, which provides exercises on the chapter material, and supplements and extends the topics discussed. For the reader interested in research on lifetime data methodology I have given fairly extensive references to recent work and outstanding problems.

Chapter 1 contains introductory material on lifetime distributions and surveys the most important parametric models. Censoring is introduced, and its ramifications for statistical inference are considered. In Chapter 2 some methods of examining univariate lifetime data and obtaining nonparametric estimates of distribution characteristics are discussed; life tables and graphical procedures play key roles. Chapters

3, 4, and 5 deal with inference for important parametric models, including the exponential, Weibull, gamma, log-normal, and generalized gamma distributions. This is extended in Chapter 6 to problems with concomitant variables, through regression models based on these distributions. Chapters 7 and 8 present nonparametric and distribution-free procedures: Chapter 7 deals with methods based on the proportional hazards regression model, and Chapter 8 gives distribution-free procedures for single- and many-sample problems. Goodness-of-fit tests for lifetime distribution models are considered in Chapter 9. Chapter 10 contains brief discussions of two important topics for which it was not feasible to give extended treatments: multivariate and stochastic process models. Several sections in this book are marked with asterisks; these contain discussions of a technical nature and can be omitted on a first reading.

A final remark concerning the methods presented is that the computer is, as always in modern statistics, a useful if not indispensible tool. For some problems, methods that do not require a computer are available, but more often access to a computer is a necessity. I have commented, wherever possible, on the computational aspects of procedures and have included additional material on computation in the Appendices.

Part of the work for the book was done during a sabbatical leave spent at Imperial College, London, and the University of Reading from 1978 to 1979; their hospitality is gratefully acknowledged. I would also like to express my appreciation to two extremely fine typists, Annemarie Nittel and Lynda Hohner, who labored long and diligently in the preparation of the manuscript.

J. F. LAWLESS

Waterloo, Ontario
June 1981

Statistical Models and Methods for Lifetime Data

CHAPTER 1

Basic Concepts and Models

1.1 INTRODUCTION

The statistical analysis of what are variously referred to as lifetime, survival time, or failure time data is an important topic in many areas, including the biomedical, engineering, and social sciences. Applications of lifetime distribution methodology range from investigations of the durability of manufactured items to studies of human diseases and their treatment. Some methods of dealing with lifetime data are quite old, but starting about 1970 the field expanded rapidly with respect to methodology, theory, and fields of application. Software packages for lifetime data analysis have been widely available since about 1980, with the frequent appearance of new features and packages.

This book presents and illustrates statistical methods for modeling and analyzing lifetime data. The aim is to provide a general treatment, and not focus exclusively on a particular field of application. Lifetime distribution methodology is widely used in the biomedical and engineering sciences, however, and most of the examples in the book come from those areas.

Throughout the book various types of data will, for convenience, be referred to as "lifetime" data. Basically, however, we consider situations in which the time to the occurrence of some event is of interest for individuals in some population. Sometimes the events are actual deaths of individuals and "lifetime" is the length of life measured from some particular starting point. In other instances "lifetime" and the words "death" or "failure," which denote the event of interest, are used in a figurative sense. In discussing applications, other terms such as "survival time" and "failure time" are also frequently used.

The following examples illustrate some ways in which lifetime data arise.

Example 1.1.1. Manufactured items with mechanical or electronic components are often subjected to life tests in order to obtain information on their durability. This involves putting items in operation, often in a laboratory setting, and observing them until they fail. It is common here to refer to the lifetimes as "failure times," since when an item ceases operating satisfactorily, it is said to have "failed."

Example 1.1.2. Demographers and social scientists are interested in the duration of certain life "states" for humans. Consider, for example, marriage and, in particular, the marriages formed during the year 1980 in a particular country. Then the lifetime of a marriage would be its duration; a marriage may end due to annulment, divorce, or death.

Example 1.1.3. In medical studies dealing with potentially fatal diseases one is interested in the survival time of individuals with the disease, measured from the date of diagnosis or some other starting point. For example, it is common to compare treatments for a disease at least partly in terms of the survival time distributions for patients receiving the different treatments.

Example 1.1.4. A standard experiment in the investigation of carcinogenic substances is one in which laboratory animals are subjected to doses of the substance and then observed to see if they develop tumors. A main variable of interest is the time to appearance of a tumor, measured from when the dose is administered.

The definition of lifetime includes a time scale and time origin, as well as a specification of the event (e.g., failure or death) that determines lifetime. In some settings it is difficult to say precisely when the event occurs: for example, this is the case for the appearance of a tumor in Example 1.1.4. The time scale is not always real or chronological time, especially where machines or equipment are concerned. For example, miles driven might be used as a time scale with motor vehicles, and number of pages of output for a computer printer or photocopier.

The main problems addressed in this book are those of specifying models to represent lifetime distributions and of making inferences based on these models. The objectives of modeling and statistical analysis include description or estimation of distributions, comparison of distributions, furthering scientific understanding, process or system improvement, prediction, and decision. Covariates or explanatory variables that can be related to lifetime usually feature prominently in these activities. In some settings there may be more than one lifetime variable associated with an individual, or an individual may die in different ways. The types of models used in lifetime data analysis range from fully parametric to nonparametric; semiparametric models that have both parametric and nonparametric features are common. The remaining sections of this chapter introduce lifetime models, but first we discuss some additional features and examples of lifetime data.

The chronological time needed to observe the lifetimes of all individuals in a study may be large enough that practical constraints prevent full observation. This leads to what is termed "censoring," in which an individual's lifetime is known only to exceed a certain value. In Example 1.1.1, for example, a life test might be terminated after, say, 28 days; if an item had not failed by that time, we would know only that its lifetime exceeded 28 days and refer to that value as a "censoring time." More generally, it may not be possible to determine exactly when a failure or death occurs, because individuals are seen only at certain times. In that case, we may know only that a lifetime lies in some interval (L, R); we refer to this as "interval censoring."

Another complication is that covariates associated with lifetime data may vary over time, and it may not be possible to observe their values at all times.

These and other features associated with lifetime data create interesting problems for analysis, and much of the development of the subject has been devoted to dealing with them. Chapter 2 considers these issues in detail. The remainder of this chapter covers the basic concepts of lifetime distributions and introduces important models. Section 1.2 discusses lifetime distributions generally, and Section 1.3 introduces important parametric univariate models. Sections 1.4 to 1.6 discuss more complex models involving covariates, multiple lifetimes, and multiple types of failure. Before turning to this, we consider a few examples of lifetime data, to illustrate some of the points just mentioned.

Example 1.1.5. Nelson (1972a) described the results of a life test experiment in which specimens of a type of electrical insulating fluid were subjected to a constant voltage stress. The length of time until each specimen failed, or "broke down," was observed. Table 1.1 gives results for seven groups of specimens, tested at voltages ranging from 26 to 38 kilovolts (kV).

The main purpose of the experiment was to investigate the distribution of time to breakdown for the insulating fluid and to relate this to the voltage level. Quite clearly, breakdown times tend to decrease as the voltage increases. In addition to the formulation of a model relating breakdown times and voltage, the estimation of the breakdown time distribution at a "normal" voltage of 20 kV was important. Breakdown times tend to be very large at 20 kV, and this involves a substantial extrapolation from the experimental data.

The experiment in Example 1.1.5 was run long enough to observe the failure of all the insulation specimens tested. Sometimes it may take a long time for all items to fail, and it is deemed necessary to terminate a study before this can happen. In this case, the lifetimes of certain items are censored. For example, if a decision had been

Table 1.1. Times to Breakdown (in minutes) at Each of Seven Voltage Levels

Voltage Level (kV)	n_i	Breakdown Times
26	3	5.79, 1579.52, 2323.7
28	5	68.85, 426.07, 110.29, 108.29, 1067.6
30	11	17.05, 22.66, 21.02, 175.88, 139.07, 144.12, 20.46, 43.40, 194.90, 47.30, 7.74
32	15	0.40, 82.85, 9.88, 89.29, 215.10, 2.75, 0.79, 15.93, 3.91, 0.27, 0.69, 100.58, 27.80, 13.95, 53.24
34	19	0.96, 4.15, 0.19, 0.78, 8.01, 31.75, 7.35, 6.50, 8.27, 33.91, 32.52, 3.16, 4.85, 2.78, 4.67, 1.31, 12.06, 36.71, 72.89
36	15	1.97, 0.59, 2.58, 1.69, 2.71, 25.50, 0.35, 0.99, 3.99, 3.67, 2.07, 0.96, 5.35, 2.90, 13.77
38	8	0.47, 0.73, 1.40, 0.74, 0.39, 1.13, 0.09, 2.38

Table 1.2. Lifetimes for 10 Pieces of Equipment

Item Number	1	2	3	4	5	6	7	8	9	10
Date of installation	11 June	21 June	22 June	2 July	21 July	31 July	31 July	1 Aug	2 Aug	10 Aug
Date of failure	13 June	—	12 Aug	—	23 Aug	27 Aug	14 Aug	25 Aug	6 Aug	—
Lifetime (days)	2	≥ 72	51	≥ 60	33	27	14	24	4	≥ 21

made in the preceding experiment to terminate testing after 180 minutes had elapsed, then two of the observations in the 26- and 28-kV sample and one each in the 30- and 32-kV samples would have been censored. In each case, we would not know the exact failure time of the item, but only that it exceeded 180 minutes.

Censoring arises in lifetime data in a variety of ways and is discussed in detail in Chapter 2. The remaining examples in this section all involve censoring of some kind.

Example 1.1.6. Bartholomew (1957) considered a situation in which pieces of equipment were installed in a system at different times. At a later date some of the pieces had failed and the rest were still in use. With the aim of studying the lifetime distribution of the equipment, Bartholomew gave the data in Table 1.2 for 10 pieces of equipment. The first item was installed on June 11 and data were collected up to August 31. At that time, three items (numbers 2, 4, and 10) had still not failed, and their failure times are therefore censored; we know for these items only that their failure times exceed 72, 60, and 21 days, respectively.

Example 1.1.7. Gehan (1965) and others have discussed the results of a clinical trial reported by Freireich et al. (1963), in which the drug 6-mercaptopurine (6-MP) was compared to a placebo with respect to the ability to maintain remission in acute leukemia patients. Table 1.3 gives remission times for two groups of 21 patients each, one group given the placebo and the other the drug 6-MP.

The starred observations are censoring times; for these patients, the disease was still in a state of remission at the end of the study. Censoring is common in clinical trials, since the trial is often terminated before all individuals have "failed." In addition, individuals may enter a study at various times, and hence may be under observation for different lengths of time. In this trial, individuals entered the study in matched pairs at different times and a sequential stopping rule was used to terminate the study (Klein and Moeschberger 1997, p. 2).

Example 1.1.8. Therneau and Hamilton (1997) discussed data that arose in a study of persons with cystic fibrosis (Fuchs et al. 1994). These individuals are susceptible to an accumulation of mucus in the lungs, which leads to pulmonary exacerbations and deterioration of lung function. In a clinical trial to investigate the efficacy of daily administration of a recombinant form of the human enzyme DNase 1 in preventing exacerbations, subjects were randomly assigned to the new treatment (called rhDNase or Pulmozyme) or a placebo. Subjects, who were exacerbation-free at ran-

Table 1.3. Lengths of Remission (in weeks) for Two Groups of Patients[a]

6-MP	6, 6, 6, 6*, 7, 9*, 10, 10*, 11*, 13, 16, 17*, 19*, 20*, 22, 23, 25*, 32*, 32*, 34*, 35*
Placebo	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

[a]Stars denote censored observations.

Table 1.4. Times to First Pulmonary Exacerbation for 10 Subjects

t (days)[a]	trt	fev[b]
168*	1	28.8
169*	1	64.0
65	0	67.2
168*	1	57.6
171*	0	57.6
166*	1	25.6
168*	0	86.4
90	0	32.0
169*	1	86.4
8	0	28.8

[a]Starred values are censoring times.
[b]fev measure is percent of predicted normal fev, based on sex, age, and height.

domization, were followed for approximately 169 days and the days at which an exacerbation period started were noted. When an exacerbation spell began, a subject was given antibiotics, and after the exacerbation had disappeared the subject was then considered at risk for a new exacerbation. Consequently, some subjects had no exacerbations over the 169-day follow-up period, some had one, and some had two or more.

There were 324 subjects assigned to the Placebo group by randomization, and 321 to the rhDNase group. The objective was to compare the two groups in terms of the avoidance of exacerbations. The simplest comparison is to note that 139 (43%) of Placebo subjects had at least one exacerbation and that 104 (32%) of rhDNase subjects did. A more comprehensive comparison can be based on the time t to the first exacerbation after randomization. Table 1.4 shows data for ten subjects: failure time t and two covariates, trt (= 0 for Placebo and 1 for rhDNase) and fev (forced expiratory volume at the time of randomization, which is a measure of initial pulmonary function). A still more comprehensive analysis might also use the data on second and subsequent exacerbations; this topic is discussed in Chapter 11.

Example 1.1.9. Table 1.5 presents survival data on 40 advanced lung cancer patients, taken from a study discussed by Prentice (1973). The main purpose of the study was to compare the effects of two chemotherapy treatments in prolonging survival time. All patients represented here received prior therapy and were then randomly assigned to one of the two treatments, termed "standard" and "test." Survival times t, measured from the start of treatment for each patient, are recorded in Table 1.5. Censored observations correspond to patients who were still alive at the time the data were collected. Concomitant variables that were thought possibly to be important are also shown for each patient. First, patients can have different types of tumors; they have been classified into four categories (squamous, small, adeno,

Table 1.5. Lung Cancer Survival Data[a,b]

t	x_1	x_2	x_3	t	x_1	x_2	x_3
	Standard, Squamous				*Test, Squamous*		
411	70	64	5	999	90	54	12
126	60	63	9	231*	50	52	8
118	70	65	11	991	70	50	7
82	40	69	10	1	20	65	21
8	40	63	58	201	80	52	28
25*	70	48	9	44	60	70	13
11	70	48	11	15	50	40	13
	Standard, Small				*Test, Small*		
54	80	63	4	103*	70	36	22
153	60	63	14	2	40	44	36
16	30	53	4	20	30	54	9
56	80	43	12	51	30	59	87
21	40	55	2		*Test, Adeno*		
287	60	66	25				
10	40	67	23	18	40	69	5
	Standard, Adeno			90	60	50	22
				84	80	62	4
8	20	61	19		*Test, Large*		
12	50	63	4				
	Standard, Large			164	70	68	15
				19	30	39	4
177	50	66	16	43	60	49	11
12	40	68	12	340	80	64	10
200	80	41	12	231	70	67	18
250	70	53	8				
100	60	37	13				

[a]Starred quantities denote censored observations.
[b]Days of survival t, performance status x_1, age in years x_2, and number of months from diagnosis to entry into the study x_3.

and large). Also given for each patient is a Karnofsky score, or performance status, assigned at the time of diagnosis. This is a measure of general medical status on a scale of 10 to 90: 10, 20, and 30 mean that the patient is completely hospitalized; 40, 50, and 60 that he is partially confined to hospital; and 70, 80, and 90 that he is able to care for himself. Finally, the age of the patient and the number of months from diagnosis of lung cancer to entry into the study are recorded.

Example 1.1.10. The data in Table 1.6 are from an experiment in which new models of a small electrical appliance were being tested (Nelson 1970b). The appliances were operated repeatedly by an automatic testing machine; the lifetimes given

Table 1.6. Failure Data for Electrical Appliance Test

Number of Cycles to Failure	Failure Code	Number of Cycles to Failure	Failure Code	Number of Cycles to Failure	Failure Code
11	1	958	10	35	15
2,223	9	7,846	9	2,400	9
4,329	9	170	6	1,167	9
3,112	9	3,059	6	2,831	2
13,403	0	3,504	9	2,702	10
6,367	0	2,568	9	708	6
2,451	5	2,471	9	1,925	9
381	6	3,214	9	1,990	9
1,062	5	3,034	9	2,551	9
1,594	2	3,034	9	2,761	6
329	6	49	15	2,565	0
2,327	6	6,976	9	3,478	9

here are the number of cycles of use completed until the appliances failed. There are two complicating factors: one is that there were many different ways in which an appliance could fail; 18 to be exact. Therefore in Table 1.6 each observation has a failure code beside it. Numbers 1 through 18 refer to the 18 different possible causes of failure for the appliances. In addition, some of the observations were censored, since it was not always possible to continue testing long enough for an appliance to fail. Appliances that have censored failure times are indicated in Table 1.6 as having a failure code of 0.

The joint distribution of failure times and failure modes is of interest. This can be used to help plan further development and testing of the appliance. The failure time distribution will change as the appliance is developed, and product improvements effectively remove certain causes of failure. In the final stages, the failure time distribution model can be used to predict the implications of a warranty plan for the appliance.

The preceding examples show some of the ways in which lifetime data arise and some of the questions that such data hope to answer. We now leave the discussion of data for the time being and turn to an examination of statistical models for lifetime distributions.

1.2 LIFETIME DISTRIBUTIONS

1.2.1 Continuous Models

We begin by considering the case of a single continuous lifetime variable, T. Specifically, let T be a nonnegative random variable representing the lifetimes of individuals in some population.

All functions, unless stated otherwise, are defined over the interval $[0, \infty)$. Let $f(t)$ denote the probability density function (p.d.f.) of T and let the (cumulative) distribution function (c.d.f.) be

$$F(t) = Pr(T \le t) = \int_0^t f(x)\,dx.$$

The probability of an individual surviving to time t is given by the survivor function

$$S(t) = Pr(T \ge t) = \int_t^\infty f(x)\,dx. \tag{1.2.1}$$

In some contexts involving systems or lifetimes of manufactured items, $S(t)$ is referred to as the reliability function. Note that $S(t)$ is a monotone decreasing continuous function with $S(0) = 1$ and $S(\infty) = \lim_{t\to\infty} S(t) = 0$. Occasionally we may wish to allow $S(\infty) > 0$ to consider settings where some individuals never fail; these will be treated as special cases.

The pth quantile of the distribution of T is the value t_p such that

$$Pr(T \le t_p) = p.$$

That is, $t_p = F^{-1}(p)$. The pth quantile is also referred to as the $100p$th percentile of the distribution. The .5 quantile is called the median of the distribution.

A very important concept with lifetime distributions is the hazard function $h(t)$, defined as

$$h(t) = \lim_{\Delta t\to 0} \frac{Pr(t \le T < t + \Delta t | T \ge t)}{\Delta t} = \frac{f(t)}{S(t)}. \tag{1.2.2}$$

The hazard function specifies the instantaneous rate of death or failure at time t, given that the individual survives up to t; $h(t)\,\Delta t$ is the approximate probability of death in $[t, t + \Delta t)$, given survival up to t. The hazard function is sometimes given other names, among them the hazard rate and the force of mortality.

The functions $f(t)$, $F(t)$, $S(t)$, and $h(t)$ give mathematically equivalent specifications of the distribution of T. It is easy to derive expressions for $S(t)$ and $f(t)$ in terms of $h(t)$: since $f(t) = -S'(t)$, (1.2.2) implies that

$$h(x) = -\frac{d}{dx}\log S(x).$$

Thus

$$\log S(x)\Big|_0^t = -\int_0^t h(x)\,dx,$$

and since $S(0) = 1$, we find that

$$S(t) = \exp\left(-\int_0^t h(x)\,dx\right). \tag{1.2.3}$$

It is also useful to define the cumulative hazard function

$$H(t) = \int_0^t h(x)\,dx,$$

which, by (1.2.3), is related to the survivor function by $S(t) = \exp[-H(t)]$. If $S(\infty) = 0$, then $H(\infty) = \infty$. Finally, in addition to (1.2.3), it follows immediately from (1.2.2) that

$$f(t) = h(t)\exp\left(-\int_0^t h(x)\,dx\right). \tag{1.2.4}$$

Example 1.2.1. Suppose T has p.d.f.

$$f(t) = \beta t^{\beta-1}\exp\left(-t^{\beta}\right) \qquad t > 0,$$

where $\beta > 0$ is a parameter; this is a Weibull distribution, discussed in Section 1.3.2. It follows easily from (1.2.1) that the survivor function for T is $S(t) = \exp(-t^{\beta})$, and then from (1.2.2) the hazard function is $h(t) = \beta t^{\beta-1}$. Conversely, if the model is specified initially in terms of $h(t)$, then $S(t)$ and $f(t)$ are readily obtained from (1.2.3) and (1.2.4).

1.2.2 Discrete Models

Sometimes, for example, when lifetimes are grouped or measured as a number of cycles of some sort, T may be treated as a discrete random variable. Suppose T can take on values $t_1, t_2, \ldots$, with $0 \le t_1 < t_2 < \cdots$, and let the probability function (p.f.) be

$$f(t_j) = Pr(T = t_j) \qquad j = 1, 2, \ldots.$$

The survivor function is then

$$S(t) = Pr(T \ge t) = \sum_{j:t_j \ge t} f(t_j). \tag{1.2.5}$$

When considered as a function for all $t \ge 0$, $S(t)$ is a left-continuous, nonincreasing step function, with $S(0) = 1$ and $S(\infty) = 0$.

The discrete time hazard function is defined as

$$\begin{aligned} h(t_j) &= Pr(T = t_j | T \ge t_j) \\ &= \frac{f(t_j)}{S(t_j)} \qquad j = 1, 2, \ldots. \end{aligned} \tag{1.2.6}$$

As in the continuous case, the probability, survivor, and hazard functions give equivalent specifications of the distribution of T. Since $f(t_j) = S(t_j) - S(t_{j+1})$, (1.2.6) implies that

$$h(t_j) = 1 - \frac{S(t_{j+1})}{S(t_j)} \qquad j = 1, 2, \ldots, \tag{1.2.7}$$

and thus

$$S(t) = \prod_{j:t_j < t} [1 - h(t_j)]. \tag{1.2.8}$$

An analog of the continuous $H(t)$ could be defined two ways in the discrete case. One would be by analogy with (1.2.3), as $-\log S(t)$, where $S(t)$ is given by (1.2.8). It is easily seen that this does not equal $\sum_{j:t_j<t} h(t_j)$, which is the second analog. The next section introduces a way to unify continuous, discrete, and mixed lifetime distributions in one framework, and it is the second definition of $H(t)$ that is adopted.

1.2.3 A General Formulation

Continuous, discrete, and mixed distributions can conveniently be treated within a single framework. To do this we introduce two types of integrals, called Riemann–Stieltjes integrals and product integrals.

Let $G(u)$ be a nondecreasing, right-continuous function with left-hand limits and a finite number of discontinuities in any finite interval. Assume that $g(u) = G'(u)$ exists except at points of discontinuity of G and that at points of discontinuity a_j we have $G(a_j) - G(a_j-) = g_j$, where $G(a-) = \lim_{\Delta a \to 0} G(a - \Delta a)$. The Riemann–Stieltjes integral of dG over the interval $(a, b]$ is then defined as

$$\int_{(a,b]} dG(u) = \int_a^b g(u)\,du + \sum_{j:a<a_j\le b} g_j, \tag{1.2.9}$$

where the first integral on the right side of (1.2.9) is a Riemann integral. We can think of $dG(u)$ as equal to $g(u)\,du + G(u) - G(u-)$.

In general, a distribution function $F(t) = Pr(T \le t)$ is a right-continuous, nondecreasing function, with jumps at points a_j for which $Pr(T = a_j) = f_j > 0$, and p.d.f. $f(u) = F'(u)$ at points where $F(u)$ is continuous. Then (1.2.9) gives

$\Pr(a < T \leq b)$ as

$$F(b) - F(a) = \int_{(a,b]} dF(u) = \int_a^b f(u)\,du + \sum_{j:a<a_j\leq b} f_j. \tag{1.2.10}$$

If $F(t)$ is continuous, there are no jump points, and if T has a discrete distribution, then $F(u)$ is a step function with $f(u) = 0$ at all continuity points. Figure 1.1 portrays discrete and continuous cases for $F(t)$.

To give a general treatment of the hazard function we introduce the product integral. Let $a = u_0 < u_1 < \cdots < u_m = b$ partition $(a, b]$, with $\Delta u_i = u_i - u_{i-1}$ and $\max(\Delta u_i) \to 0$ when $m \to \infty$. The product integral of a function $dG(u)$ as defined earlier is

$$\prod_{(a,b]} \{1 + dG(u)\} = \lim_{m\to\infty} \prod_{i=1}^{m} \{1 + G(u_i) - G(u_{i-1})\}. \tag{1.2.11}$$

If $G(u)$ is continuous for all u in $(a, b]$, then $dG(u) = g(u)\,du$ and (1.2.11) gives

$$\prod_{(a,b]} \{(1 + g(u)\,du\} = \lim_{m\to\infty} \prod_{i=1}^{m} \{1 + g(u_i)\,\Delta u_i + o(\Delta u_i)\}$$
$$= \lim_{m\to\infty} \prod_{i=1}^{m} \{1 + g(u_i)\,\Delta u_i\},$$

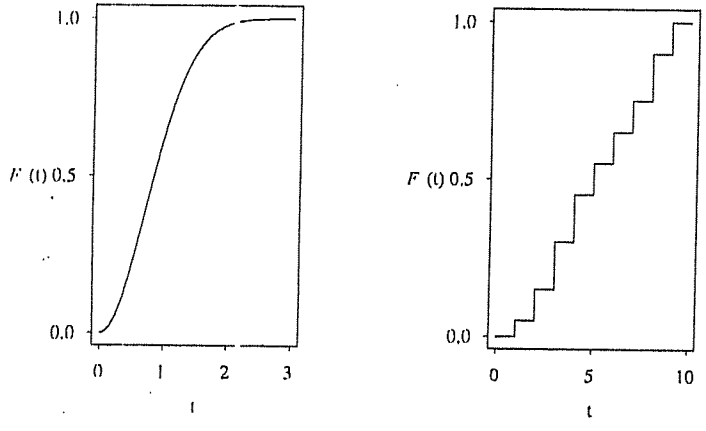

Figure 1.1. Cumulative distribution functions.

where $o(x)$ means a function $w(x)$ such that $w(x)/x \to 0$ as $x \to 0$. By noting that $\log\{1 + g(u_i)\,\Delta u_i\} = g(u_i)\,\Delta u_i + o(\Delta u_i)$ for Δu_i small and taking the log of the product integral, we see that in the continuous case

$$\prod_{(a,b]} \{1 + g(u)\,du\} = \exp\left\{\int_a^b g(u)\,du\right\}, \tag{1.2.12}$$

which relates the product and Riemann integrals.

If $G(u)$ has jumps at points a_j $(j = 1, 2, \ldots)$ of sizes g_j, then (1.2.11) gives

$$\prod_{(a,b]} \{1 + dG(u)\} = \prod_{(a,b]} \{1 + g(u)\,du\} \prod_{j:a<a_j\leq b} (1 + g_j). \tag{1.2.13}$$

Note that if $G(u)$ is a step function, then $g(u) = 0$ at all continuity points and the first term on the right side disappears.

We are now in a position to consider the hazard function. Let $h(u) = f(u)/S(u)$ represent the hazard function for T at points where $F(u)$ (or $S(u)$) is continuous, and $h_j = Pr(T = a_j | T \geq a_j)$ be the discrete hazard values at times a_j for which a jump in F occurs. The cumulative hazard function is then defined by a Riemann–Stieltjes integral of the form (1.2.9):

$$H(t) = \int_0^t dH(u) = \int_0^t h(u)\,du + \sum_{j:a_j\leq t} h_j. \tag{1.2.14}$$

Given the cumulative hazard function, we can obtain the survivor function through the fundamental result that for any sequence of values $0 = u_0 < u_1 < \cdots < u_m = t$,

$$Pr(T \geq t) = \prod_{i=1}^{m} Pr(T \geq u_i | T \geq u_{i-1}). \tag{1.2.15}$$

Now for $\Delta u_i = u_i - u_{i-1}$ sufficiently small, $[u_{i-1}, u_i)$ contains either 0 or 1 jump points, and

$$Pr(T \geq u_i | T \geq u_{i-1}) = 1 - \frac{Pr(u_{i-1} \leq T < u_i)}{Pr(T \geq u_{i-1})}$$
$$= 1 - [H(u_i-) - H(u_{i-1}-)] + o(\Delta u_i).$$

Therefore by (1.2.15) and (1.2.11),

$$Pr(T \geq t) = \prod_{(0,t)} [1 - dH(u)]. \tag{1.2.16}$$

Note that in (1.2.16) the product integral is over the open interval $(0, t)$ since $S(t) = Pr(T \geq t)$ and is left-continuous, whereas $H(u)$ is right-continuous. Sometimes $S(t)$ is defined as $Pr(T > t)$ and in that case (1.2.16) is replaced with the product integral over $(0, t]$.

The relationships (1.2.14) and (1.2.16) apply to all types of distribution. We get (1.2.3) from (1.2.16) in the case of a continuous distribution by using (1.2.12), and we get (1.2.8) for a discrete distribution by using (1.2.13). In general (1.2.16) gives

$$Pr(T \geq t) = \exp\left\{-\int_0^t h(u)\,du\right\} \prod_{j:a_j<t} (1 - h_j). \tag{1.2.17}$$

Finally, we note a useful result following immediately from (1.2.16): for $a < b$,

$$Pr(T \geq b | T \geq a) = \prod_{[a,b)} [1 - dH(u)]. \tag{1.2.18}$$

1.2.4 Some Remarks on the Hazard Function

The hazard function is a particularly important characteristic of a lifetime distribution. It indicates the way the risk of failure varies with age or time, and this is of interest in most applications. Prior information about the shape of the hazard function can help guide model selection. Finally, if factors affecting an individual's lifetime vary over time, it is often essential to approach modeling through the hazard function.

Figure 1.2 shows hazard functions and p.d.f.'s for four continuous distributions. The shapes of the hazard functions are qualitatively different; distribution (*a*) has a monotone increasing hazard function, distribution (*b*) has a monotone decreasing hazard function, (*c*) has a so-called bathtub-shaped, or U-shaped, hazard function, and (*d*) displays an inverse bathtub shape. Models with these and other shapes are all useful in practice. If, for example, individuals in a population are followed right from actual birth to death, a bathtub-shaped hazard function is often seen. We are familiar with this pattern in human populations: after an initial period in which deaths result primarily from birth defects or infant diseases, the death rate drops and is relatively constant until the age of 30 or so, after which it increases with age. This pattern also manifests itself in other biological populations and in populations of manufactured items, some of which contain defects.

Distributions with increasing hazard functions are seen for individuals for whom some kind of aging or wearout takes place. Also, populations that display a bathtub-shaped hazard function are sometimes purged of weak individuals, leaving a reduced population with an increasing hazard function. For example, manufacturers may use inspection or a burn-in process, in which items are subjected to a brief period of operation before being sent to customers. In this way defective or poor-quality items that would fail early are removed from the population; this frequently leaves a residual population that exhibits an increasing hazard function.

Certain types of electronic devices display a decreasing hazard as items with defects fail and are removed from the population. Roughly constant hazard functions

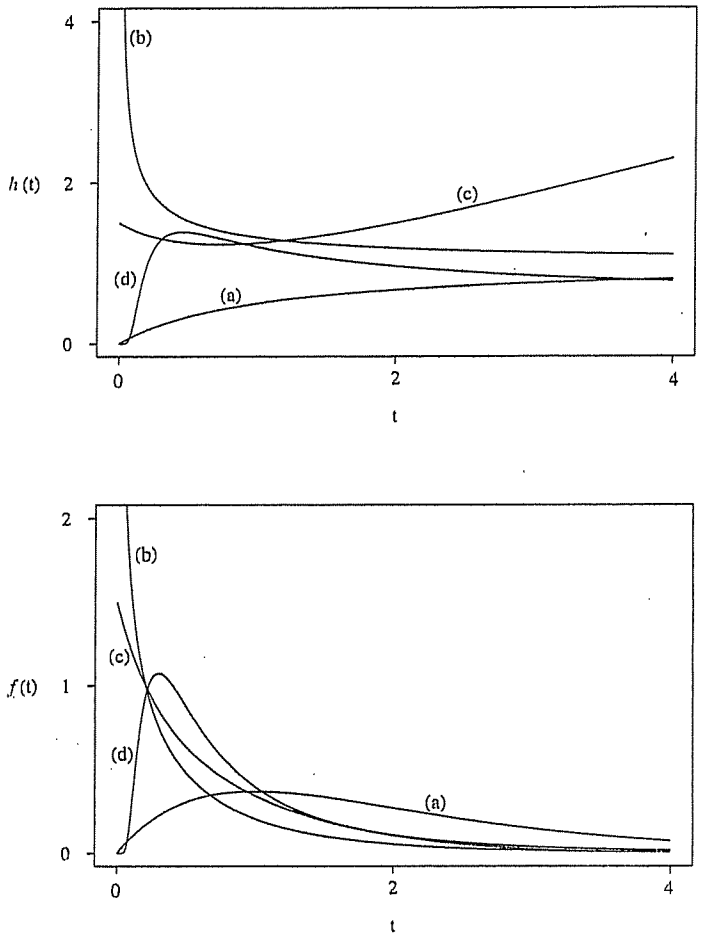

Figure 1.2. Some hazard and probability density functions.

tend to occur in stable settings where failure or death is due to random phenomena such as shocks or accidents, which are external to the individual. Shape (d) in Figure 1.2, where $h(t)$ first increases to a maximum and then decreases, is encountered in many applications, for example, in the case of survival after treatment for cancer, where some individuals are cured, and in connection with the duration of marriage.

In many settings factors or covariates affecting an individual's lifetime vary over time; we refer to them as "time-varying" or "time-dependent" covariates. For example, in life tests of electrical insulation (see Example 1.1.5) the voltage level that items are subjected to is sometimes changed over time according to a fixed

schedule. In studies of the age at which smokers develop chronic diseases, the type and level of smoking for each individual can vary over time. The duration of a marriage (see Example 1.1.2) may be affected by the presence of children or the couple's employment status, both of which can change over time. When there are time-varying covariates, it is usually essential to think about models in terms of their hazard functions. We cannot discuss a lifetime's relationship to covariates without considering the covariate "history," that is, the values the covariates take over time; a generally useful approach is to consider the hazard function at time t conditional on previous covariate values. Subtle issues arise in connection with the modeling and interpretation of time-varying covariate effects. Discussion is provided in Chapters 6, 7, and 8, and introductory remarks are given in Section 1.4, where regression models are discussed.

1.3 SOME IMPORTANT MODELS

Various parametric families of models are used in the analysis of lifetime data and the modeling of aging or failure processes. Among univariate models, a few distributions occupy a central position because of their demonstrated usefulness in a wide range of situations. Foremost in this category are the exponential, Weibull, log-normal, log-logistic, and gamma distributions. This section introduces these and some other models.

Sometimes there is information about the aging or failure process in a population that suggests a particular distribution, though this information is rarely specific enough to narrow consideration to just one family of models. Some references providing theoretical motivation for certain models are provided in the Bibliographic Notes section at the end of the chapter. The motivation for using a particular model in a given situation is often empirical, it having been found that the model satisfactorily describes the distribution of lifetimes in populations like the one under study. Convenience of analysis can also be a factor. Section 1.6 provides some general remarks on model selection and analysis.

We make one additional preliminary remark. Models are presented here without the inclusion of a so-called threshold parameter, or guarantee time. Briefly, this is a time $\gamma \geq 0$ before which it is assumed that an individual cannot die. Occasionally a situation calls for the inclusion of such a parameter. The distributions considered can all be extended to include a threshold parameter by replacing the lifetime t by $t' = t - \gamma$, with t' satisfying the restriction $t' \geq 0$. For example, we consider in Section 1.3.1 the exponential distribution, in which T has p.d.f. $f(t) = \lambda \exp(-\lambda t)$, with $t \geq 0$. If a threshold parameter were introduced, the p.d.f. would be

$$f(t) = \lambda e^{-\lambda(t-\gamma)} \qquad t \geq \gamma.$$

Properties of the latter distribution follow immediately from those of the former, since $T' = T - \gamma$ has p.d.f. $\lambda \exp(-\lambda t')$, with $t' \geq 0$.

1.3.1 The Exponential Distribution

The exponential distribution is characterized by a constant hazard function

$$h(t) = \lambda \qquad t \geq 0, \tag{1.3.1}$$

where $\lambda > 0$. The p.d.f. and survivor function are found from (1.2.4) and (1.2.3) to be

$$f(t) = \lambda e^{-\lambda t} \quad \text{and} \quad S(t) = e^{-\lambda t}, \tag{1.3.2}$$

respectively. The distribution is also often written using the parameterization $\theta = \lambda^{-1}$, in which case the p.d.f. becomes

$$f(t) = \theta^{-1} e^{-t/\theta} \qquad t \geq 0. \tag{1.3.3}$$

We will sometimes use the notation $T \sim \text{Exp}(\theta)$ to indicate that a random variable T has distribution (1.3.3). The mean and variance of the distribution are θ and θ^2, respectively, and the pth quantile is $t_p = -\theta \log(1 - p)$. The distribution where $\theta = 1$ is called the standard exponential distribution; its p.d.f. is shown in Figure 1.3. Clearly, if T has p.d.f. (1.3.2), then $\lambda T \sim \text{Exp}(1)$.

Historically, the exponential was the first widely discussed lifetime distribution model. This was in part because of the availability of simple statistical methods for it. The assumption of a constant hazard function is very restrictive, so the model's applicability is fairly limited. Statistical inference under an exponential model is considered in Chapter 3.

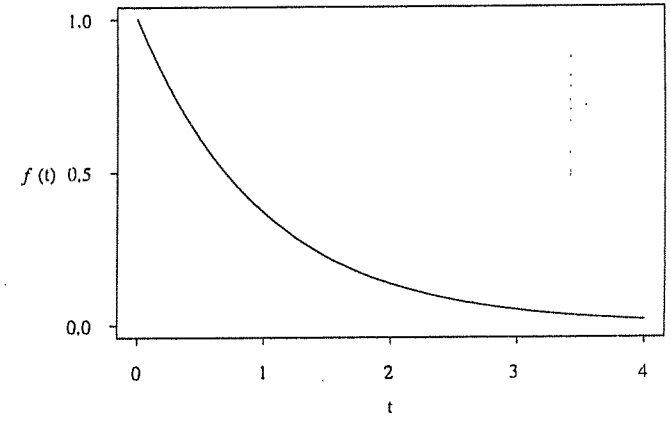

Figure 1.3. The standard exponential p.d.f.

1.3.2 The Weibull Distribution

The Weibull distribution is perhaps the most widely used lifetime distribution model. Application to the lifetimes or durability of manufactured items is common, and it is used as a model with diverse types of items, such as ball bearings, automobile components, and electrical insulation. It is also used in biological and medical applications, for example, in studies on the time to the occurrence of tumors in human populations or in laboratory animals.

The Weibull distribution has a hazard function of the form

$$h(t) = \lambda\beta(\lambda t)^{\beta-1}, \tag{1.3.4}$$

where $\lambda > 0$ and $\beta > 0$ are parameters. It includes the exponential distribution as the special case where $\beta = 1$. By (1.2.4) and (1.2.3), the p.d.f. and survivor function of the distribution are

$$f(t) = \lambda\beta(\lambda t)^{\beta-1}\exp[-(\lambda t)^{\beta}] \qquad t > 0 \tag{1.3.5}$$

and

$$S(t) = \exp[-(\lambda t)^{\beta}] \qquad t > 0. \tag{1.3.6}$$

The rth moment $E(X^r)$ of the distribution is $\lambda^{-r}\Gamma(1 + r/\beta)$, where

$$\Gamma(k) = \int_0^{\infty} u^{k-1}e^{-u}\,du \qquad k > 0$$

is the gamma function (see Appendix B). The mean and variance are thus $\lambda^{-1}\Gamma(1 + 1/\beta)$ and $\lambda^{-2}[\Gamma(1 + 2/\beta) - \Gamma(1 + 1/\beta)^2]$.

The Weibull hazard function is monotone increasing if $\beta > 1$, decreasing if $\beta < 1$, and constant for $\beta = 1$. The model is fairly flexible (see Fig. 1.4) and has been found to provide a good description of many types of lifetime data. This and the fact that the model has simple expressions for the p.d.f. and survivor and hazard functions partly account for its popularity. The Weibull distribution arises as an asymptotic extreme value distribution (see Problem 1.12), and in some instances this can be used to provide motivation for it as a model.

The scale parameter $\alpha = \lambda^{-1}$ is often used in place of λ. The pth quantile corresponding to (1.3.6) is then

$$t_p = \alpha[-\log(1 - p)]^{1/\beta}, \tag{1.3.7}$$

and by putting $p = 1 - e^{-1} = .632$ into (1.3.7) we see that α is the .632 quantile of the distribution, regardless of the value of β. In some areas, especially in engineering, α is termed the characteristic life of the distribution.

The notation $T \sim \text{Weib}(\alpha, \beta)$ will occasionally be used to indicate that a random variable T has distribution (1.3.5) with $\lambda = \alpha^{-1}$. The shape of the Weibull p.d.f. and

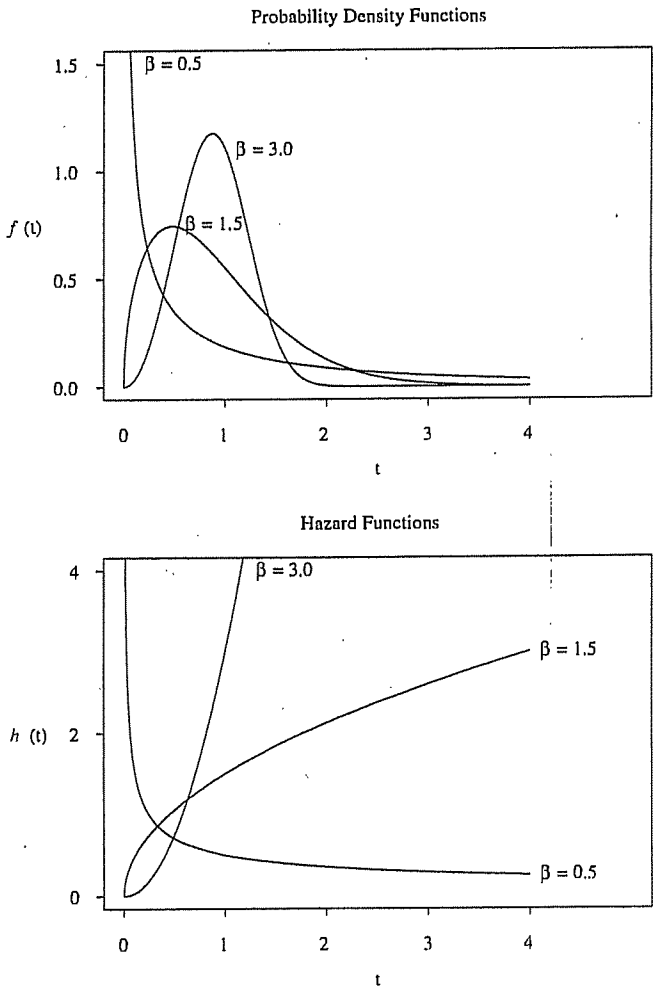

Figure 1.4. Weibull p.d.f.'s and hazard functions for $\lambda = 1$ and $\beta = 0.5$, 1.5, and 3.0.

hazard function depends only on β, which is sometimes called the shape parameter for the distribution. Typical β values vary from application to application, but in many situations distributions with β in the range 0.5 to 3 are appropriate. Figure 1.4 shows some Weibull p.d.f.'s and the corresponding hazard functions for $\lambda = 1$ and several values of β. Note that the effect of different values of λ in Figure 1.4 is just to change the scale on the horizontal (t) axis, and not the basic shape of the graph.

1.3.2.1 *The Extreme Value Distribution*

It is convenient at this point to introduce a distribution that is closely related to the Weibull distribution. This is the so-called first asymptotic distribution of extreme values, hereafter referred to simply as the extreme value distribution. This distribution is also sometimes referred to as the Gumbel distribution. Our interest in it arises because if T has a Weibull distribution, then $\log T$ has an extreme value distribution.

The p.d.f. and survivor function for the extreme value distribution are, respectively,

$$f(y) = b^{-1} \exp\left[\frac{y-u}{b} - \exp\left(\frac{y-u}{b}\right)\right] \qquad -\infty < y < \infty \qquad (1.3.8)$$

$$S(y) = \exp\left[-\exp\left(\frac{y-u}{b}\right)\right] \qquad -\infty < y < \infty, \qquad (1.3.9)$$

where $b > 0$ and $u (-\infty < u < \infty)$ are parameters. It is easily seen that if T has a Weibull distribution with p.d.f. (1.3.5), then $Y = \log T$ has an extreme value distribution with $b = \beta^{-1}$ and $u = -\log\lambda = \log\alpha$. In analyzing data it is often convenient to work with log lifetimes, so the extreme value distribution is frequently encountered.

We use the notation $Y \sim EV(u, b)$ to indicate that the random variable Y has p.d.f. (1.3.8). The extreme value distribution $EV(0, 1)$ with $u = 0$ and $b = 1$ is termed the standard extreme value distribution. A graph of its p.d.f. is given in Figure 1.5. Clearly, if $Y \sim EV(u, b)$ then $(Y - u)/b \sim EV(0, 1)$. Since u is a location and b a scale parameter, values of u and b different from 0 and 1 do not affect the shape of $f(y)$, but only the location and scale.

Moments of the distribution are conveniently obtained via the moment generating function. For the standard extreme value distribution, this is

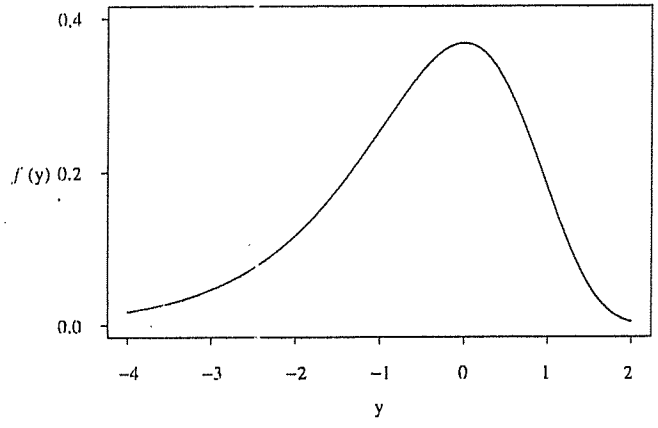

Figure 1.5. The standard extreme value p.d.f.

$$M(\theta) = \int_{-\infty}^{\infty} e^{\theta x} \exp(x - e^x)\, dx.$$

Letting $u = e^x$, we have

$$M(\theta) = \int_0^{\infty} u^{\theta} e^{-u}\, du$$
$$= \Gamma(1 + \theta).$$

The mean of the standard extreme value distribution is found from this to be $\Gamma'(1) = -\gamma$, where $\gamma = 0.5772\cdots$ is known as Euler's constant; the variance is $\Gamma''(1) - \gamma^2 = \pi^2/6$ (see Appendix B). The mean and variance of the general distribution (1.3.8) are $u - \gamma b$ and $(\pi^2/6)b^2$, since $(Y - u)/b$ has the standard extreme value distribution. The pth quantile of (1.3.8) is

$$y_p = u + b \log[-\log(1 - p)]$$

which implies that the location parameter u is the .632 quantile.

The statistical analysis of data under a Weibull distribution model is discussed in Chapter 5, and the extreme value distribution is considered further there.

1.3.3 The Log-Normal Distribution

The log-normal distribution has been used as a model in diverse applications in engineering, medicine, and other areas. The lifetime T is said to be log-normally distributed if $Y = \log T$ is normally distributed, say with mean μ, variance σ^2, and p.d.f.

$$\frac{1}{(2\pi)^{1/2}\sigma} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right] \qquad -\infty < y < \infty.$$

From this the p.d.f. of $T = \exp Y$ is easily found to be

$$f(t) = \frac{1}{(2\pi)^{1/2}\sigma t} \exp\left[-\frac{1}{2}\left(\frac{\log t - \mu}{\sigma}\right)^2\right] \qquad t > 0. \qquad (1.3.10)$$

The survivor and hazard functions for the log-normal distribution involve the standard normal distribution function

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{(2\pi)^{1/2}} e^{-u^2/2}\, du.$$

The log-normal survivor function is easily seen to be

$$S(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right), \tag{1.3.11}$$

and the hazard function is given as $h(t) = f(t)/S(t)$.

The hazard function can be shown (see Problem 1.4) to have the value 0 at $t = 0$, increase to a maximum, and then decrease, approaching 0 as $t \to \infty$. This shape arises in many situations, for example, when a population consists of a mixture of

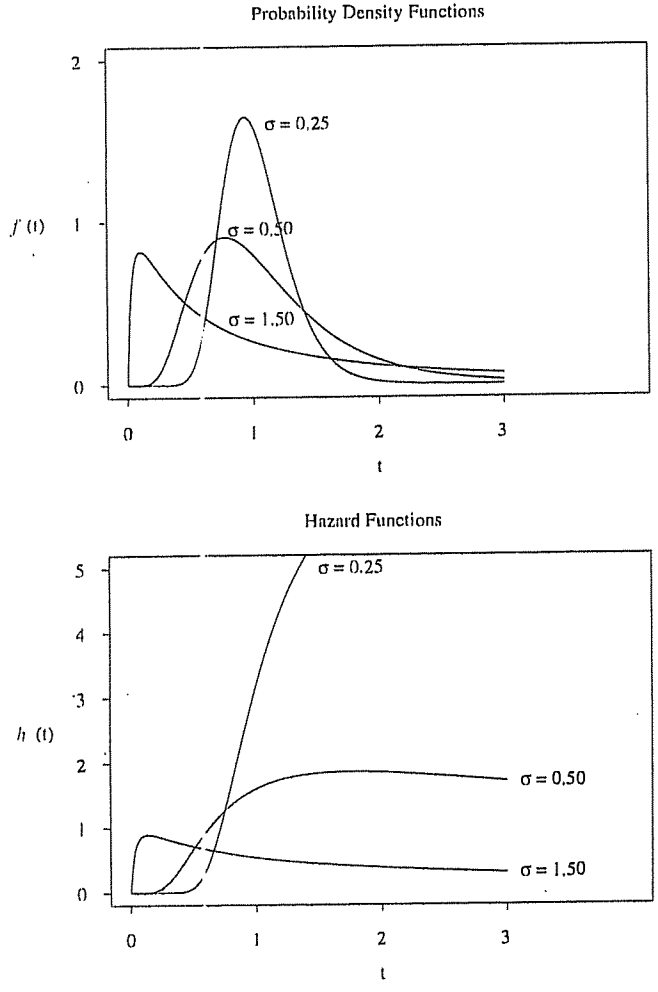

Figure 1.6. Log-normal p.d f.'s and hazard functions for $\mu = 0$ and $\sigma = 0.25$, 0.5, and 1.5.

individuals who tend to have short and long lifetimes, respectively. Examples include survival after treatment for some forms of cancer, where persons who are cured become long-term survivors, and the duration of marriages, where after a certain number of years the risk of marriage dissolution due to divorce tends to decrease.

The notation $Y \sim N(\mu, \sigma^2)$ is used to denote that Y is normal with mean μ and variance σ^2, and $T \sim \text{Log}\,N(\mu, \sigma^2)$ is used to denote that T has p.d.f. (1.3.10). Figure 1.6 shows some log-normal p.d.f.'s and hazard functions for $\mu = 0$ and different values of σ. It should be noted that a nonzero value of μ just changes the scale on the time axis, and not the basic shapes of the functions portrayed, since μ is a location parameter for Y $(= \log T)$ and e^{μ} is a scale parameter in (1.3.10).

Some additional properties of the log-normal distribution are discussed in Problem 1.4, including the fact that the mean and variance are $\exp(\mu + \sigma^2/2)$ and $[\exp(\sigma^2) - 1][\exp(2\mu + \sigma^2)]$, respectively. The median $(t_{.50})$ is $\exp(\mu)$. Statistical inference for log-normal distributions is considered in Chapter 5.

1.3.4 The Log-Logistic Distribution

The log-logistic distribution has p.d.f. of the form

$$f(t) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{[1 + (t/\alpha)^{\beta}]^2}, \qquad t > 0, \tag{1.3.12}$$

where $\alpha > 0$ and $\beta > 0$ are parameters. The survivor function and hazard function are, respectively,

$$S(t) = [1 + (t/\alpha)^{\beta}]^{-1}$$
$$h(t) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{[1 + (t/\alpha)^{\beta}]}. \tag{1.3.13}$$

The log-logistic gets its name from the fact that $Y = \log T$ has a logistic distribution with p.d.f.

$$f(y) = \frac{b^{-1}\exp[(y-u)/b]}{[1 + \exp[(y-u)/b]]^2} \qquad -\infty < y < \infty, \tag{1.3.14}$$

where $u = \log\alpha$ and $b = \beta^{-1}$, so that $-\infty < u < \infty$ and $b > 0$. We use the notation $Y \sim \text{Logist}(u, b)$ to indicate that Y has p.d.f. $f(y)$, and $T \sim \text{LLogist}(\alpha, \beta)$ to indicate that T has p.d.f. (1.3.12).

The rth moment of T exists and is given by $E(T^r) = \alpha^r\Gamma(1 + r/\beta)\Gamma(1 - r/\beta)$, provided $\beta > r$ (see Problem 1.5). The mean therefore exists only if $\beta > 1$, in which case $E(T) = \alpha\Gamma(1 + \beta^{-1})\Gamma(1 - \beta^{-1})$. The moments of $Y = \log T$ are easily found via its moment generating function; the mean equals u and the variance is $\pi^2 b^2/3$ (Problem 1.5). Figure 1.7 shows p.d.f.'s and hazard functions of T for $b = 0.14$, 0.28, and 0.83. These are chosen so that the variance of Y is roughly the same as the

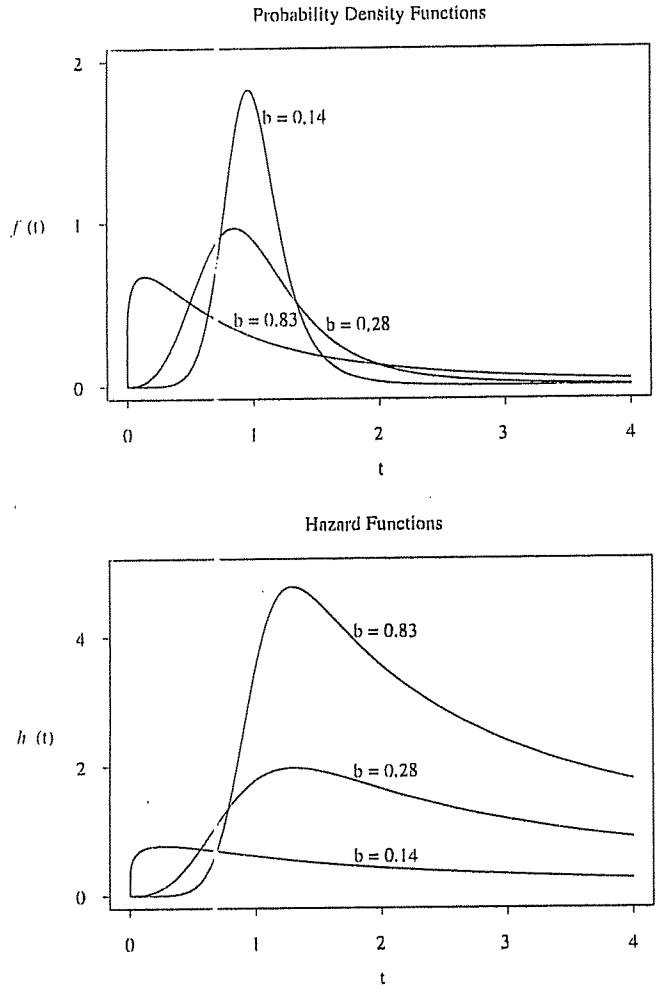

Figure 1.7. Log-logistic p.d.f.'s and hazard functions for $a = 0$ and $b = 0.14$, 0.28, and 0.83.

variance of a normal distribution with $\sigma = 0.25$, 0.5 and 1.5, respectively. Note the similarity of Figures 1.6 and 1.7.

The logistic and normal distributions have similar shapes, and it is easily seen (Problem 1.5) that for $\beta > 1$ the hazard function has the same characteristic shape as the log-normal: it has $h(0) = 0$, increases to a maximum, and then approaches 0 monotonically as $t \to \infty$. For $\beta \leq 1$ the hazard function is monotone decreasing. Inference for the log-logistic and logistic distributions is discussed in Chapter 5.

1.3.5 The Gamma Distribution

The gamma distribution has a p.d.f. of the form

$$f(t) = \frac{\lambda(\lambda t)^{k-1}e^{-\lambda t}}{\Gamma(k)} \qquad t > 0, \tag{1.3.15}$$

where $k > 0$ and $\lambda > 0$ are parameters; λ^{-1} is a scale parameter and k is sometimes called the index or shape parameter. This distribution, like the Weibull distribution, includes the exponential as a special case ($k = 1$). The survivor and hazard functions involve the incomplete gamma function [see (B.12) of Appendix B]

$$I(k, x) = \frac{1}{\Gamma(k)} \int_0^x u^{k-1}e^{-u}\, du. \tag{1.3.16}$$

Integrating (1.3.15), we find that the survivor function is

$$S(t) = 1 - I(k, \lambda t).$$

The hazard function is $h(t) = f(t)/S(t)$; it can be shown (see Problem 1.6) to be monotone increasing for $k > 1$, with $h(0) = 0$ and $\lim_{t\to\infty} h(t) = \lambda$. For $0 < k < 1$, $h(t)$ is monotone decreasing, with $\lim_{t\to 0} h(t) = \infty$ and $\lim_{t\to\infty} h(t) = \lambda$.

The distribution with $\lambda = 1$ is called the one-parameter gamma distribution and has p.d.f.

$$f(t) = \frac{t^{k-1}e^{-t}}{\Gamma(k)} \qquad t > 0, \tag{1.3.17}$$

Its c.d.f. is given by (1.3.16). The notation $Y \sim Ga(k)$ will be used to indicate that a random variable Y has p.d.f. (1.3.17). Note that if T has p.d.f. (1.3.15), then $\lambda T \sim Ga(k)$. The one-parameter gamma distribution is closely related to the chi-squared (χ^2) distribution: if $Y \sim Ga(k)$, then $2Y$ has a χ^2 distribution with $2k$ degrees of freedom, henceforth simply referred to as $\chi^2_{(2k)}$. Figure 1.8 shows p.d.f.'s and hazard functions for a few gamma distributions.

The moment generating function of (1.3.17) is

$$M(\theta) = \int_0^\infty \frac{e^{\theta t}t^{k-1}e^{-t}dt}{\Gamma(k)}$$
$$= (1-\theta)^{-k},$$

and that of (1.3.15) is $(1 - \theta/\lambda)^{-k}$. The moments of the distribution can be found from this; for example, $E(T^r) = k(k+1)\cdots(k+r-1)$ for (1.3.17).

The gamma distribution is not used as a lifetime model as much as the Weibull, log-normal, and log-logistic distributions. It does fit a variety of lifetime data adequately, however. It also arises in some situations involving the exponential distribution, because of the well-known result that sums of independent and identically

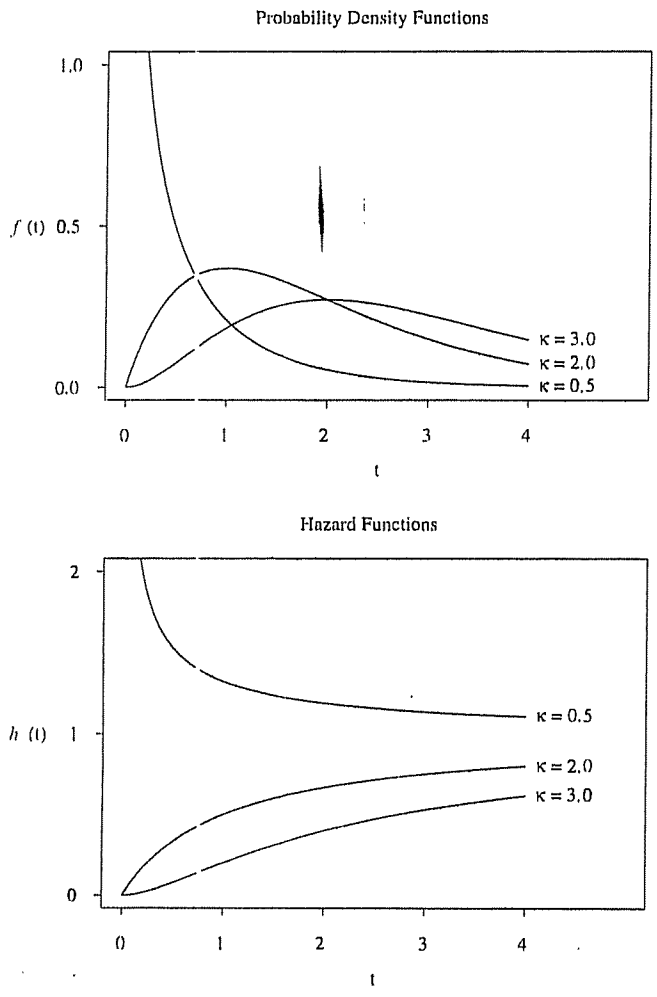

Figure 1.8. Gamma p.d.f.'s and hazard functions for $\lambda = 1$ and $k = 0.5$, 2.0, and 3.0.

distributed (i.i.d.) exponential random variables have a gamma distribution. Specifically, if $T_1, \ldots, T_n$ are independent, each with p.d.f. (1.3.2), then $T_1 + \cdots + T_n$ has a gamma distribution with parameters λ and $k = n$.

1.3.6 Log-Location-Scale Models

A parametric location-scale model for a random variable Y on $(-\infty, \infty)$ is a distribution with p.d.f. of the form

$$f(y) = \frac{1}{b} f_0\left(\frac{y-u}{b}\right) \qquad -\infty < y < \infty, \tag{1.3.18}$$

where $u(-\infty < u < \infty)$ and $b > 0$ are location and scale parameters, and $f_0(z)$ is a specified p.d.f. on $(-\infty, \infty)$. The distribution and survivor functions for Y are $F_0[(y-u)/b]$ and $S_0[(y-u)/b]$, respectively, where

$$F_0(z) = \int_{-\infty}^{z} f_0(w)\, dw = 1 - S_0(z).$$

The standardized random variable $Z = (Y-u)/b$ clearly has p.d.f. and survivor functions $f_0(z)$ and $S_0(z)$, and (1.3.18) with $u = 0$, $b = 1$ is called the standard form of the distribution.

The lifetime distributions introduced in Sections 1.3.2 to 1.3.4 all have the property that $Y = \log T$ has a location-scale distribution: the Weibull, log-normal, and log-logistic distributions for T correspond to extreme value, normal, and logistic distributions for Y. The survivor functions for $Z = (Y-u)/b$ are, respectively,

$$\begin{aligned} S_0(z) &= \exp(-e^z) && \text{extreme value} \\ S_0(z) &= 1 - \Phi(z) && \text{normal} \\ S_0(z) &= (1+e^z)^{-1} && \text{logistic,} \end{aligned}$$

where $-\infty < z < \infty$ and $\Phi(z)$ is given just before (1.3.11). By the same token, any location-scale model (1.3.18) gives a lifetime distribution through the transformation $T = \exp(Y)$. Note that the survivor function for T can in this case be expressed as

$$\begin{aligned} Pr(T \geq t) &= S_0\left(\frac{\log t - u}{b}\right) \\ &= S_0^*\left[\left(\frac{t}{\alpha}\right)^{\beta}\right], \end{aligned} \tag{1.3.19}$$

where $\alpha = \exp(u)$, $\beta = b^{-1}$, and $S_0^*(x)$ is a survivor function defined on $(0, \infty)$ by the relationship $S_0^*(x) = S_0(\log x)$.

Families of distributions with three or more parameters can be obtained by generalizing (1.3.18) to let $f_0(z)$, $F_0(z)$, or $S_0(z)$ include one or more "shape" parameters. We mention two such families that are useful because they include common two-parameter lifetime distributions as special cases.

The first model is the generalized log-Burr family, for which the standardized variable $(Y-u)/b$ has survivor function of the form

$$S_0(z; k) = \left(1 + \frac{1}{k} e^z\right)^{-k} \qquad -\infty < z < \infty, \tag{1.3.20}$$

where $k > 0$ is a third parameter; it is easily verified that (1.3.20) is a survivor function for all $k > 0$. The special case $k = 1$ gives the standard logistic distribution (see (1.3.14)), and the limit as $k \to \infty$ gives the extreme value distribution (see (1.3.9)). The family of lifetime distributions obtained from (1.3.20) is given by (1.3.19) and has

$$Pr(T \geq t) = \left[1 + \frac{1}{k}\left(\frac{t}{\alpha}\right)^{\beta}\right]^{-k}. \tag{1.3.21}$$

The log-logistic survivor function is given by $k = 1$, and the Weibull survivor function is given by the limit as $k \to \infty$. Figure 1.9 shows p.d.f.'s for log-Burr distributions (1.3.20) with $k = .5$, 1, 10, and ∞. Note that $E(Z)$ and $\mathrm{Var}(Z)$ vary with k (see Problem 1.9) so that the distributions in Figure 1.9 do not have identical means and standard deviations.

Since the generalized log-Burr family includes the log-logistic and Weibull distributions, it allows discrimination between them. It is also a flexible model for fitting to data; inference for it is discussed in Chapter 5.

A second extended model is the generalized log-gamma distribution, which includes the Weibull and log-normal distributions as special cases. The model was originally introduced by specifying that $(T/\alpha)^{\beta}$ has a one-parameter gamma distribution (1.3.17) with index parameter $k > 0$. Equivalently, $W = (Y - u_1)/b_1$, where $Y = \log T$, $u_1 = \log \alpha$ and $b_1 = \beta^{-1}$, has a log-gamma distribution. However, the mean and the variance for the gamma distribution both equal k, and as k increases, the gamma and log-gamma distributions do not have limits. The mean and variance for W are (see Problem 1.10) $E(W) = \psi(k)$ and $\mathrm{Var}(W) = \psi'(k)$, where ψ and ψ' are the digamma and trigamma functions (see Appendix B.2). For large k they

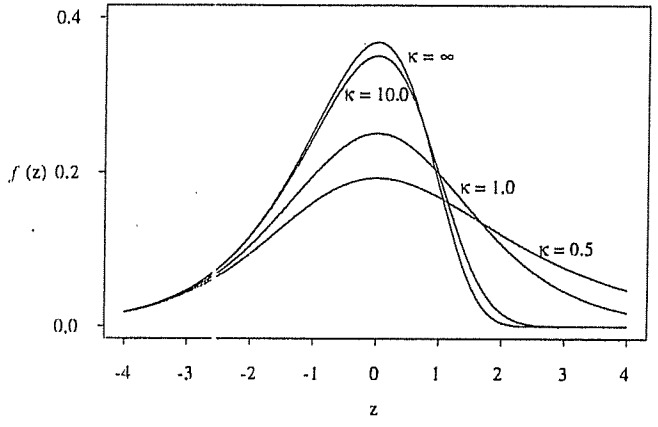

Figure 1.9. P.d.f.'s of log-Burr distributions for $k = .5, 1, 10, \infty$.

behave like $\log k$ and k^{-1}, respectively (see (B9)), and it is therefore convenient and customary to define a transformed log-gamma variate $Z = k^{1/2}(W - \log k)$, which has p.d.f. (see Problem 1.10)

$$f_0(z; k) = \frac{k^{k-1/2}}{\Gamma(k)} \exp(k^{1/2}z - ke^{k^{-1/2}z}) \qquad -\infty < z < \infty. \tag{1.3.22}$$

The generalized log-gamma model is then the three-parameter family of distributions for which $Z = (Y - u)/b$ has p.d.f. (1.3.22); the corresponding distribution of $T = \exp(Y)$ is obtained from this, and is called the generalized gamma model. Figure 1.10 shows p.d.f.'s (1.3.22) for $k = .5$, 1, 10, and ∞. As for the log-Burr distributions in Figure 1.9, note that $E(Z)$ and $\mathrm{Var}(Z)$ vary with k.

For the special case $k = 1$, (1.3.22) becomes the standard extreme value p.d.f. (see (1.3.8)). It can also be shown (see Problem 1.10) that as $k \to \infty$, (1.3.22) converges to the standard normal p.d.f., and thus the generalized gamma model includes the Weibull and log-normal distributions as special cases. The two-parameter gamma distribution (1.3.15) also arises as a special case; in the original (α, β, k) parameterization this corresponds to $\beta = 1$, and in the (u, b, k) parameterization with (1.3.22), to $b = k^{-1/2}$. Inference for the generalized gamma and log-gamma distributions is discussed in Chapter 5.

Other extended families may be useful from time to time. For example, one might take $Z = (Y - u)/b$ to have a Student t distribution with k degrees of freedom. Kalbfleisch and Prentice (1980, Sec. 2.2.7) consider a four-parameter model in which Z is a rescaled $\log F$ random variable; it includes the generalized log-Burr and log-gamma families as special cases.

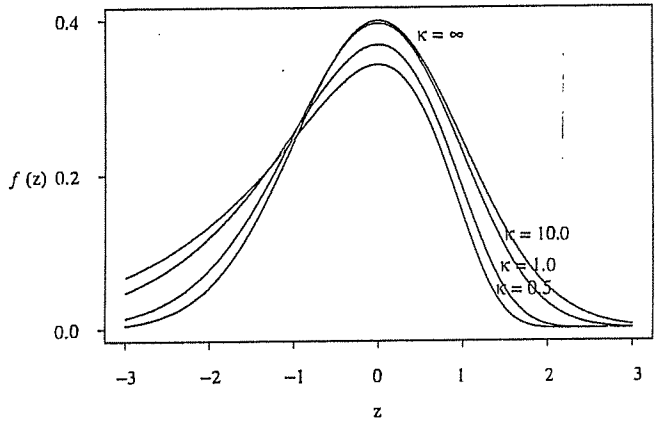

Figure 1.10. P.d.f.'s of log-gamma distributions for $k = .5, 1, 10, \infty$.

1.3.7 The Inverse Gaussian Distribution

The inverse Gaussian distribution arises as the time until a continuous-time Wiener process with drift parameter $\gamma > 0$ and dispersion parameter σ^2 first crosses a given threshold level $d > 0$. The Wiener process is a Gaussian stochastic process $\{X(t),\ t > 0\}$, with $X(0) = 0$, and one of its properties is that $X(t) \sim N(\gamma t, \sigma^2 t)$ for any specified $t > 0$ (e.g., Ross 1983). The random variable $T = \inf(t : X(t) = d)$ can be shown to have p.d.f.

$$f(t) = \frac{d}{\sigma(2\pi t^3)^{1/2}} \exp\left[\frac{-(d-\gamma t)^2}{2\sigma^2 t}\right] \qquad t > 0.$$

This p.d.f. depends only on d/γ and d/σ, and it is common to reparameterize it by defining $\mu = d/\gamma$, $\lambda = d^2/\sigma^2$ so that

$$f(t) = \frac{\lambda^{1/2}}{(2\pi t^3)^{1/2}} \exp\left[\frac{-\lambda(t-\mu)^2}{2t\mu^2}\right] \qquad t > 0. \tag{1.3.23}$$

The mean and variance of T may be shown to be $E(T) = \mu$ and $\mathrm{Var}(T) = \mu^3/\lambda$, and the c.d.f. is

$$F(t) = \Phi\left[\left(\frac{t}{\mu} - 1\right)\left(\frac{\lambda}{t}\right)^{1/2}\right] + e^{2\lambda/\mu}\Phi\left[-\left(\frac{t}{\mu} + 1\right)\left(\frac{\lambda}{t}\right)^{1/2}\right], \tag{1.3.24}$$

where $\Phi(z)$ is the standard normal c.d.f. We will denote this model by $IG(\mu, \lambda)$.

The inverse Gaussian distribution is sometimes a plausible model in settings where failure occurs when a deterioration process reaches a certain level. More generally, it is a reasonably flexible two-parameter family of models with properties that are rather similar to those of the log-normal distribution. Figure 1.11 shows the hazard and density functions for $\mu = 1$ and several values of λ.

1.3.8 Models with Piecewise Constant or Polynomial Hazard Functions

Let $a_0 < a_1 < \cdots < a_m$ be specified values with $a_0 = 0$ and $a_m = \infty$. If the hazard function for T is of the form

$$h(t) = \lambda_j, \qquad a_{j-1} \le t < a_j \tag{1.3.25}$$

where the λ_j are positive values for $j = 1, \ldots, m$, then T is said to have a piecewise-constant hazard function. This model may seem implausible, since $h(t)$ is discontinuous at the cut points $a_1, \ldots, a_{m-1}$, but with an appropriate value of m and selection of cut points, it can approximate arbitrary shapes of hazard functions and survivor functions. Further, as we will see in Chapter 4, statistical methods based on the model are straightforward, and it provides a convenient link between parametric and nonparametric methods.

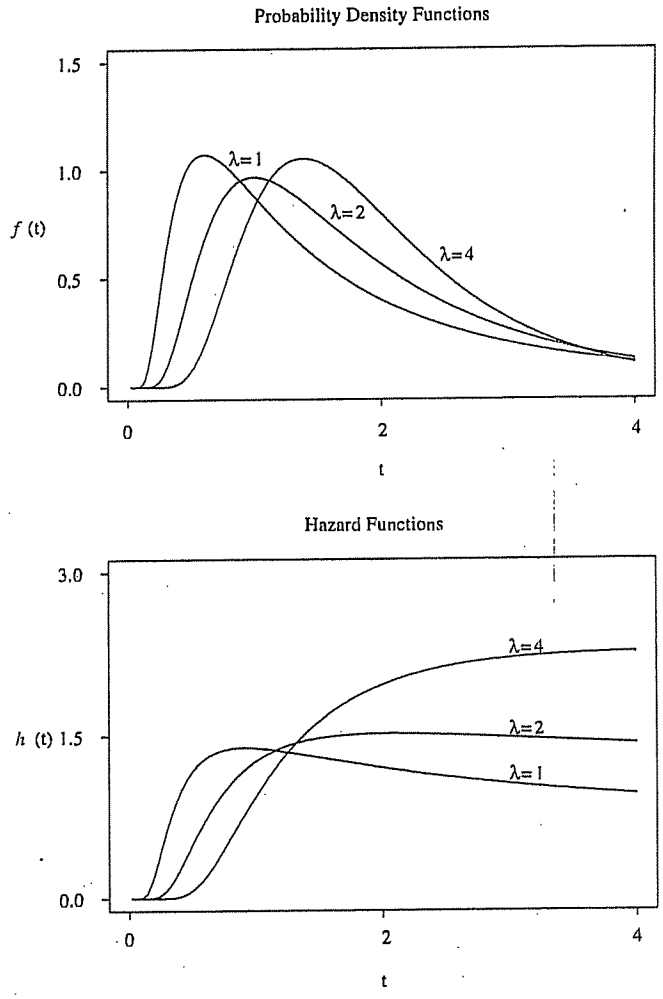

Figure 1.11. Inverse Gaussian p.d.f.'s and hazard functions for $\mu = 1$ and $\lambda = 1, 2$, and 4.

The survivor function corresponding to (1.3.25) is readily found from (1.2.3) to be

$$S(t) = \exp\left[-\sum_{j=1}^{m(t)-1} \lambda_j(a_j - a_{j-1}) - \lambda_{m(t)}(t - a_{m(t)-1})\right],$$

where $m(t)$ is defined by $a_{m(t)-1} \le t < a_{m(t)}$. The p.d.f. $f(t) = \lambda_{m(t)}S(t)$ is piecewise exponential. The survivor function is conveniently expressed by introducing the

notation below, which will be useful in the book:

$$\Delta_j(t) = \int_{a_{j-1}}^{a_j} I(u \le t)\, du \qquad j = 1, \ldots, m. \tag{1.3.26}$$

Then,

$$S(t) = \exp\left[-\sum_{j=1}^{m} \lambda_j \Delta_j(t) \right]. \tag{1.3.27}$$

Piecewise-constant hazard functions and the corresponding probability density functions are discontinuous, which makes them unappealing in many settings. Another way to obtain flexible hazard functions is to use spline functions, which consist of polynomial pieces joined at the cut points $a_1, \ldots, a_{m-1}$. A cubic spline $g(t)$ on (a_0, a_m) consists of cubic polynomial pieces that are designed so that $g(t)$ and its first two derivatives are everywhere continuous on (a_0, a_m). In particular, they are continuous at the cut points $a_1, \ldots, a_{m-1}$, which are referred to as knots in spline terminology. One can also define splines of other orders (e.g., quadratic or linear), but cubic splines are favored for a number of reasons, and here we restrict discussion to them.

It can be seen that only $m+3$ parameters are needed to specify a cubic spline with $m-1$ knots $a_1, \ldots, a_{m-1}$. One specification is

$$g(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \sum_{j=1}^{m-1} \beta_j (t - a_j)_+^3,$$

where $x_+ = \max(x, 0)$. This is not a particularly good representation when fitting a model (i.e., when estimating $\alpha_0, \ldots, \alpha_3$ and $\beta_1, \ldots, \beta_{m-1}$), however. In general, we can write $g(t)$ as

$$g(t) = \sum_{j=1}^{m+3} \theta_j B_j(t),$$

where the $B_j(t)$ are specified piecewise cubic basis functions. A common approach is to use the so-called B-spline basis functions (e.g., de Boor 1978); software exists for generating such functions.

Spline-based hazard models usually take $h(t)$ or $\log h(t)$ as a cubic spline. An alternative is to use a cubic spline for the p.d.f. $f(t)$. Models with as few as one or two well-chosen knots provide considerable flexibility. Models where $h(t)$ or $f(t)$ are splines have the advantage that cumulative hazard and distribution functions $H(t)$ and $F(t)$ are easily obtained, but parameters may need to be constrained to keep $h(t)$ and $f(t)$ nonnegative. With a larger number of knots the end pieces over (a_0, a_1) and (a_{m-1}, a_m) can be difficult to estimate, and so linear instead of cubic pieces are often used for those intervals. When this is done, the spline is termed a

natural cubic spline. It can be seen that only $m-1$ parameters are needed to specify a natural cubic spline $g(t)$ with $m-1$ knots.

Spline models are discussed further in Section 4.2.3.

1.3.9 Some Other Models

The distributions described in preceding sections are the most frequently used parametric models. Other models are, however, sometimes used in applications. We list a few models with references from which more information can be obtained.

1. Distributions in which either $h(t)$ or $\log[h(t)]$ is a low-order polynomial (e.g., Bain 1974; Gehan and Siddiqui 1973). Models of this type include the Gompertz distribution, with $h(t) = \exp(\alpha + \beta t)$, and the Rayleigh, or linear hazard rate, distribution, with $h(t) = a + bt$. Note that for the Weibull distribution, the log hazard function is linear in $\log t$. Models where $h(t)$ or $\log h(t)$ is a linear combination of specified functions can also be considered.
2. Models with bathtub-shaped hazards. Glaser (1980) and Hjorth (1980) discuss distributions with nonmonotone hazard functions, and provide references.
3. Discrete distributions. Usually, when a discrete model is used with lifetime data, it is a multinomial distribution. Methods for multinomial models are discussed in Chapter 3. Occasionally, discrete parametric distributions are used; often these are based on one of the common continuous life distributions.

1.3.10 Mixture Models

Discrete mixture models arise when individuals in a population are each one of k distinct types, with a proportion, p_i, of the population being of the ith type; the p_i satisfy $0 < p_i < 1$ and $\sum p_i = 1$. Individuals of type i are assumed to have a lifetime distribution with survivor function $S_i(t)$. An individual randomly selected from this population then has survivor function

$$S(t) = p_1 S_1(t) + \cdots + p_k S_k(t). \tag{1.3.28}$$

Models of this kind are termed discrete mixture models, and are useful in situations where the population is heterogeneous but it is not possible to distinguish between individuals of different types. Often the $S_i(t)$ in (1.3.28) are taken to be from the same parametric family, though this is, of course, unnecessary. The properties of a mixture model follow from the properties of the k distributions, or components, involved in the mixture. Estimation can be difficult, and models with k larger than 3 are rarely used.

Two special models with $k = 2$ are important. One has a degenerate component with a probability mass at $T = \infty$. The survivor function for this model is

$$S(t) = pS_1(t) + 1 - p \qquad t \ge 0, \tag{1.3.29}$$

where $0 < p < 1$, and $S_1(t)$ is a survivor function with $S_1(\infty) = 0$. This is used in settings where some fraction $1-p$ of individuals in a population have very long life-

times, which for convenience are assumed infinite. In medical applications involving treatment of disease, (1.3.29) is sometimes referred to as a cure-rate model.

The second special model has a degenerate component with a probability mass at $T = 0$. The survivor function is

$$S(t) = pS_1(t) \qquad t > 0, \tag{1.3.30}$$

where $0 < p < 1$, and $S_1(\cdot)$ is a survivor function. This is used in settings where some fraction $1 - p$ of individuals in a population die or fail at $t = 0$; one application is to manufactured items that defects may render inoperative.

Continuous mixture models can also be considered. They have survivor functions of the form

$$S(t) = \int_{-\infty}^{\infty} S_1(t|z)g(z)\,dz, \tag{1.3.31}$$

where z is an unobservable random variable with p.d.f. $g(z)$ and $S_1(t|z)$ is the survivor function for T, given z. The most widely studied and used models assume that the hazard function for T given z is $zh_0(t)$, where $z > 0$ and $h_0(t)$ is a baseline hazard function. In this case

$$S(t) = \int_{0}^{\infty} e^{-zH_0(t)}g(z)\,dz, \tag{1.3.32}$$

where $H_0(t) = \int_0^t h_0(u)\,du$. Such models are called "frailty" models; the name comes from thinking of z as a factor that renders an individual's hazard function $zh_0(t)$ bigger or smaller than the baseline $h_0(t)$.

Discrete and continuous mixture models can be discussed within a single framework by replacing (1.3.31) with the Riemann–Stieltjes integral

$$S(t) = \int_{0}^{\infty} S_1(t|z)\,dG(z), \tag{1.3.33}$$

where $G(z)$ is a distribution function as in (1.2.10).

Since the random variable z is unobservable, (1.3.33) can be viewed as merely a way to generate new models $S(t)$. However, insight into the effects of heterogeneity in populations can be gained from such models; see Problem 1.14. Inference for mixture models is discussed in Chapter 4.

1.4 REGRESSION MODELS

The use of explanatory variables, or covariates, in a regression model is an important way to represent heterogeneity in a population. Indeed, the main objective in many studies is to understand and exploit the relationship between lifetime and covariates. Thus, data often include covariates that might be related to lifetime: for example, in a survival study for lung cancer patients (see Example 1.1.9) factors such as the age and general condition of the patient, and the type of tumor, were recorded. In experiments on the time to failure of electrical insulation an important factor is the voltage the insulation is subjected to (see Example 1.1.5). In clinical trials in medicine, the treatment assigned to a patient may be considered a covariate (Example 1.1.8).

Regression models for lifetimes can be formulated in many ways, and several types are in common use. Regression analysis is discussed at length in Chapters 6, 7, and 8. We provide a brief introduction here.

Any of the parametric models discussed in this chapter can be made into a regression model by specifying a relationship between the model parameters and covariates. Suppose that each individual in a population has a lifetime T and a column vector $\mathbf{x} = (x_1, \ldots, x_p)'$ of covariates. Then, for example, an exponential distribution model (see Section 1.3.1) would assume that given $\mathbf{x}$, the distribution of T is exponential with survivor function

$$S(t|\mathbf{x}) = \exp[-\lambda(\mathbf{x})t]. \tag{1.4.1}$$

The specification of the model also involves a functional form for $\lambda(\mathbf{x})$. A common form is $\lambda(\mathbf{x}) = \exp(\boldsymbol{\beta}'\mathbf{x})$, where $\boldsymbol{\beta}$ is a $p \times 1$ column vector of regression coefficients; this has the convenient property that $\lambda(\mathbf{x}) \geq 0$ for all real vectors $\boldsymbol{\beta}$ and $\mathbf{x}$.

Often only certain parameters in a lifetime distribution are assumed to depend on covariates. For example, log-location-scale models (1.3.18), with only u depending on $\mathbf{x}$, are useful. The specification $u(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$ gives a model where $Y = \log T$ has survivor function of the form

$$Pr(Y \geq y|\mathbf{x}) = S_0\left(\frac{y - \boldsymbol{\beta}'\mathbf{x}}{b}\right), \tag{1.4.2}$$

where $b > 0$ is a scale parameter. Such models are familiar in ordinary regression analysis, particularly for the case where $S_0(z)$ is the standard normal survivor function.

Semiparametric models are also widely used; they specify the dependence of T or Y on $\mathbf{x}$ parametrically, but leave the actual distribution arbitrary. For example, ordinary least-square estimation of $\boldsymbol{\beta}$ when $E(Y|\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$ can be viewed as an estimation procedure for a model (1.4.2) where $S_0(z)$ is unspecified aside from the assumption about $E(Y|\mathbf{x})$. The best known semiparametric lifetime regression model is the proportional hazards model introduced by Cox (1972a), which takes the hazard function for T given $\mathbf{x}$ to be of the form

$$h(t|\mathbf{x}) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{x}), \tag{1.4.3}$$

where $h_0(t)$ is an arbitrary "baseline" hazard function.

As discussed in Section 1.2.4, covariates may vary over time. In this case models cannot simply be specified in terms of survivor functions like (1.4.1) and (1.4.2), because there is an entire "history," $X = \{\mathbf{x}(t),\ t \geq 0\}$, to consider for a covariate. Sometimes a time-varying covariate may be linked physically with the lifetime

process: for example, blood pressure may be linked to the time or age at which an individual has a first stroke. Such covariates are termed internal, and their treatment requires care. A covariate process $X = \{\mathbf{x}(t), t \geq 0\}$, which develops independently of the lifetime process, is termed external; factors such as air pollution or climatic conditions, or applied stresses such as voltage or temperature in life test experiments, are examples. We restrict further discussion in this section to external covariates; note that fixed (constant) covariates are external.

A convenient approach to modeling with time-varying covariates is through the hazard function, which may be allowed to depend on previous covariate history. Let $X(t) = \{\mathbf{x}(s),\ 0 \leq s \leq t\}$ denote the history up to time t, with $X(\infty) = X$. It is a natural assumption that the hazard function for T given X depend only on $X(t)$; we denote this as $h(t|X(t))$. A simple but flexible approach is to define a vector $\mathbf{w}(t)$ that represents features of $X(t)$, then specify $h(t|X(t))$ as a function of t and $\mathbf{w}(t)$. The multiplicative formulation

$$h(t|X(t)) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{w}(t)) \tag{1.4.4}$$

is useful; it is an extension of the proportional hazards model (1.4.3).

A connection between the hazard function and survival probabilities can still be made in the usual way. If T has hazard function $h(t|X) = h(t|\mathbf{w}(t))$, then it follows from the argument leading to (1.2.16) that

$$Pr(T \geq t|X) = \prod_{(0,t)}[1 - dH(u|\mathbf{w}(u))]. \tag{1.4.5}$$

In the case where T is continuous, $dH(u|\mathbf{w}(u)) = h(u|\mathbf{w}(u))\,du$, where $h(u|\mathbf{w}(u))$ is assumed continuous, except possibly at a finite set of points in any interval, and (1.2.12) gives

$$Pr(T \geq t|X) = \exp\left\{-\int_0^t h(u|\mathbf{w}(u))\,du\right\}. \tag{1.4.6}$$

Regression models for the case of discrete lifetimes T are also conveniently formulated via the hazard function, giving $h(t|\mathbf{w}(t))$ for fixed or time-varying covariates. The relationship (1.4.5) still holds, with $dH(t|\mathbf{w}(t)) = h(a_j|\mathbf{w}(a_j))$ at times a_j, where $Pr(T = t|X) > 0$, and 0 elsewhere.

1.5 MULTIPLE LIFETIMES AND MULTIPLE MODES OF FAILURE

Sometimes two or more lifetime variables $T_1, \ldots, T_k$ are of interest simultaneously; there are several types of settings where this may occur. One is where there are k separate lifetimes for an individual: for example, the lifetimes of left and right front brake pads in a car or the times until a particular condition appears in the left and right eyes of a person. A similar situation occurs when individuals fall into clusters or groups, with the expectation being that the lifetimes $T_1, \ldots, T_k$ of k individuals

in the same group have some degree of association. In carcinogenicity experiments involving the time to the appearance of a tumor in laboratory animals, for example, we would expect some association for animals in the same litter (i.e., with the same parents).

These situations lead to a consideration of multivariate lifetime distributions, which can be specified in terms of a joint p.d.f. or a joint survivor function,

$$S(t_1, \ldots, t_k) = Pr(T_1 \geq t_1, \ldots, T_k \geq t_k). \tag{1.5.1}$$

A full treatment of multivariate models is beyond the scope of this book; separate treatments of this area exist (e.g., Joe 1997, Hougaard 2000). We provide a brief discussion in Section 11.1, focusing on problems where methods based on univariate lifetimes may be adopted.

Another class of situations is where a sequence of times $T_1, T_2, \ldots$ can occur for a single individual. For a repairable system, for example, T_j could be the time between the $(j-1)$st and jth failures. In Example 1.1.8, patients with cystic fibrosis could experience successive intervals of time free from pulmonary infections, of lengths $T_1, T_2, \ldots$. In these circumstances the lifetime T_j can be observed only if $T_1, \ldots, T_{j-1}$ have already been observed, and it is natural to consider a sequence of models $f(t_1)$, $f(t_2|t_1)$, $f(t_3|t_1, t_2)$, and so on. This allows univariate lifetime models to be used; we consider examples in Section 11.2.

A third kind of multivariate problem occurs when there is a single lifetime for each individual, but failure or death may be of different modes or types. Often the modes refer to causes of failure, in which case the term "competing risks" is sometimes used. For example, an individual in a demographic study might be recorded as dying at age T from one of cancer, cardiovascular disease, or "other" causes; a marriage may end due to death of one partner, death of both partners, or divorce; an appliance may fail for any of several different reasons (see Example 1.1.10).

The distinguishing feature of the multiple failure modes setting is that each individual has a lifetime T and a mode of failure C, so we require a joint model for T and C. This can be approached by specifying models for $Pr(T \leq t,\ C = j)$ or by specifying mode-specific hazard functions

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{Pr(T < t + \Delta t,\ C = j | T \geq t)}{\Delta t} \tag{1.5.2}$$

for failure modes $j = 1, \ldots, K$. The analysis of multiple modes of failure is of considerable importance, and it turns out to be closely related to the analysis of ordinary univariate lifetime data. Chapter 9 is devoted to this topic.

1.6 SOME COMMENTS ON MODEL SELECTION AND STATISTICAL ANALYSIS

A number of factors enter into the process of modeling and analyzing lifetime distributions. These include the level of detail needed to address specific objectives, background information about the variables and distributions in question, the type of

data available to fit and check models, the availability of software and, more generally, the ease of analysis and interpretation. The planning of experiments and other processes for data collection is also closely linked to modeling issues; observational schemes and matters of design are discussed in Chapter 2.

Two sets of choices are whether to use discrete or continuous-time models, and whether to use parametric or nonparametric assumptions. Most of the standard lifetime data methodology and software is for continuous-time models, and so even when time is discrete (e.g., number of cycles to failure) we often use continuous models. This book deals primarily with continuous models, but discrete distributions are described and used in several chapters.

The choice between parametric and nonparametric specifications is influenced by the amount and type of data available, by background knowledge that may point to specific parametric forms, by assumptions about the regularity or smoothness of the distributions in question, and by the objectives of analysis. Personal taste influences what approach is adopted, but analysis of data usually involves both parametric and nonparametric aspects. This book deals with both parametric and nonparametric methods.

Advantages of parametric models include simplicity, the availability of likelihood-based inference procedures, and ease of use for description, comparison, prediction, or decision. The selection of a specific parametric model is often dictated by its tractability and how well it fits the data. A primary requirement is that the model adequately capture features of the lifetime distribution that are apparent from empirical data. The ability to represent perceived features of the density and hazard functions is often important, as is the behavior of the model in either the left or right tails of the distribution. Convenient representation and comparison of distribution characteristics such as quantiles and survival probabilities is another consideration.

Even when no covariates are present, rather large samples are often needed before the superiority of one model over another in terms of fit is indicated, and severe right censoring limits the comparison of models. This increases an already strong tendency to use models that are mathematically or computationally convenient, and to a large extent this accounts for the extensive use of Weibull, log-logistic, and log-normal models. As the number and complexity of fixed covariates increases, the emphasis on distributional shape is usually much reduced, the primary focus being on location and dispersion aspects of T or $\log T$. Many software packages include methodology for exponential, Weibull, log-normal, log-logistic, and gamma distributions. We discuss how to deal conveniently with them and other models throughout the book.

Nonparametric and semiparametric methods are somewhat less fettered by assumptions than fully parametric methods. They are also useful for checking parametric modeling assumptions. Many software packages provide such methodology.

It is important to bear in mind that models only approximate reality, and that in a given situation several models may provide a good description of observed data. A question of considerable importance is whether alternative models provide consistent conclusions or output from the analysis. Observed data may admit more than one interpretation, but beyond this, we should recognize which conclusions or actions are

sensitive to the choice of model and which are not. This book emphasizes applications in which the objectives are "scientific," for example, to increase understanding of some lifetime process, to estimate important characteristics, or to develop a good model. Lifetime distributions are also used to make decisions in many fields; see, for example, Klugman et al. (1998, Ch. 2) for applications to insurance, and Ascher and Feingold (1984, Ch. 7) and Blischke and Murthy (2000, Chs. 16–18) for applications to system maintenance planning and to product warranty design. The soundness of such decisions naturally depends on the soundness of the models on which they are based.

BIBLIOGRAPHIC NOTES

Many of the origins of lifetime data analysis are in demography and actuarial science, in particular the use of the hazard function or "force of mortality," as it is often called in those disciplines. Over the twentieth century, problems arising in the engineering, life, physical, economic, and social sciences motivated extensive development of models and methodology. Properties of hazard functions and applications to reliability were considered by Barlow, Proschan, and others (Barlow and Proschan 1975). Cox (1972a) stressed the use of the hazard function in problems involving covariates. The use of product integrals to represent lifetime distributions became common from about 1980; Gill and Johansen (1990) survey this topic. Fleming and Harrington (1991) and Andersen et al. (1993) are fundamental references on mathematical aspects of lifetime and event history processes.

Parametric lifetime distributions were used a good deal in the 1930s and 1940s, and started to be very widely studied from about 1950, though some of the models in question had been used much earlier in other contexts. The encyclopedic volumes by Johnson, Kotz, and Balakrishnan (1994, 1995) provide numerous references and properties of many parametric models, as do articles in the *Encyclopedia of Statistical Sciences* (Kotz et al. 1988). We provide only a few key references here.

Davis (1952) described applications of the exponential distribution to reliability; Feigl and Zelen (1965) provided an early application of an exponential model with covariates to medical survival data. The Weibull distribution was studied by Weibull (1951) in connection with the strength of materials; Lieblein and Zelen (1956) and Kao (1959) considered applications in reliability, and Pike (1966) applications in medicine. Gumbel (1958) considered the Weibull and the extreme value distribution in extreme value theory. The log-normal distribution was used by Boag (1949) and Glasser (1965) for cancer survival data and by Nelson and Hahn (1972) for reliability data. Bennett (1983a) described medical applications of the log-logistic distribution. Buckland (1964) and Cox (1962) discussed the gamma distribution in connection with failure times.

The generalized log-Burr model (1.3.21) derives from the work of Burr (1942); Tadikamalla (1980) considers the form here. The generalized gamma distribution was introduced by Stacy (1962), but Prentice (1974) gave the form (1.3.22) considered here, and Farewell and Prentice (1977) discussed applications to reliability.

Chhikara and Folks (1977, 1989) discuss the inverse Gaussian model and its use as a lifetime distribution. Desmond (1985) studies physical models of failure and the relationship between the inverse Gaussian and Birnbaum-Saunders distribution (Birnbaum and Saunders 1969). The piecewise exponential model has been used for a long time in demography; Holford (1976) is an important modern reference. Kooperberg and Stone (1992) and Rosenberg (1995) consider cubic-spline models for density and hazard functions. Titterington et al. (1985) discuss discrete-mixture models. Early examples of applications in reliability were given by Cox (1959) and Kao (1959); Maller and Zhou (1996) discuss cure-rate models of the form (1.3.27). Vaupel et al. (1979), Aalen (1988 1994), and Hougaard (2000) discuss continuous mixtures and the concept of frailty.

Parametric regression models for lifetime data have been widely used since about 1960; early references include Feigl and Zelen (1965), Zippin and Armitage (1966), Pike (1966), Nelson (1970a), Nelson and Hahn (1972), Prentice (1973) and Breslow (1974). Cox (1972a) introduced the semiparametric proportional hazards model and the incorporation of time-varying covariates. Semiparametric location-scale models are closely linked with rank-based methodology; an early reference to problems involving censored data is Prentice (1978).

Multivariate lifetime distributions are considered at length by Hougaard (2000) and also by Joe (1997). Multiple failure modes have a long history in connection with competing risk, or multiple decrement, models in actuarial science or demography. Important modern references include Cox (1959), Altschuler (1970), Nelson (1969, 1972b), and Prentice et al. (1978). Crowder (2001) gives a detailed account. Finally, numerous books discuss the application of lifetime distributions to specific fields. In addition to the preceding references, we mention Klugman et al. (1998) for applications to insurance and actuarial science, Lancaster (1990) for applications to economics, and Blossfeld and Rohwer (1995) for applications to the social sciences.

COMPUTATIONAL NOTES

Software to compute the density function, distribution or survivor function, and quantiles of many of the parametric families in this chapter is widely available, as is the capability to simulate observations from these distributions. Procedures for computing special functions such as gamma, digamma, and trigamma functions are also available. Software that implements statistical methodology in subsequent chapters is also widely available in packages such as SAS, S-Plus, BMDP, SPSS, Systat, and Stata. This book does not give instructions on how to use specific packages, since there are so many choices available. However, specific procedures in S-Plus (Mathsoft, Inc.) and SAS (SAS Institute) will be mentioned in some places, since these packages are very popular and were used in preparing the examples in the book. Brief comments are also provided in the Computational Notes at the ends of chapters. Various surveys of software for lifetime data analysis also exist. For example, Collett (1994) and Harrell and Goldstein (1997) provide general overviews.

PROBLEMS AND SUPPLEMENTS

1.1 *Mean residual lifetime.* Let T be a continuous random variable with survivor function $S(t)$. The mean residual life function $m(t)$ is defined as

$$m(t) = E(T - t|T \geq t).$$

(a) Prove that if $m(t)$ exists, then

$$m(t) = \frac{\int_t^\infty S(x)dx}{S(t)}.$$

The case $t = 0$ gives the well-known result

$$E(T) = \int_0^\infty S(x)\,dx.$$

Also obtain $S(t)$ in terms of $m(t)$, showing that $m(t)$ uniquely defines the distribution of T, through

$$S(t) = \frac{m(0)}{m(t)} \exp\left[-\int_0^t m(u)^{-1}\,du\right].$$

(b) Prove that

$$\lim_{t\to\infty} m(t) = \lim_{t\to\infty} \left(-\frac{d}{dt}\log f(t)\right)^{-1},$$

where $f(t) = -S'(t)$ is the p.d.f. of T. Use this to show that for the log-normal distribution $m(t) \to \infty$ as $t \to \infty$.

(Sections 1.2.1, 1.3.4)

1.2 *Classifying life distributions.* Suppose a continuous lifetime distribution has survivor function $S(t)$, hazard function $h(t)$, cumulative hazard function $H(t)$, and mean residual life function $m(t)$. Consider the following properties that a distribution might have:

I. $h(t)$ is nondecreasing for $t \geq 0$. Distributions with this property are often said to have the increasing failure rate (IFR) property.

II. $H(t)/t$ is nondecreasing for $t > 0$. Distributions with this property are often said to have the increasing failure rate on the average (IFRA) property.

III. $m(t) \leq m(0)$ for all $t \geq 0$. Distributions with this property are often said to have the "new better than used" property.

IV. $m(t)$ is a decreasing function for $t \geq 0$. This is called the decreasing mean residual life property.

(a) Prove that I $\Rightarrow$ II $\Rightarrow$ III.

(b) Prove that I $\Rightarrow$ IV $\Rightarrow$ III.

(It is sometimes useful to classify distributions according to criteria like these, for example, in applications to system reliability.)

(Section 1.2; Bryson and Siddiqui 1969; Barlow and Proschan 1975)

1.3 *Distributions with decreasing failure rates.* A continuous lifetime distribution is said to have the decreasing failure rate (DFR) property if its hazard function $h(t)$ is nonincreasing for $t \geq 0$.

(a) Show that $h'(t) < 0$ only if $f'(t) < 0$, and thus a necessary condition for a distribution to have a DFR is that its p.d.f. have a unique mode at $t = 0$.

(b) Prove that a discrete mixture of distributions that all have DFRs itself has a DFR. Show that a discrete mixture of exponential distributions therefore has a DFR and also that a mixture of IFR distributions does not necessarily have an IFR.

(Sections 1.2, 1.3.10; Proschan 1963)

1.4 *The log-normal distribution.* Consider the log-normal distribution with p.d.f. (1.3.10).

(a) Show that the mean and variance of the distribution are

$$E(T) = e^{\mu+\sigma^2/2}, \qquad \mathrm{Var}(T) = (e^{\sigma^2} - 1)(e^{2\mu+\sigma^2}).$$

(b) Show that the log-normal hazard function $h(t)$ has $h(0) = 0$, increases to a maximum, then decreases, with $h(t) \to 0$ as $t \to \infty$.

(c) Show that the turning point t^* for $h(t)$ satisfies the equation

$$h(t^*) = \frac{1}{\sigma^2 t^*}(\sigma^2 + \log t^* - \mu),$$

and use this to show that

$$e^{\mu-\sigma^2} \leq t^* \leq e^{\mu-\sigma^2+1}.$$

(Section 1.3.3; Watson and Wells 1961; Goldthwaite 1961)

1.5 *The logistic and log-logistic distributions.* Consider the log-logistic distribution (1.3.12), and the corresponding logistic distribution with location parameter u and scale parameter b for $Y = \log T$.

(a) Show that the moment generating function for $W = (Y - u)/b$ is $M(\theta) = E[\exp(\theta W)] = \Gamma(1+\theta)\Gamma(1-\theta)$, and deduce from this that the mean and variance of W are 0 and $\pi^2/3$, respectively. Thus deduce the mean and variance of Y. (Note that $C(\theta) = \log M(\theta)$ is the "cumulant" generating function, that $E(W) = C'(0)$, $\mathrm{Var}(W) = C''(0)$, and see Appendix B.)

(b) Show that the hazard function (1.3.13) is monotone decreasing if $\beta \leq 1$ and that it behaves like the log-normal hazard function if $\beta > 1$. That is, for $\beta > 1$, $h(t)$ has $h(0) = 0$, increases to a maximum, then approaches 0 monotonically as $t \to \infty$.

(c) Find the pth quantile of T and show that when $\beta > 1$ the turning point for $h(t)$ occurs at the $(\beta - 1)/\beta$ quantile.

(d) Show that $E(T^r)$ exists if and only if $\beta > r$, and in that case equals $\alpha^r \Gamma(1 + \beta^{-1})\Gamma(1 - \beta^{-1})$.

(Section 1.3.4)

1.6 *The gamma distribution.* Consider the gamma distribution with p.d.f. $f(t)$ given by (1.3.15).

(a) Show that the hazard function for this distribution is strictly monotone increasing if $k > 1$ and strictly monotone decreasing if $k < 1$. In both cases show that $\lim_{t\to\infty} h(t) = \lambda$.

(b) Show that the mean residual life function $m(t)$ as defined in Problem 1.1 satisfies

$$\lim_{t\to\infty} m(t) = \lambda^{-1}.$$

(c) For the case in which the index parameter k is an integer, prove by repeated integration by parts that

$$\int_t^\infty f(x)\,dx = \sum_{i=0}^{k-1} \frac{e^{-\lambda t}(\lambda t)^i}{i!}.$$

In other words, if T has p.d.f. (1.3.15), then $P(T \geq t) = P(Y_{\lambda t} < k)$, where $Y_{\lambda t}$ has a Poisson distribution with mean λt. Note that this result also follows directly from well-known properties of the Poisson process.

(Section 1.3.5)

1.7 *The generalized Pareto distribution.* Consider the three-parameter distribution with hazard function of the form

$$h(t) = \alpha + \frac{\beta}{t + \gamma}.$$

Examine the range of values that α, β, and γ can take. Investigate $h(t)$ and show that it can be monotone increasing or monotone decreasing, according to the values of the parameters. Give the p.d.f. and survivor function for the distribution.

(Davis and Feldstein 1979)

1.8 *A model capable of bathtub-shaped hazards.* Consider the model that has hazard function

$$h(t) = \frac{\beta}{t + \gamma} + \delta t.$$

Show that the hazard function may be bathtub-shaped, and consider the flexibility of the model in allowing a variety of shapes for the hazard function.

(Hjorth 1980)

1.9 Show for the generalized log-Burr distribution (1.3.20) that the moment generating function $M(\theta) = E[\exp(\theta Z)]$ is

$$M(\theta) = k^{\theta+1}\frac{\Gamma(k-\theta)\Gamma(1+\theta)}{\Gamma(k+1)}.$$

Show that $E(Z) = \log k - \psi(k) + \psi(1)$ and that $\mathrm{Var}(Z) = \psi'(k) + \psi'(1)$, where $\psi(z)$ and $\psi'(z)$ are the digamma and trigamma functions, respectively (Appendix B). Examine the values of $E(Z)$ and $\mathrm{Var}(Z)$ as k ranges from 1 to ∞.

(Section 1.3.6)

1.10 *The log-gamma distribution.* Suppose T has a gamma distribution (1.3.15) with $\lambda = 1$. Show that the moment generating function $M(\theta) = E[\exp(\theta W)]$ for $W = \log T$ is

$$M(\theta) = \frac{\Gamma(k+\theta)}{\Gamma(k)}.$$

Show further that $E(W) = \psi(k)$ and $\mathrm{Var}(W) = \psi'(k)$, where ψ and ψ' are as in Problem 1.9. Derive (1.3.22) as the distribution of $Z = k^{1/2}(W - \log k)$, and show that as $k \to \infty$, it approaches the standard normal p.d.f.

(Bartlett and Kendall 1946; Prentice 1974; Section 1.3.6)

1.11 Let $Y = \log T$ have a logistic distribution with $u = 0$, $b = 1$.

(a) Determine the specific extreme value and normal distributions that have the same mean and variance as Y. Graph and compare the p.d.f.'s of the three distributions. Comment on the similarities and dissimilarities in the models, with a view to discriminating among them.

(b) Compare in a similar way the p.d.f.'s of $T = \exp Y$ in the three cases.

(Section 1.3.6)

1.12 Let $X_1, X_2, \ldots$ be i.i.d. random variables with continuous distribution function $F(x) = P(X_i \le x)$ that satisfies the conditions

1. $F(0) = 0$.

2. For some $\beta > 0$, $\lim_{t\to 0^+}[F(xt)/F(t)] = x^\beta$, with $x > 0$.

The second condition specifies that $F(x) \sim \alpha x^\beta$, where $\alpha > 0$, as $x \to 0^+$.

(a) Let $Y_n = \min(X_1, \ldots, X_n)$. Determine the survivor function of Y_n and hence the survivor function of $Z_n = n^{1/\beta}Y_n$. Show that as $n \to \infty$ the distribution of Z_n converges to a Weibull distribution.

(b) Examine whether or not condition 2 holds when the X_is have (1) a Weibull distribution (2) a gamma distribution, and (3) a uniform distribution on $(0, a)$.

(These results are sometimes put forward as motivation for the Weibull model, as, for example, when an individual is assumed to die at the point at which one of many factors reaches a critical level. The approach here is not totally realistic, since the X_is have been assumed to be i.i.d., but the limiting Weibull form may hold under weaker conditions.)

(Section 1.3.2)

1.13 *A mixed exponential model.* Suppose that a population contains individuals for which lifetimes T are exponentially distributed, but that the hazard function λ varies across individuals. Specifically, suppose that the distribution of T given λ has p.d.f.

$$f(t|\lambda) = \lambda e^{-\lambda t} \qquad t \ge 0,$$

and that λ itself has a gamma distribution with p.d.f.

$$g(\lambda) = \frac{\lambda^{k-1}e^{-\lambda/\alpha}}{\alpha^k\Gamma(k)} \qquad \lambda > 0.$$

(a) Find the unconditional p.d.f. and survivor function for T and show that the unconditional hazard function is

$$h(t) = \frac{k\alpha}{1+\alpha t}.$$

Note that this is a special case of the generalized Pareto model of Problem 1.7. Show that $h(t)$ is monotone decreasing.

(b) Prove that if the distribution of T, given λ, is exponential and λ has a continuous distribution on $(0, \infty)$, then the hazard function for the marginal distribution of T is monotone decreasing.

(c) Prove more generally that if the distribution of T, given λ, has a hazard function $h(t; \lambda)$ that is monotone decreasing for any $\lambda > 0$, and λ has a distribution on $(0, \infty)$, then the hazard function for the marginal distribution of T is monotone decreasing. This generalizes results in Problem 1.3.

(Section 1.3.10; Proschan 1963; Barlow et al. 1963)

1.14 *Burr distributions and properties of mixtures.* The results in the preceding problem can be generalized. Consider frailty models with survivor functions of the form (1.3.32), where Z has a gamma distribution with mean 1 and variance ϕ. That is, the p.d.f. $g(z)$ is given by (1.3.15) with $\lambda = k = \phi^{-1}$.

(a) Show that the survivor (1.3.32) in this case becomes

$$S(t) = \left[1 + \frac{1}{k}H_0(t)\right]^{-k},$$

so that the Burr distribution (1.3.21) arises by taking $H_0(t) = (t/\alpha)^\beta$ of Weibull form.

(b) Consider the case where $H_0(t) = t^2$ is of Weibull form. Obtain the hazard function $h(t) = -S'(t)/S(t)$ and plot it for the values $k = 1, 2, \infty$.

(c) Determine and plot $E(Z|T \geq t)$ as a function of t.

(d) Use the preceding results to comment on the effects of heterogeneity on hazard functions for lifetimes.

(Sections 1.3.6, 1.3.10)

1.15 *Discrete models.*

(a) For the Poisson distribution with probability function

$$Pr(X = j) = e^{-\lambda}\frac{\lambda^j}{j!} \qquad j = 0, 1, \ldots,$$

show that the hazard function is monotone increasing.

(b) For the negative binomial model with probability function

$$Pr(X = j) = \binom{-\alpha}{j} p^{\alpha}(p-1)^j \qquad j = 0, 1, \ldots$$

where $\alpha > 0$ and $0 < p < 1$, show that the hazard function is monotone decreasing (increasing) if $\alpha < 1 (\alpha > 1)$. What happens if $\alpha = 1$?

(Section 1.3.7)

1.16 *Failure rate in multivariate lifetime distributions.* There are various ways in which the hazard function (failure rate) concept can be extended to multivariate distributions. One approach to the idea of increasing hazard functions (Brindley and Thompson 1972) is as follows: suppose that continuous random variables $T_1, \ldots, T_n$ have the joint survivor function

$$S(t_1, \ldots, t_n) = Pr(T_1 \geq t_1, \ldots, T_n \geq t_n) \qquad t_i \geq 0.$$

Suppose that for any subset $\{i_1, \ldots, i_m\}$ of $\{1, \ldots, n\}$ the joint survivor function $S_{i_1 \ldots i_m}(t_{i_1}, \ldots, t_{i_m})$ of $T_{i_1}, \ldots, T_{i_m}$ is such that

$$\frac{S_{i_1 \ldots i_m}(t_{i_1} + x, \ldots, t_{i_m} + x)}{S_{i_1 \ldots i_m}(t_{i_1}, \ldots, t_{i_m})} \tag{1.6.1}$$

is monotone decreasing in $t_{i_1}, \ldots, t_{i_m}$, for any $x > 0$. Then $(T_1, \ldots, T_n)$ is said to have the multivariate increasing failure rate (MIFR) property.

(a) For a univariate distribution with survivor function $S(t)$ the MIFR property states that $S(t+x)/S(t)$ is decreasing in t for all fixed $x > 0$. Show that this is equivalent to the statement that the hazard function $h(t) = -S'(t)/S(t)$ is monotone increasing; that is, the distribution has an IFR.

(b) Prove that $Y = \min(T_1, \ldots, T_n)$ has an IFR if $(T_1, \ldots, T_n)$ has the MIFR property.

(c) The standard bivariate logistic distribution has distribution function

$$F(y_1, y_2) = (1 + e^{-y_1} + e^{-y_2})^{-1} \qquad -\infty < y_1, y_2 < \infty.$$

Obtain the joint survivor function for $T_1 = \exp Y_1$ and $T_2 = \exp Y_2$, and examine (1.7.1) in this case. Does (T_1, T_2) have the MIFR property?

(Section 1.5)

1.17 *Multiple modes of failure.* Consider the definition of mode-specific hazard functions (1.5.2).

(a) Show that the hazard function for T is $\sum_{j=1}^{k} \lambda_j(t)$, and thus obtain the marginal-survivor function $S(t)$ for T.

(b) The mode-specific subdensity functions $f_j(t)$ are defined by

$$f_j(t) = \lim_{\Delta t \to 0} \frac{Pr(t \leq T < t + \Delta t, C = j)}{\Delta t}.$$

Show that $f_j(t) = \lambda_j(t)S(t)$.

(c) Find $F_j(t) = Pr(T \leq t, C = j)$ and thereby also obtain $Pr(C = j)$ and $Pr(C = j|T \leq t)$.

(Section 1.5)

CHAPTER 2

Observation Schemes, Censoring, and Likelihood

2.1 INTRODUCTION

Section 1.1 showed that lifetime data often come with the feature known as right-censoring. As we will see in this chapter, other restrictions on the information available about a set of lifetimes can also occur. A major challenge of lifetime data analysis is to develop methodology that deals with censoring and other conditions. The statistical inference procedures in this book use likelihood functions based on observed data. This chapter establishes the form of the likelihood under censoring and other conditions associated with the selection and observation of individuals in a study, and serves as a basis for the methodology presented in subsequent chapters.

We begin with some preliminary discussion of likelihood; a general summary is given in Appendix C. Suppose that the probability distribution of potentially observable data in a study is specified up to the parameter vector $\boldsymbol{\theta}$. A likelihood function for $\boldsymbol{\theta}$ is, as a function of $\boldsymbol{\theta}$, proportional to the probability of data that were observed. That is,

$$L(\boldsymbol{\theta}) \propto Pr(\text{Data}; \boldsymbol{\theta}), \tag{2.1.1}$$

where Data denotes observed data, and Pr denotes the probability density or mass function from which the data are assumed to arise. A more formal notation for $L(\boldsymbol{\theta})$, which we will use only if necessary, is $L(\boldsymbol{\theta}; \text{Data})$.

Standard likelihood-based methodology applies to models where $\boldsymbol{\theta}$ is a finite dimensional vector, and includes maximum likelihood estimation of $\boldsymbol{\theta}$ and the construction of confidence intervals and tests. The likelihood is, in conjunction with a prior distribution for $\boldsymbol{\theta}$, also the basis for Bayesian analysis. Inferences for nonparametric or semiparametric models can be developed by likelihood theory as well; in this case, the parameter specifying the model is infinite dimensional, and often uncountable. As discussed in Appendix C, more than one likelihood function may

be obtainable from a given data set and model, by employing subsets of the full data or by conditioning on certain outcomes. This is sometimes a way to avoid nuisance parameters, and is associated with the terms marginal and conditional likelihood. In this chapter we focus on "ordinary" likelihood functions, but these terms and the more general concept of partial likelihood (Appendix C) are encountered briefly.

The following example illustrates some of the points mentioned.

Example 2.1.1. Suppose that lifetimes for individuals in some population follow a distribution with probability density function (p.d.f.) $f(t)$ and distribution function $F(t)$, and that the lifetimes $t_1, \ldots, t_n$ for a random sample of n individuals are observed. In the format of (2.1.1), Data $= (t_1, \ldots, t_n)$ and

$$Pr(\text{Data}) = \prod_{i=1}^{n} f(t_i). \tag{2.1.2}$$

If it is assumed that $f(t)$ has a specific parametric form $f(t; \boldsymbol{\theta})$, then the likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i; \boldsymbol{\theta}). \tag{2.1.3}$$

This can be maximized to give an estimate $\hat{\boldsymbol{\theta}}$, and consequently an estimate $F(t; \hat{\boldsymbol{\theta}})$ of the distribution function. A nonparametric approach would be to assume that $F(t)$ is discrete, say with unspecified probabilities $f(t) = F(t) - F(t-1)$ at the jump points $t = 1, 2, 3, \ldots$; this is not very restrictive, since lifetime measurements are in practice discrete. In this case, we consider $\mathbf{f} = (f(1), f(2), \ldots)$ as the model parameter and the right side of (2.1.2) as the likelihood. It is easily seen that (2.1.2) is maximized subject to $f(t) \geq 0$, $\sum_{s=1}^{\infty} f(s) = 1$ by

$$\hat{f}(t) = \frac{1}{n} \sum_{i=1}^{n} I(t_i = t),$$

where $I(A)$ is the indicator function that equals 1 if event A is true and 0 if it is not true. Although $\hat{f}(t)$ may not be a highly appealing estimate because of its roughness, the corresponding estimate $\hat{F}(t) = \hat{f}(1) + \cdots + \hat{f}(t)$, or

$$\hat{F}(t) = \frac{1}{n} \sum_{i=1}^{n} I(t_i \leq t),$$

is an appealing estimate of F. It is known as the empirical distribution function.

Now suppose that $t_1, \ldots, t_n$ are not from an unrestricted random sample of individuals, but rather a random sample of those with lifetimes 1 year or less, with there being no information about the number of individuals with lifetimes greater than 1 year. Truncated samples of this type sometimes arise in reliability and in epidemiology (e.g., Kalbfleisch and Lawless 1989). In this case, the data in (2.1.1) include the nonignorable information that $t_i \leq 1$ for $i = 1, \ldots, n$. The likelihood function is then given by

$$\prod_{i=1}^{n} Pr(t_i | T_i \leq 1) = \prod_{i=1}^{n} \left\{ \frac{f(t_i)}{F(1)} \right\},$$

rather than by (2.1.2).

We now turn to the question of how individuals are selected and observed in studying lifetime distributions. This may be done in a variety of ways, depending on factors such as the (chronological) time needed to observe the events that define lifetime, the feasibility of following individuals over time, and the mechanism for recording lifetimes and covariate values.

Many studies follow individuals longitudinally over time. This is referred to as a prospective study, and examples include life tests, clinical trials, and other types of follow-up studies (see Examples 1.1.1 and 1.1.3–1.1.10). The group or cohort of individuals in such studies is often, but not necessarily, randomly selected from a population of individuals who are at the time origin ($t = 0$) for the lifetime variable T. Limitations on the information collected may be imposed by time, cost, and other constraints. Termination of follow-up before an individual fails causes their lifetime to be right censored. In some settings it may be possible only to determine whether an individual is unfailed or failed at a succession of time points $a_1 < a_2 < \cdots < a_m$; in this case, the lifetime is known only to lie in some interval $[a_{j-1}, a_j)$, a feature known as interval-censoring. The case where the interval is $[0, a_1)$ is known as left censoring. The values of time-varying covariates may likewise be observable only at certain times.

Sometimes individuals cannot be randomly selected and followed from $t = 0$. One possibility is that they are randomly selected from a population of individuals who are alive, and then followed. If u is an individual's t-value at the time of selection, then it is an initial condition on the data that $T > u$, and this must be reflected in the likelihood function. Another possibility is that the observation of data for an individual is at least in part retrospective; this means that at least some of Data used in the likelihood function arises chronologically before the time individuals are selected for the study. In this case, there may be nonignorable conditions that apply to the lifetimes of individuals who are selected.

These observational features are discussed in subsequent sections. Right censoring is the most prevalent complication with lifetime data, and Section 2.2 considers its sources and effects. Sections 2.3 and 2.4 deal with other forms of incomplete data and with nonignorable conditions that arise because of the way individuals are selected for a study. Section 2.5 discusses issues pertaining to the planning of studies. With this material in place, we will be in a position to develop methodology for a wide range of settings and models.

2.2 RIGHT CENSORING AND MAXIMUM LIKELIHOOD

Right censoring, whereby only lower bounds on lifetime are available for some individuals, can occur for various reasons. It may be planned, as when a decision is made to terminate a life test before all items have failed, or unplanned, as when a person in a prospective study is "lost to follow-up" because they move away from the region where the study takes place. To obtain a likelihood function (2.1.1) or the properties of statistical procedures based on censored data it is necessary to consider the process by which both lifetimes and censoring times arise. To do this we apparently need a probability model for the censoring mechanism. Interestingly, it turns out that the observed likelihood function for lifetime parameters takes the same form under a wide variety of mechanisms. We consider some specific types of censoring in the next section and then give a general formulation.

We first introduce some notation for censored data. Suppose that n individuals have lifetimes represented by random variables $T_1, \ldots, T_n$. Instead of the observed values for each lifetime, however, we have a time t_i which we know is either the lifetime or a censoring time. Let us define a variable $\delta_i = I(T_i = t_i)$ that equals 1 if $T_i = t_i$ and 0 if $T_i > t_i$; this is called the censoring or status indicator for t_i, since it tells us if t_i is an observed lifetime ($\delta_i = 1$) or censoring time ($\delta_i = 0$). The observed data then consist of (t_i, δ_i), $i = 1, \ldots, n$. With this notation we occasionally let t_i represent either a random variable or a realized value. This violates the convention where capital letters represent random variables and lowercase letters represent realized values, but no confusion should arise.

The most important result of this section is that for a variety of censoring mechanisms the observed likelihood function takes the form

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i+)^{1-\delta_i}.$$

This is derived below as expression (2.2.3) for the most basic type of censoring, and subsequently for some other censoring mechanisms.

2.2.1 Some Types of Right Censoring

Several censoring mechanisms and the likelihood function obtained for each are described in this section. For simplicity we ignore covariates and assume that lifetimes T_i are independent and identically distributed; extensions to allow covariates are straightforward.

2.2.1.1 Type 1 Censoring

A Type 1 censoring mechanism is said to apply when each individual has a fixed potential censoring time $C_i > 0$ such that T_i is observed if $T_i \leq C_i$; otherwise, we know only that $T_i > C_i$. Type 1 censoring often arises when a study is conducted over a specified time period. In Example 1.1.5, termination of a life test on electrical

insulation specimens after 180 minutes would, for example, mean that $C_i = 180$ for each item. In clinical trials there is often staggered entry of individuals to the study combined with a specified end-of-study date. Example 1.1.7 discussed a clinical trial concerning the duration of remission for patients with leukemia, which was planned to run for one year, with patients entering the trial over that period. The lifetime variable T_i for a patient was the duration of their remission measured from time of entry to the study, and C_i would be the time between their date of entry and the end of study. Example 1.1.6 involved a similar design for a study of equipment reliability.

In our general notation, we have

$$t_i = \min(T_i, C_i), \qquad \delta_i = I(T_i \leq C_i) \tag{2.2.1}$$

for Type 1 censoring. The likelihood function for a Type 1 censored sample is based on the probability distribution of (t_i, δ_i), $i = 1, \ldots, n$. Both t_i and δ_i are random variables in (2.2.1), and their joint p.d.f. is

$$f(t_i)^{\delta_i} Pr(T_i > C_i)^{1-\delta_i}. \tag{2.2.2}$$

To see this, note that the C_i are fixed constants and that t_i can take on values $\leq C_i$, with

$$Pr(t_i = C_i, \delta_i = 0) = Pr(T_i > C_i)$$
$$Pr(t_i, \delta_i = 1) = f(t_i) \qquad t_i \leq C_i,$$

where Pr in the second expression denotes either a p.d.f. or probability mass function according to whether the T_i distribution is continuous or discrete at t_i. Assuming that the lifetimes $T_1, \ldots, T_n$ are statistically independent, we obtain the likelihood function from (2.2.2) as

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i+)^{1-\delta_i}. \tag{2.2.3}$$

The term $S(t_i+)$ appears in (2.2.3) because it equals $Pr(T_i > t_i)$ in general; if $S(t)$ is continuous at t_i, then $S(t_i+) = S(t_i)$.

The adjustment to (2.2.3) when fixed covariates x_i are present in the model is simply to replace $S(t)$ and $f(t)$ with $S_i(t) = Pr(T \geq t|x_i)$ and $f_i(t) = f(t|x_i)$. In the case of external time-varying covariates (Section 1.4), $S_i(t)$ is given by (1.4.6) and $f_i(t)$ by $h_i(t)S_i(t)$, where $h_i(t) = h(t|X_i)$.

Exact sampling properties of estimates or tests based on a likelihood function of the form (2.2.3) are generally intractable mathematically, but standard large sample results for maximum likelihood (described in Appendix C) apply, and finite sample properties can be investigated by simulation. Asymptotic theory and statistical inference from likelihoods based on censored data are discussed in Section 2.2.3 and, for specific models, in later chapters throughout the book.

Example 2.2.1. Suppose that lifetimes T_i are independent and follow an exponential distribution with p.d.f. $f(t) = \lambda \exp(-\lambda t)$ and survivor function $S(t) = \exp(-\lambda t)$. Then (2.2.3) gives

$$L(\lambda) = \prod_{i=1}^{n} (\lambda e^{-\lambda t})^{\delta_i} (e^{-\lambda t})^{1-\delta_i}$$

$$= \lambda^r \exp\left(-\lambda \sum_{i=1}^{n} t_i\right), \tag{2.2.4}$$

where $r = \sum \delta_i$ is the observed number of uncensored lifetimes, or failures. The log-likelihood function $\ell(\lambda) = \log L(\lambda)$ is

$$\ell(\lambda) = r \log \lambda - \lambda \sum_{i=1}^{n} t_i. \tag{2.2.5}$$

The maximum likelihood estimate is given by solving $d\ell/d\lambda = 0$, and is $\hat{\lambda} = r/\sum_{i=1}^{n} t_i$. The exact distribution of $\hat{\lambda}$ is rather intractable, as is the distribution of the minimal sufficient statistic $(r, \sum t_i)$.

For the Type 1 censoring scheme the censoring times C_i are specified fixed values. In many settings they are actually random. For example, in the clinical trial described in Example 1.1.7 and discussed earlier, individuals entered the study in a more or less random fashion according to their time of diagnosis with leukemia, so their censoring times were effectively random. In fact, the study was actually terminated early, based on the accumulating data, thus altering the original censoring times. We consider a simple model for random censoring next, and a more general model in Section 2.2.2.

2.2.1.2 *Independent Random Censoring*

A very simple random censoring process that is often realistic is one in which each individual is assumed to have a lifetime T and a censoring time C, with T and C independent continuous random variables, with survivor functions $S(t)$ and $G(t)$, respectively. All lifetimes and censoring times are assumed mutually independent, and it is assumed that $G(t)$ does not depend on any of the parameters of $S(t)$. As in the case of Type 1 censoring, $t_i = \min(T_i, C_i)$ and $\delta_i = 1$ if $T_i \leq C_i$ and $\delta_i = 0$ if $T_i > C_i$. The data from observations on n individuals is assumed to consist of the pairs (t_i, δ_i), $i = 1, \ldots, n$; the same final result is obtained if C_i is available for all $i = 1, \ldots, n$. The p.d.f. of (t_i, δ_i) is easily obtained: if $f(t)$ and $g(t)$ are the p.d.f.'s for T_i and C_i, then

$$\begin{aligned} Pr(t_i = t, \delta_i = 0) &= Pr(C_i = t, T_i > C_i) \\ &= g(t)S(t) \\ Pr(t_i = t, \delta_i = 1) &= Pr(T_i = t, T_i \leq C_i) \\ &= f(t)G(t). \end{aligned}$$

These can be combined into the single expression

$$Pr(t_i = t, \delta_i) = [f(t)G(t)]^{\delta_i}[g(t)S(t)]^{1-\delta_i},$$

and thus the distribution of (t_i, δ_i), $i = 1, \ldots, n$, is

$$\prod_{i=1}^{n} [f(t_i)G(t_i)]^{\delta_i}[g(t_i)S(t_i)]^{1-\delta_i}.$$

Since $G(t)$ and $g(t)$ do not involve any of the parameters in $f(t)$, they can be neglected and the likelihood function taken to be

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i},$$

which is of the same form as (2.2.3). The earlier result for Type 1 censoring can in fact be considered as a special case of this if we allow the C_i to have degenerate distributions, each with mass at one fixed point. Another approach that leads directly to this likelihood function is to argue that if $G(t)$ and $g(t)$ do not involve any parameters of $f(t)$, then $C_1, \ldots, C_n$ are ancillary and one should condition on the realized censoring times when making inferences about the distribution of T. This takes us back to the Type 1 censoring framework. A point to note is that although it may be desirable to make inferences conditional on the C_i in any given situation, the properties of procedures averaged over the distribution of the C_i may be of interest when planning studies, and in some applications.

Although the independent random-censorship model is often reasonable, in many situations the censoring process is linked to the failure time process. Suppose, for example, that the termination date for a medical trial is not fixed before the study commences, but is chosen later, with the choice influenced by the results of the study up to that time. In such instances it may be difficult to write down a model that fully represents the process under study. Fortunately, the likelihood function (2.2.3) is still applicable in many such complicated situations. This is discussed in Section 2.2.2.

2.2.1.3 *Type 2 Censoring*

The term Type 2 censoring refers to the situation where only the r smallest lifetimes $t_{(1)} \leq \cdots \leq t_{(r)}$ in a random sample of n are observed; here r is a specified integer between 1 and n. This censoring scheme arises when n individuals start on study at the same time, with the study terminating once r failures (or lifetimes) have been observed. Although some life tests are formulated with Type 2 censoring, they have the practical disadvantage that the total time $t_{(r)}$ that the test will run is random and hence unknown at the start of the test. Type 1 censoring is therefore much more common in planned experiments. The exact sampling properties of statistical procedures based on a Type 2 censored sample are, however, tractable in many cases and this censoring scheme is often discussed in theoretical work.

With Type 2 censoring the value of r is chosen before the data are collected, and the data consist of the r smallest lifetimes in a random sample $T_1, \ldots, T_n$. For continuous distributions we can ignore the possibility of ties and denote the r smallest lifetimes as $T_{(1)} < T_{(2)} < \cdots < T_{(r)}$. If the T_i have p.d.f. $f(t)$ and survivor function $S(t)$, then from general results on order statistics (Appendix B.3) the joint p.d.f. of $T_{(1)}, \ldots, T_{(r)}$ is

$$\frac{n!}{(n-r)!}\left\{\prod_{i=1}^{r} f(t_{(i)})\right\} S(t_{(r)})^{n-r}. \tag{2.2.6}$$

The likelihood function is based on (2.2.6). By dropping the constant $n!/(n-r)!$ and noting that in terms of the (δ_i, t_i) notation we have $\delta_i = 0$ and $t_i = t_{(r)}$ for those individuals whose lifetimes are censored, we see that (2.2.6) gives a likelihood of the same form (2.2.3) as for Type 1 censoring. The sampling properties are, however, different in finite samples.

Example 2.2.2. Consider the exponential distribution as in Example 2.2.1, but suppose lifetimes are Type 2 censored. The log-likelihood is still of the form (2.2.5), but here it can be written as

$$\ell(\lambda) = r \log \lambda - \lambda\left[\sum_{i=1}^{r} t_{(i)} + (n-r)t_{(r)}\right]$$

and the maximum likelihood estimate for λ can be written as $\hat{\lambda} = r/W$, where

$$W = \sum_{i=1}^{r} t_{(i)} + (n-r)t_{(r)}.$$

Since r is fixed, the statistic W is sufficient for λ, and it is readily shown (see Section 4.1.2) that with the data considered as random variables, $2\lambda W = 2r\lambda/\hat{\lambda} \sim \chi^2_{(2r)}$, a chi-squared distribution with $2r$ degrees of freedom. This allows exact confidence intervals and tests for λ to be developed.

Progressive Type 2 Censoring

Progressive Type 2 censoring is a generalization of Type 2 censoring. In this case, the first r_1 failures in a life test of n items are observed; then n_1 of the remaining $n - r_1$ unfailed items are removed from the experiment, leaving $n - r_1 - n_1$ items still present. When a further r_2 items have failed, n_2 of the still unfailed items are removed, and so on. The experiment terminates after some prearranged series of repetitions of this procedure.

This scheme is of more theoretical than practical interest, but let us obtain the likelihood function, assuming that lifetimes are independent and indentically distributed (i.i.d.) with p.d.f. $f(t)$ and survivor function $S(t)$. For simplicity we suppose the censoring has only two stages: at the time of the r_1th failure, n_1 of the remaining $n - r_1$

unfailed items are randomly selected and removed. The experiment then terminates when a further r_2 items have failed. At this point there will be $n - r_1 - n_1 - r_2$ items still unfailed. The observations in this case are the r_1 failure times $T_{(1)} < \cdots < T_{(r_1)}$ in the first stage of the experiment and the r_2 failure times in the second stage of the experiment, which we will denote by $T^*_{(1)} < \cdots < T^*_{(r_2)}$. The experiment is represented in Figure 2.1.

The distribution of the data can be written as

$$g_1(t_{(1)}, \ldots, t_{(r_1)}) g_2(t^*_{(1)}, \ldots, t^*_{(r_2)} | t_{(1)}, \ldots, t_{(r_1)}), \tag{2.2.7}$$

where g_1 and g_2 represent p.d.f.'s of the variables indicated. The joint p.d.f. $g_1(t_{(1)}, \ldots, t_{(r_1)})$ of $T_{(1)}, \ldots, T_{(r_1)}$ is given by (2.2.6), with $r = r_1$. To write down the second term in (2.2.7) we observe that given $t_{(1)}, \ldots, t_{(r_1)}$, the lifetimes of the items left in the experiment have a left-truncated distribution with p.d.f. and survivor functions

$$f_1(t) = \frac{f(t)}{S\left(t_{(r_1)}\right)}, \qquad S_1(t) = \frac{S(t)}{S\left(t_{(r_1)}\right)} \qquad t \geq t_{(r_1)},$$

respectively. Thus $T^*_{(1)}, \ldots, T^*_{(r_2)}$ are the r_2 smallest observations in a random sample of size $n - n_1 - r_1$ from this truncated distribution. By (2.2.6), the second term in (2.2.7) is therefore

$$\frac{(n-r_1-n_1)!}{(n-r_1-n_1-r_2)!} f_1(t^*_{(1)}) \cdots f_1(t^*_{(r_2)}) \left[S_1\left(t^*_{(r_2)}\right)\right]^{n-r_1-n_1-r_2}.$$

Combining the two parts of (2.2.7), we obtain the likelihood function as

$$c f(t_{(1)}) \cdots f(t_{(r_1)}) \left[S(t_{(r_1)})\right]^{n_1} f(t^*_{(1)}) \cdots f(t^*_{(r_2)}) \left[S(t^*_{(r_2)})\right]^{n-r_1-n_1-r_2}, \tag{2.2.8}$$

where $c = n!(n-r_1-n_1)!/[(n-r_1)!(n-r_1-n_1-r_2)!]$. Once again, using the (t_i, δ_i) notation, we find that (2.2.8) is of the same form as (2.2.3).

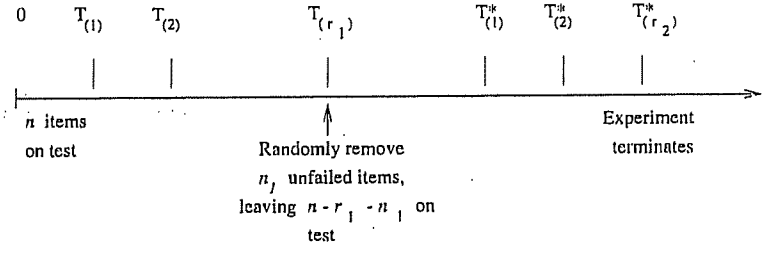

Figure 2.1. Progressive Type 2 censoring.

2.2.2 A General Formulation of Right Censoring

The censoring process is often not any of the types discussed so far, and may be sufficiently complicated to make modeling it impossible. For example, a decision to terminate a life test or clinical trial at time t, or to withdraw certain individuals, might be based on failure information prior to time t. Fortunately it can be shown that under rather general conditions the observed likelihood is of the form (2.2.3) and can be used in the normal way to make inferences about the lifetime distribution under study.

The key idea for a general approach is to consider the failure and censoring processes for a group of individuals as time goes by. We develop results for discrete-time models first; the general case is then obtained by limiting arguments similar to those used in Section 1.2.3 for the general formulation of lifetime distributions.

Suppose that n individuals are followed from $t = 0$ until each fails or is censored. Assume that lifetimes and censoring times are discrete; for convenience and with no loss of generality we assume allowable values for each are $t = 0, 1, 2, \ldots$. Suppose for now that covariates do not vary over time and let $h_i(t)$ and $S_i(t)$ be the hazard and survivor functions (see (1.2.7) and (1.2.8)) for individual i, conditional on observed covariate values.

We introduce some additional notation directed at the evolution of the failure and censoring processes over time, as follows. For $t = 0, 1, 2, \ldots$ let

$$Y_i(t) = I(T_i \geq t, \text{ Individual } i \text{ is not censored before } t)$$

$$dN_i(t) = Y_i(t)I(T_i = t)$$

$$dC_i(t) = Y_i(t)I(\text{Individual } i \text{ is censored at } t)$$

The variable $Y_i(t)$ is often called the at risk indicator; it equals 1 if and only if individual i is alive and uncensored just before time t, and hence at risk of being observed to fail at t. The variables $dN_i(t)$ and $dC_i(t)$ record observed failure and censoring events at time t, respectively. Among all the values $\{dN_i(t), dC_i(t), t \geq 0\}$, only one is nonzero for any individual.

We also define vectors $d\mathbf{N}(t) = (dN_1(t), \ldots, dN_n(t))$, $d\mathbf{C}(t) = (dC_1(t), \ldots, dC_n(t))$, and

$$\mathcal{H}(t) = \{(d\mathbf{N}(s), d\mathbf{C}(s)), \qquad s = 0, 1, \ldots, t-1\}.$$

We refer to $\mathcal{H}(t)$ as the history of the failure and censoring processes at time t. It consists of the information about all failures and censoring events that occurred up to time $t-1$. The important point is that the data that we observe (aside from the covariate values) can be represented as

$$\text{Data} = (d\mathbf{N}(t), d\mathbf{C}(t);\ t = 0, 1, 2, \ldots).$$

Furthermore, we can decompose $Pr(\text{Data})$ as

$$Pr(\text{Data}) = \prod_{t=0}^{\infty} Pr(d\mathbf{N}(t)|\mathcal{H}(t))\, Pr(d\mathbf{C}(t)|d\mathbf{N}(t), \mathcal{H}(t)), \tag{2.2.9}$$

where $\mathcal{H}(0)$ is null. In (2.2.9) all probabilities are conditional on covariate values, but for simplicity this is suppressed in the notation.

So far we have made no assumptions about the censoring mechanism, but to proceed further it is necessary to do this. Assumptions that have become standard in lifetime data analysis require that

$$Pr(d\mathbf{N}(t)|\mathcal{H}(t)) = \prod_{i=1}^{n} h_i(t)^{dN_i(t)}[1 - h_i(t)]^{Y_i(t)(1-dN_i(t))}. \tag{2.2.10}$$

Effectively, this requires that given $\mathcal{H}(t)$ and covariate values, the failure mechanisms for individuals at risk at time t operate independently, and that for $t = 0, 1, 2, \ldots$

$$Pr(dN_i(t) = 1|\mathcal{H}(t)) = Y_i(t)h_i(t). \tag{2.2.11}$$

The notational convention $0^0 = 1$ is used in (2.2.10), corresponding to the fact that if $Y_i(t) = 0$ there is no information about individual i at time t, and the term in the likelihood should equal one. Note that the value of $Y_i(t)$ is determined by the information in $\mathcal{H}(t)$.

The condition (2.2.11) represents a conditional (on $\mathcal{H}(t)$ and covariate values) independence between failure and censoring at time t, and mechanisms that satisfy it are often termed independent censoring mechanisms. Under (2.2.11), the probability that an individual who is alive and uncensored just prior to time t is observed to fail at t is $h_i(t)$, the same as if there were no censoring.

If the terms $Pr(d\mathbf{C}(t)|d\mathbf{N}(t), \mathcal{H}(t))$ in (2.2.9) do not involve any of the parameters that specify the $h_i(t)$, the censoring scheme is called noninformative. These terms can then be dropped from the likelihood, and by inserting (2.2.10) into (2.2.9), we get

$$L = \prod_{i=1}^{n}\prod_{t=0}^{\infty} h_i(t)^{dN_i(t)}[1 - h_i(t)]^{Y_i(t)(1-dN_i(t))}. \tag{2.2.12}$$

Each individual is observed either to fail or be censored at some time t. In the case of failure at t, $dN_i(t) = 1$ and $Y_i(s) = I(s \leq t)$; in the case of censoring at t, $dN_i(t) = 0$ and $Y_i(s) = I(s \leq t)$. Since (see (1.2.6) and (1.2.8))

$$S_i(t) = \prod_{s=0}^{t-1}(1 - h_i(s)), \qquad f_i(t) = h_i(t)S_i(t),$$

we find that (2.2.12) gives, in the (t_i, δ_i) notation,

$$L = \prod_{i=1}^{n} f_i(t_i)^{\delta_i} S_i(t_i + 1)^{1-\delta_i}. \tag{2.2.13}$$

Since $S_i(t+1) = S_i(t+)$, the likelihood is exactly of the form (2.2.3) encountered previously for Type 1 and other forms of censoring.

To obtain the likelihood in the case of continuous or mixed distributions we use the same ideas as in Section 1.2.3. We associate $dN_i(t)$ and $dC_i(t)$ with a short interval $[t, t + dt)$ in a partition of the time axis, and in (2.2.10) and (2.2.11) we replace $h_i(t)$ with $dH_i(t)$, where $H_i(t)$ is the cumulative hazard function (1.2.14). The preceding arguments go through essentially unchanged as we take the product limit of (2.2.12) and then use (1.2.16) to obtain the product integrals in

$$L = \prod_{i=1}^{n} \prod_{(0,\infty)} dH_i(t)^{dN_i(t)}[1 - dH_i(t)]^{Y_i(t)(1-dN_i(t))}$$
$$= \prod_{i=1}^{n} f_i(t_i)^{\delta_i} S_i(t_i+)^{1-\delta_i}, \tag{2.2.14}$$

exactly as in (2.2.3).

2.2.2.1 *Discussion*

The independent censoring condition (2.2.11) requires that censoring in $[t, t+dt)$ not depend on $dN(t)$. In the discrete-time case we usually assume that a censoring event at time t means that the physical censorship is just after time t, so that if $Y_i(t) = 1$, an individual who dies at t is observed to do so, as (2.2.11) indicates. Consistent with this, for an individual censored at t it is assumed that $T_i > t$, as in (2.2.13). Censoring is typically noninformative in this case as well. More generally, (2.2.11) means that censoring at time t cannot be related to failure information at or after time t, so it cannot selectively discriminate among individuals according to when they will fail in the future. This seems an obvious requirement for valid estimation of the lifetime distribution in the presence of censoring, but is one that is uncheckable solely on the basis of the data (t_i, δ_i), $i = 1, \ldots, n$. It may also appear hard for it to be violated by a real censoring process, but that is not so. For example, if a covariate x that affects lifetime also affects the censoring process, then failure to include x in the model for T can cause a violation of (2.2.11). Another setting where (2.2.11) could be violated is in the discrete case where $t = 0, 1, 2, \ldots$ refers to equally spaced points in continuous time, and where an individual alive and uncensored at time t may be lost to follow-up between t and $t + 1$. In studies where the observation times are far apart, the event that an individual is lost to follow-up (and therefore censored, so that $Y(t + 1) = 0$) may not be independent of whether they fail in $(t, t + 1]$.

There are two additional important features of the preceding development. One is that the censoring mechanism at time t is allowed to depend on the history of censoring and failure before t. The Type 2 censoring process is actually of this type. More generally, it would be permissible in a study to make a decision about censoring individuals (i.e., removing them from the study) or terminating the study at time t according to failure information up to that time. A second point is that the likelihood (2.2.14) is available without a specific model of the censoring process. As long as the terms $Pr(d\mathbf{C}(t)|d\mathbf{N}(t), \mathcal{H}(t))$ in (2.2.9) do not involve parameters of interest, they drop out of the likelihood. With independent censoring this is generally the case, but even if these terms do contain information about the $h_i(t)$, it can be shown that

(2.2.14) is a partial likelihood (see Appendix C) and can still be used for inference in the usual way.

The observed likelihood (2.2.14) thus has the same form for a variety of censoring schemes. Moreover, inference procedures based on maximum likelihood large-sample theory can be applied in a straightforward way, as we will discuss in Section 2.2.3. However, the probability distributions upon which (2.2.14) is based can differ substantially according to the censoring process, and small sample properties of estimates or tests may therefore be different. In general, (2.2.14) is based on the joint distribution for the censoring and failure processes, and it is only in special cases such as Type 1 or independent random censoring that the censoring times can be viewed as fixed values.

If there are external time-varying covariates $\mathbf{x}(t)$, then the preceding development can be extended. We now assume that models for the hazard function of T given $X = \{\mathbf{x}(t), t \geq 0\}$ are as in Section 1.4, say of the form (1.4.4). The preceding argument goes through essentially unchanged, adding the $X_i(t)$ to the history $\mathcal{H}(t)$ in (2.2.9), (2.2.10), and (2.2.11). The likelihood function is then of the form (2.2.14), with $S_i(t)$ given by (1.4.6) and $f_i(t)$ by $h_i(t)S_i(t)$, where $h_i(t) = h(t|X_i) = h(t|\mathbf{w}_i(t))$.

2.2.3 Likelihood Inference with Censored Data

Statistical inference for parametric models can in standard settings be based on well-known maximum likelihood methodology, described in general terms in Appendix C. Let $\boldsymbol{\theta}$ be a $p \times 1$ parameter vector, and let $L(\boldsymbol{\theta})$ represent the likelihood and $\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$ the log-likelihood function. The $p \times 1$ vector $\mathbf{U}(\boldsymbol{\theta}) = \partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is usually called the score vector and the $p \times p$ matrix

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{-\partial^2\ell}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \tag{2.2.15}$$

is called the information matrix. The maximum likelihood estimate (m.l.e.) $\hat{\boldsymbol{\theta}}$ maximizes $L(\boldsymbol{\theta})$ and $\ell(\boldsymbol{\theta})$, and usually satisfies the score equation $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$. The Fisher or expected information matrix is defined as

$$\mathcal{I}(\boldsymbol{\theta}) = E\{\mathbf{I}(\boldsymbol{\theta})\}, \tag{2.2.16}$$

where the expectation in (2.2.16) is with respect to the random Data, which specifies the likelihood (see (2.1.1)), and is calculated under the probability distribution that generates the observed data. Well-known large sample or asymptotic results that are used for inference throughout this book include the fact that, considered as random variables, $\hat{\boldsymbol{\theta}}$ is approximately normally distributed in large-samples and the likelihood ratio statistic $\Lambda(\boldsymbol{\theta}) = 2\ell(\hat{\boldsymbol{\theta}}) - 2\ell(\boldsymbol{\theta})$ is approximately χ^2. Details of these and other results are given in Appendix C.

Standard large-sample procedures for maximum likelihood can be shown to apply to all of the settings described in Sections 2.2.1 and 2.2.2. With Type 1 censoring, asymptotic results of the usual type hold under essentially the same conditions as

for the case of complete (i.e., uncensored) random samples. An added requirement is that the sequence of fixed censoring times $C_1, \ldots, C_n$ satisfy conditions so that as $n \to \infty$, the expected information $\mathcal{I}(\boldsymbol{\theta})$ increases at rate n; a sufficient condition in most instances is that the expected number of observed (i.e., uncensored) lifetimes approach infinity at rate n as $n \to \infty$.

Independent random censoring (Section 2.2.1.2) is subject to essentially the same requirements as Type 1 censoring, the only distinction being that the C_i are treated as random variables rather than fixed constants. Type 2 censoring is also straightforward to deal with; the usual assumption for the development of asymptotic results is that as $n \to \infty$, we have $r \to \infty$ with r/n approaching a limiting constant p. In a few cases involving Type 2 censoring, exact distributional results for m.l.e.'s or likelihood-based procedures can be obtained.

Asymptotic results can be derived in an elegant fashion for the general censoring framework of Section 2.2.2 by the use of counting processes and martingale theory (Appendix F). This approach obviates the need for special treatments of Type 1 or Type 2 censoring. Detailed mathematical treatments are available, and we merely outline the main ideas.

Consider a continuous lifetime distribution with parametrically specified hazard functions $h_i(t; \boldsymbol{\theta})$ for $i = \ , \ldots, n$. In this case, $dH_i(t; \boldsymbol{\theta}) = h_i(t; \boldsymbol{\theta})\, dt$ and the product integral expression for the likelihood (2.2.14) is proportional to

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{(0,\infty)} h_i(t; \boldsymbol{\theta})^{dN_i(t)} [1 - h_i(t; \boldsymbol{\theta})\, dt]^{Y_i(t)(1-dN_i(t))}.$$

By (1.2.12) this equals

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ \prod_{(0,\infty)} h_i(t; \boldsymbol{\theta})^{dN_i(t)} \right\} \exp\left\{ -\int_0^\infty Y_i(t) h_i(t; \boldsymbol{\theta})\, dt \right\}.$$

Defining the counting process $N_i(t) = \int_0^t dN_i(u)$, we can then write the log-likelihood as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^\infty \log h_i(t; \boldsymbol{\theta})\, dN_i(t) - \int_0^\infty Y_i(t) h_i(t; \boldsymbol{\theta})\, dt \tag{2.2.17}$$

and the score function as

$$\begin{aligned} \mathbf{U}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \int_0^\infty \frac{\partial h_i(t; \boldsymbol{\theta})/\partial \boldsymbol{\theta}}{h_i(t; \boldsymbol{\theta})}\, dN_i(t) - \int_0^\infty Y_i(t) \frac{\partial h_i(t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\, dt \\ &= \sum_{i=1}^{n} \int_0^\infty \frac{\partial \log h_i(t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} [dN_i(t) - Y_i(t) h_i(t; \boldsymbol{\theta})\, dt], \end{aligned} \tag{2.2.18}$$

assuming that we can differentiate through the integral sign. It is noted that because

$$E\{dN_i(t) - Y_i(t) h_i(t; \boldsymbol{\theta})\, dt | \mathcal{H}(t)\} = 0,$$

the score function satisfies $E_{\boldsymbol{\theta}}\{\mathbf{U}(\boldsymbol{\theta})\} = \mathbf{0}$, and is therefore an unbiased estimating function (Appendix C.2). Moreover, the processes

$$M_i(t) = \int_0^t [dN_i(u) - Y_i(u) h_i(u; \boldsymbol{\theta})\, du]$$

are martingales (see Appendix F.2) and by applying standard results we are able to show that the log-likelihood (2.2.17) and score function (2.2.18) give the same asymptotic results as standard settings.

Maximum likelihood large-sample methods will be used throughout the book. It should be noted that although some approaches utilize the expected information (2.2.16), it is in many settings impossible to calculate this because there is not a tractable or sufficiently detailed model of the censoring process. It is appropriate and customary in most applications to use the ordinary information matrix (2.2.15) or the observed information matrix $\mathbf{I}(\hat{\boldsymbol{\theta}})$ in large-sample methodology.

Example 2.2.3. Consider once again the exponential distribution of Example 2.2.1. The observed log-likelihood function is given by (2.2.5) under all of the censoring processes satisfying the conditions of Section 2.2.2:

$$\ell(\lambda) = r \log \lambda - \lambda \sum_{i=1}^{n} t_i,$$

where $r = \sum \delta_i$ is the number of observed (uncensored) lifetimes. The 1×1 information matrix is

$$I(\lambda) = \frac{-d^2\ell}{d\lambda^2} = \frac{r}{\lambda^2},$$

and the expected information matrix is $\mathcal{I}(\lambda) = E(r)/\lambda^2$. For Type 1 or Type 2 censoring we can evaluate $E(r)$, but for complicated censoring processes in which decisions to end follow-up are based on previous lifetimes, or when individuals are lost to follow-up by an unknown process, it is not feasible to determine $E(r)$. In this case, we would use $I(\lambda)$; this turns out here to be the same as if we estimated $E(r)$ in $\mathcal{I}(\lambda)$ by r.

2.3 OTHER TYPES OF INCOMPLETE DATA

2.3.1 Intermittent Observation and Interval Censoring

Because lifetime data occur over chronological time, a variety of schemes are used to obtain data according to prevailing time and resource constraints. This can produce other forms of incompleteness besides right censoring. A common occurrence is for

individuals in a study to be observed intermittently, at discrete-time points. We begin by considering a framework where each individual $i = 1, \ldots, n$ is observed at a prespecified set of times $0 = a_{i0} < a_{i1} < \cdots < a_{im_i} < \infty$. If an individual has not failed by time $a_{i,j-1}$ $(j = 1, \ldots, m_i)$, they are observed next at a_{ij}, and it is determined whether or not failure occurred in the interval $(a_{i,j-1}, a_{ij}]$. The observed data then consist of an interval $(U_i, V_i]$ for each individual, with the information that $U_i < T_i \leq V_i$, and the lifetime is said to be interval censored. If failure has not occurred by time a_{im_i}, then $V_i = \infty$ and $U_i = a_{im_i}$ is a right-censoring time for T_i.

The observed likelihood function from a sample of N independent individuals under this observation scheme is

$$L = \prod_{i=1}^{n} [F_i(V_i) - F_i(U_i)], \tag{2.3.1}$$

where $F_i(t)$ is the distribution function for T_i and we assume that $F_i(0) = 0$. An easy way to obtain (2.3.1) is to notice that the observation for individual i is in effect multinomial $(1; p_{i1}, \ldots, p_{im_i})$, where $p_{ij} = F_i(a_{ij}) - F_i(a_{i,j-1})$. For parametric models, inference based on the likelihood (2.3.1) falls under the standard theory of Appendix C. Nonparametric estimation is more complicated; some special problems are discussed in Chapters 3 and 7. The case where observation times are the same for all individuals (i.e., $a_{ij} = a_j$) is often referred to as grouped data.

The interval-censoring framework just described covers many situations; a few examples follow.

Example 2.3.1. Sometimes the failure of a piece of material or equipment can be determined only by inspection. For example, a lifetime often associated with metal components such as airplane bodies, pressure tubes in nuclear reactors, or railway track is the time until a defined type of flaw (e.g., a crack) appears. Components are usually examined periodically, so the exact time of appearance is interval censored.

Example 2.3.2. In many longitudinal studies on humans it is feasible to see individuals only at rather widely spaced intervals; in longitudinal surveys individuals may be seen only every one or two years. The timing of some types of events can be determined retrospectively, but some cannot. For example, the determination that a child has reached puberty may rely on tests carried out at the observation times, so that the age of onset of puberty is interval censored.

Example 2.3.3. Current Status Data. This term refers to interval censored lifetimes where the interval for an individual is either $(0, C_i]$ or (C_i, ∞). Such data arise when individual i is examined once, at time C_i, at which point it is determined whether failure has already occurred (i.e., $T_i \leq C_i$) or not (i.e., $T_i > C_i$). In shelf-life problems involving food or drugs, an item may degrade over time, with failure being defined in terms of the amount of degradation. To determine whether an item has failed it may be necessary to destroy the item, for example, to open a sealed container or to carry out chemical analysis of the item. Current status data also arise in animal carcinogenicity studies in which the time to occurrence of a tumor is of interest, but where tumors can be detected only by autopsy when the animal is dead. In demography, studies on female fertility in underdeveloped countries often use current-status data on items such as the age at which a woman becomes fertile or first gives birth, because accurate information about timing of events is hard to determine retrospectively.

The assumption that observation times a_{ij} are fixed ahead of time, or even that they are determined independently of the process that generates lifetimes, is unsupportable in many settings. For example, a decision regarding when to see an individual next in a clinical study may be based on current information about the individual. A more general process analogous to the censoring process in Section 2.2.2 can be considered. Suppose that if an individual is alive and uncensored at time $a_{i,j-1}$, a decision about the next observation time a_{ij} is based on observed failure, covariate, and observation time history $\mathcal{H}(a_{i,j-1})$ up to $a_{i,j-1}$. The choice of a_{ij} is, however, conditionally independent of failure and covariate information beyond $a_{i,j-1}$, given $\mathcal{H}(a_{i,j-1})$. In this case

$$Pr(a_{i,j-1} < T_i \leq a_{ij} | \mathcal{H}(a_{i,j-1}), a_{ij}) = \frac{F_i(a_{ij}) - F_i(a_{i,j-1})}{1 - F_i(a_{i,j-1})}, \tag{2.3.2}$$

where $\mathcal{H}(a_{i,j-1})$ is understood to include the information that individual i is alive and uncensored at $a_{i,j-1}$. Under this observation process the data consist of the observation times $0 = a_{i0} < \cdots < a_{i,m_i} \leq \infty$, and the information that $a_{i,m_i-1} < T_i \leq a_{i,m_i}$. Note that $a_{i,m_i} = \infty$ corresponds to right-censoring of the lifetime at time a_{i,m_i-1}. Assuming that the terms $Pr(a_{ij} | \mathcal{H}(a_{i,j-1}))$, $j = 1, \ldots, m_i$, do not contain information about the lifetime distribution, the observed likelihood function for individual i is proportional to

$$\left\{ \prod_{j=1}^{m_i-1} Pr(T_i > a_{ij} | \mathcal{H}(a_{i,j-1}), a_{ij}) \right\} Pr(T_i \leq a_{i,m_i} | \mathcal{H}(a_{i,m_i-1}), a_{i,m_i}).$$

Because of (2.3.2), this reduces to the earlier likelihood (2.3.1), where $U_i = a_{i,m_i-1}$ and $V_i = a_{i,m_i}$.

The probabilities in (2.3.1) and (2.3.2) are conditional on observed covariate values, but for convenience this is suppressed in the notation. If covariates are time varying, then values are needed over $(a_{i,j-1}, a_{ij}]$; it is often necessary to estimate or impute values.

The condition (2.3.2) requires that the process determining a_{ij} is unrelated to failure information beyond $a_{i,j-1}$. This can be violated if covariates related to both the observation time and lifetime processes are not included in the lifetime model. Another potential problem is where an individual last seen at $a_{i,j-1}$ is scheduled

for observation at some future time, but is lost to follow-up; this often happens in longitudinal surveys and in some clinical studies. The result is that $a_{ij} = \infty$, and the concern is whether loss to follow-up between the two scheduled observation times is conditionally independent of T_i. This can only be assessed by tracking some of the individuals lost to follow-up

Another form of interval-censoring arises in connection with life tables, where an individual may fail or be censored between specified observation times; it is known which occurred, but not the time of occurrence. This is discussed in Section 3.6.

2.3.2 Double Censoring

Other forms of interval censoring can arise. In many applications the lifetime T_i is the time between two events, for example, the time between infection with the human immunodeficiency virus (HIV) and the diagnosis of AIDS, the time between the beginning and end of a period of unemployment, or the time between the appearance of a crack in a metal specimen and its growth to a critical size. If the timing of the initial event is interval censored, then even if the exact time of the failure or censoring event is observed, the exact lifetime or censoring time for T is known only to lie in an interval.

Specifically, let U_i^* be the time of the initial event and suppose we observe only that $L_i^* < U_i^* \leq R_i^*$ under a scheme satisfying the conditions just specified. Let y_i be an observed censoring or failure time (i.e., time of the second event), measured on the same scale as U_i^*. Then the failure or censoring time for T_i is $t_i = y_i - U_i^*$, and we know only that $y_i - R_i^* \leq t_i < y_i - L_i^*$. This is known as double censoring. The likelihood function is not given by (2.3.1) with $U_i = y_i - R_i^*$ and $V_i = y_i - L_i^*$, in spite of the seeming similarity with standard interval-censoring. To see this, consider the p.d.f. $g_i(u)$ for the distribution of U_i^*, given that $U_i^* \in (L_i^*, R_i^*]$. If T_i is independent of U_i^*, the likelihood contributions are then given by

$$Pr(y_i, \delta_i | U_i^* \in (L_i^*, R_i^*]) = \int_{L_i^*}^{R_i^*} g_i(u) f_i(y_i - u)^{\delta_i} S_i(y_i - u)^{1-\delta_i} \, du. \qquad (2.3.3)$$

A difficulty in this case is the necessity to specify $g_i(u)$.

2.3.3 Remarks on Missing Data

Censoring is an example of incomplete or missing data: the exact values of lifetimes are unavailable for certain individuals. As such it may be considered in the context of general formulations for incomplete data (e.g., Little and Rubin 1987). A crucial issue is whether data are missing at random in some sense; if they are not, then a model that represents the process by which data are missing is necessary in order to obtain appropriate likelihood functions and inference procedures.

A mechanism that leads to missing data is sometimes completely independent of the lifetime process. The missing data are then said to be missing completely at random, or MCAR. Type 1 right censoring (Section 2.2.1.1), random independent censoring (Section 2.2.1.2) and interval censoring induced by prescheduled intermittent observation of individuals fall into this category. When data are MCAR we do not have to consider a model for the "missingness" process in order to obtain the observed likelihood, though it may be challenging to compute the probability distribution of the observed data.

The MCAR model is too stringent in many settings. A weaker requirement in the general theory is that data be missing at random (MAR), which means the probability that data are missing may depend on data that are observed, but not on data that are unobserved. In this case, it also proves possible to avoid specific modeling of the missingness process. The general censoring mechanism in Section 2.2.2 and the interval-censoring mechanism leading to (2.3.2) both satisfy the MAR condition, since censoring at any time is allowed to depend only on events observed in the past. Fortunately the observed likelihoods for both the MCAR and MAR censoring processes are not hard to obtain and, as we have seen, the observed likelihoods (2.2.3) and (2.3.1) are the same for the MCAR and MAR processes. It should be noted, however, that although we can avoid modeling the censoring process for inference based on the observed likelihood, we need a censoring model to evaluate exact frequency properties for small samples or for purposes of study design.

Information on censoring times or covariates may also be missing. A treatment of this topic is beyond the scope of this book; a few remarks and references are given in the Bibliographic Notes of Chapters 3, 4, and 7.

2.4 TRUNCATION AND SELECTION EFFECTS

In Section 2.1 we mentioned that in some studies individuals are not randomly selected and followed prospectively from their time origin ($t = 0$). In this case, it is necessary to consider the selection mechanism in writing down the likelihood function for observed data. This section considers some settings that involve selection effects.

2.4.1 Delayed Entry and Left Truncation

Individuals are sometimes selected and followed prospectively until failure or censoring, but their current lifetime at selection is not $t = 0$, but some value $u > 0$. The definition of a prospective study is that lifetime information after the time of selection forms the response. Selection of an individual at time u_i thus requires that $T_i \geq u_i$, and the observed data for individual i consist of $(u_i, t_i, \delta_i, \mathbf{x}_i)$, where $t_i \geq u_i$ is a lifetime or censoring time and $\mathbf{x}_i$ represents covariates. We say that the lifetime T_i is left truncated (at u_i) in this setting.

Let $S(t|\mathbf{x})$ be the survivor function for T given $\mathbf{x}$. The crucial issue affecting inference is whether the distribution of T given $\mathbf{x}$, u, and the fact that $T \geq u$ is given by the truncated distribution with survivor function $S(t|\mathbf{x})/S(u|\mathbf{x})$ for $t \geq u$. More specifically, in terms of the hazard function we need

$$Pr(T = t | T \geq t, u, T \geq u, \mathbf{x}) = Pr(T = t | T \geq t, \mathbf{x}). \qquad (2.4.1)$$

If this is the case and the conditions on the censoring process described in Section 2.2.2 hold, then the likelihood function arising from n individuals with independent lifetimes is given by (2.2.14) as

$$L = \prod_{i=1}^{n} \left[\frac{f_i(t_i)}{S_i(u_i)} \right]^{\delta_i} \left[\frac{S_i(t_i+)}{S_i(u_i)} \right]^{1-\delta_i}. \tag{2.4.2}$$

It is not always easy to determine whether (2.4.1) is a plausible assumption, especially when individuals in a study are selected from an ongoing process that defines T; Example 2.4.3 provides an illustration. One setting where (2.4.1) is valid is when observation of a process switches on at u_i in such a way that u_i is a stopping time with respect to the lifetime process (Andersen et al. 1993, Sec. 3.4); this is sometimes referred to as independent delayed entry. If we broaden the definition of $Y_i(t)$ in (2.2.12) and (2.2.14) to

$$Y_i(t) = I(u_i \le t \le t_i), \tag{2.4.3}$$

then (2.2.12) continues to hold and the product integral expression in (2.2.14) reduces directly to (2.4.2). The concept of independent delayed entry allows u_i to depend on prior lifetime history or on covariates $\mathbf{x}$ included in the lifetime model.

The following examples provide illustrations of delayed entry.

Example 2.4.1. A lifetime T can be viewed as the time between an initial event E_1 and a subsequent event E_2, for an individual. In many settings the events E_1 for different individuals occur at different points in calendar time, and individuals are selected for a prospective study by randomly choosing from those who have experienced E_1 but not E_2. For example, E_1 may refer to the onset of some disease that is typically fatal, and E_2 to death; T is the survival time from disease onset.

The selection mechanism and lifetime process are illustrated in Figure 2.2. Individuals experience E_1 at calendar time X and E_2 at time $X + T$; selection occurs at calendar time τ from individuals with $X \le \tau$ and $X + T > \tau$. The distribution of T is assumed to depend on y on information observable at τ; this could include the value of X, which would then be included among the covariates $\mathbf{x}$ in $S(t|\mathbf{x})$. The probability of selection could also depend on X or other covariates, provided they are included in the model for T. In this case, the conditions for independent delayed entry are met, with $u_i = \tau - X_i$, and (2.4.1) holds.

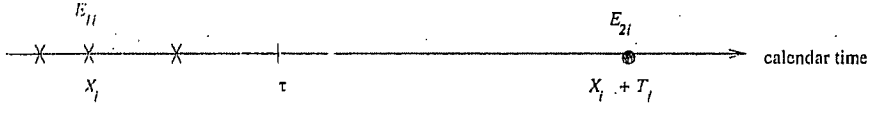

Figure 2.2. Selection conditional on survival to calendar time τ.

Example 2.4.2. A setup similar to that in the preceding example arose in a study on the lifetime of automobile brake pads (Kalbfleisch and Lawless 1992). The pads have a nominal lifetime, which is the number of miles or kilometers driven before the pads are reduced to a specified minimum thickness. To study the lifetime distribution, a manufacturer selected a random sample of vehicles sold over the preceding 12 months at a specific group of dealers. Only cars that still had the initial pads were selected. For each car the brake pad lifetime t_i could have then been observed by following the cars prospectively. Instead, to save time the current odometer reading u_i (in km) and the remaining pad thickness above the minimum were used in conjunction with the initial pad thickness to estimate the lifetime t_i (in km); this was treated as the actual lifetime in the analysis. In any case the selection framework is similar to that in Figure 2.2, and the lifetime t_i is left truncated at u_i.

Table 2.1 gives u_i and t_i values (in 1000-km units) for the left front brake pads on a sample of 98 vehicles.

Table 2.1. Brake Pad Life (t) and Odometer Readings (u) for 98 Cars

u	t	u	t	u	t	u	t
22.2	38.7	16.5	69.6	18.4	86.7	10.9	79.5
23.0	49.2	15.7	74.8	18.2	43.8	25.5	55.0
24.0	42.4	28.0	32.9	15.9	100.6	12.4	46.8
28.6	73.8	13.3	51.5	16.4	67.6	39.9	124.5
21.8	46.7	16.5	31.8	23.6	89.5	17.7	92.5
17.0	44.1	24.2	77.6	19.2	60.3	26.3	110.0
26.0	61.9	17.6	63.7	23.3	103.6	14.1	101.2
23.2	39.3	27.8	83.0	20.4	82.6	21.0	59.4
18.9	49.8	18.3	24.8	20.9	88.0	11.2	27.8
21.9	46.3	17.7	68.8	28.5	42.4	10.8	33.6
27.3	56.2	20.0	68.8	23.2	68.9	25.7	69.0
13.8	50.5	13.2	89.1	17.9	95.7	32.4	75.2
24.0	54.9	16.9	65.0	46.1	78.1	13.6	58.4
20.1	54.0	14.9	65.1	39.3	83.6	19.1	105.6
15.7	49.2	15.5	59.3	11.8	18.6	16.1	56.2
26.8	44.8	7.0	53.9	17.7	92.6	53.3	55.9
27.9	72.2	15.8	79.4	30.9	42.4	57.3	83.8
15.3	107.8	15.0	47.4	22.4	34.3	36.5	123.5
28.8	81.6	38.3	61.4	45.0	105.6	19.7	69.0
16.0	45.2	11.2	72.8	18.2	20.8	20.8	101.9
23.6	124.6	38.2	54.0	30.2	52.0	30.8	87.6
53.8	64.0	26.7	37.2	21.8	77.2	20.0	38.8
21.7	83.0	17.1	44.2	18.2	68.9	39.6	74.7
28.8	143.6	29.0	50.8	23.0	78.7		
17.0	43.4	18.3	65.5	27.2	165.5		

Note: Units are 1000 km.

Example 2.4.3. Cook et al. (1999) described a clinical trial involving patients with chronic bronchitis, for whom periods of exacerbation of symptoms alternated with periods of good respiratory health. Persons entering the study had to be undergoing an exacerbation spell at the time of selection, and were then assigned randomly to one of two treatments, A and B. The duration of exacerbation spells was a primary response for the comparison of A and B; let us consider the remaining duration of the initial exacerbation spell. The setup is on the surface similar to that in Figure 2.2, but because patients have a prior history of exacerbation spells, the selection mechanism will tend to pick patients with longer exacerbation periods, so the study population is not representative of the population of all patients; it is what we refer to as a length-biased sample (see Problem 2.7).

Because of the randomization, the study populations assigned to treatments A and B are comparable. One satisfactory approach is therefore to compare the distributions of the times V_i from randomization to the end of the initial exacerbation spell. Assuming that the duration U_i of the spell at the time of randomization is known, considering V_i given U_i is equivalent to considering T_i given U_i, where $T_i = U_i + V_i$ is the total duration of the initial spell. However, it would not be appropriate to treat T_i as a truncated response arising from independent delayed entry at U_i. Since treatment is not assigned until time U_i, and since the sample selected is not representative of the population of all patients anyway, basing comparisons on the marginal distribution of V_i seems the best approach.

The selection of individuals in the preceding example was not independent of their lifetimes. We consider some other such selection effects in the following section.

2.4.2 Retrospective Observation and Selection Effects

In a prospective study of lifetimes, individuals are typically followed from the entry time u_i to a failure or censoring time $t_i > u_i$. In some studies, the observational plan is retrospective to some degree. That is, part or all of the observation period (u_i, t_i) occurs chronologically prior to the selection of the individual. Such plans are attractive when it is not feasible to follow individuals long enough prospectively to obtain desired information, but they frequently impose conditions on the lifetimes of those selected. The next two examples provide illustrations of retrospective observation.

Example 2.4.4. Kalbfleisch and Lawless (1989) analyzed data on persons infected with HIV via blood transfusion, who were subsequently diagnosed with AIDS. The data were used to estimate the distribution of the time T between HIV infection and AIDS diagnosis.

The way the data were obtained was retrospective. In particular, the study group was assembled in 1987 and consisted of individuals who had a diagnosis of AIDS prior to July 1, 1986. For each person the date of HIV infection could also be ascertained, because the individuals selected were deemed to have contracted the HIV through a blood transfusion on a particular date. The condition for being included in the data set was therefore that $T_i \le v_i$, where

$$v_i = \text{Time between the individual's HIV infection and July 1, 1986.}$$

This is referred to as right truncation of the lifetime T_i, and the likelihood function based on n independent individuals is given by

$$\prod_{i=1}^{n} Pr(t_i | v_i, T_i \le v_i) = \prod_{i=1}^{n} \frac{f_i(t_i)}{F_i(v_i)}. \tag{2.4.4}$$

Example 2.4.5. Consider the estimation of distributions for the duration of first marriages. Because of the long duration of many marriages, retrospective ascertainment of data is attractive. One study design would be to sample couples married in the past, perhaps stratifying on specific time periods. (In any case, the duration of a marriage is likely related to the year of marriage, so the latter would be considered a covariate.) If couples were sampled from marriage records, without reference to whether the marriage was still intact or even if the individuals were still alive, then assuming that it was possible to trace the couples and determine the fate of the marriages, no selection effect would be present and a likelihood of the form (2.2.14) would apply. Tracing couples could be difficult and expensive, however, and an alternative plan would be to sample randomly individuals or married couples alive at the present time. By determining the previous history of marriage for such individuals, data on first marriage duration and related covariates could be obtained. However, it is a condition of selection that an individual be alive, and there is an association between duration of life and duration of marriage. Consequently we would expect that the distribution of marriage durations in the sampled population would not be exactly the same as in the population consisting of all couples who got married over the period of interest. In order to deal with this situation, we would need to formulate a model describing the ways individuals or couples are deleted from the population used for selection. Such issues take us into the realm of event history analysis (see Chapter 11).

2.5 INFORMATION AND DESIGN ISSUES

The planning of studies or experiments requires decisions about numbers of individuals, the duration of the study, the modes of sample selection and observation, and settings for controllable covariates. The decisions are based on time, resource, and physical constraints plus an assessment of the information on important parameters or hypotheses that the data will provide. This section contains some general remarks about information and design; specific applications are considered in subsequent chapters.

Studies are carried out for a variety of reasons, which may include furthering scientific understanding, the development of models for prediction or decision making,

and the improvement of processes or systems. Study objectives can often be linked to estimation or hypothesis-testing problems for specific quantities, which are then considered during the planning process. Thus, suppose there is a parametric lifetime distribution $f(t; \boldsymbol{\theta})$ and that some parameter $\psi = g(\boldsymbol{\theta})$ is of interest. For example, this might be a distribution quantile or survival probability. Putting aside questions of model adequacy, we consider the precise estimation or testing of ψ.

The sampling properties of tests or estimation procedures can sometimes be determined analytically, and in general can be examined through simulation. Most of the procedures in this book are likelihood-based, as described in Section 2.1, and the discussion in this section focuses on them. Appendix C summarizes likelihood-based inference and important large-sample results that underlie the following discussion.

Confidence intervals or tests for a scalar ψ are often based either on (1) the likelihood ratio statistic

$$\Lambda(\psi) = 2\ell(\hat{\boldsymbol{\theta}}) - 2\ell(\tilde{\boldsymbol{\theta}}(\psi)), \tag{2.5.1}$$

where $\ell(\boldsymbol{\theta})$ is the log-likelihood function, $\hat{\boldsymbol{\theta}}$ is the m.l.e. that maximizes $\ell(\boldsymbol{\theta})$, and $\tilde{\boldsymbol{\theta}}(\psi)$ is the vector $\boldsymbol{\theta}$ that maximizes $\ell(\boldsymbol{\theta})$ under the constraint $g(\boldsymbol{\theta}) = \psi$, or (2) the standardized quantity

$$Z(\psi) = (\hat{\psi} - \psi)/\hat{V}_{\psi}^{1/2}, \tag{2.5.2}$$

where $\hat{\psi} = g(\hat{\boldsymbol{\theta}})$ is the m.l.e. of ψ and $\hat{V}_{\psi}^{1/2}$ is its standard error. The latter is usually based on

$$\hat{V}_{\psi} = (\partial\hat{\psi}/\partial\hat{\boldsymbol{\theta}})'\hat{\mathbf{V}}_{\theta}(\partial\hat{\psi}/\partial\hat{\boldsymbol{\theta}}), \tag{2.5.3}$$

where $\hat{\mathbf{V}}_{\theta}$ is an estimate of the asymptotic covariance matrix for $\hat{\boldsymbol{\theta}}$, typically either $\mathbf{I}(\hat{\boldsymbol{\theta}})^{-1}$ or $\mathcal{I}(\hat{\boldsymbol{\theta}})^{-1}$, where $I(\boldsymbol{\theta})$ and $\mathcal{I}(\boldsymbol{\theta})$ are the observed and expected information matrices for $\boldsymbol{\theta}$, respectively, discussed in Section 2.2.3 and Appendix C.

$\Lambda(\psi)$ and $Z(\psi)$ are approximate pivotal quantities in standard settings, with asymptotic $\chi^2_{(1)}$ and $N(0, 1)$ distributions, respectively. Two-sided α confidence intervals based on Z are of the form $\hat{\psi} \pm z\hat{V}_{\psi}^{1/2}$, where z is the $.5(1 + \alpha)$ quantile for $N(0, 1)$; two-sided α confidence intervals based on Λ are obtained as the set of values ψ such that $\Lambda(\psi) \le \chi^2_{(1),\alpha}$, the α quantile of $\chi^2_{(1)}$.

The following example illustrates some important points in the simple context of an exponential distribution.

Example 2.5.1. Suppose that in a test environment a piece of equipment has an exponentially distributed lifetime T, with p.d.f. $f(t; \theta) = \theta^{-1}\exp(-t/\theta)$. From (2.2.5), the log-likelihood function for θ under a variety of prospective observation schemes with right-censoring is

$$\ell(\theta) = -r\log\theta - \frac{1}{\theta}\sum_{i=1}^{n} t_i, \tag{2.5.4}$$

where t_i denotes a lifetime or censoring time and $r = \sum \delta_i$ is the number of t_i that are lifetimes.

The m.l.e. from (2.5.4) is $\hat{\theta} = \sum t_i/r$, and the information is

$$I(\theta) = \frac{-d^2\ell}{d\theta^2} = \frac{-r}{\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^{n} t_i. \tag{2.5.5}$$

The expected information depends on the censoring process, but it turns out that a simple general expression is available. Since $E(d\ell/d\theta) = 0$ for standard maximum likelihood (Appendix C), it follows from (2.5.4) that $E(\sum t_i) = \theta E(r)$, and from (2.5.5) we then get

$$\mathcal{I}(\theta) = E\{I(\theta)\} = \frac{E(r)}{\theta^2}. \tag{2.5.6}$$

A two-sided approximate .95 confidence interval for θ based on (2.5.2) with $\hat{V}_{\theta} = I(\hat{\theta})^{-1} = \hat{\theta}^2/r$ consists of values of θ that satisfy $|Z(\theta)| \le 1.96$, or

$$Z(\theta)^2 = r\left(1 - \frac{\theta}{\hat{\theta}}\right)^2 \le 3.84. \tag{2.5.7}$$

The analogous interval based on (2.5.1) consists of values satisfying

$$\Lambda(\theta) = 2r\left[\frac{\hat{\theta}}{\theta} - 1 - \log(\hat{\theta}/\theta)\right] \le 3.84. \tag{2.5.8}$$

The two intervals can be seen to agree more and more closely as r increases. When the intervals are not in close agreement, the one based on (2.5.8) is preferred, as discussed in Appendix C and in Section 4.1.1. In either case, it is the number of uncensored lifetimes r that determines the relative precision with which θ is estimated. For example, the relative width of the .95 confidence interval (i.e., the width divided by $\hat{\theta}$) based on (2.5.7) is $3.92r^{-1/2}$.

Let us consider design issues. It is possible to design studies where r is fixed; Type 2 censored life test plans (Section 2.2.1.3) are of this type. However, the duration of the study is random when r is fixed, and it is more common to use designs for which r is a random variable and the study duration is fixed. For example, if we test each of n items over a specified time period $(0, C)$, then $Pr(T_i \le C) = 1 - \exp(-C/\theta)$, and r has a binomial distribution with

$$E(r) = n(1 - e^{-C/\theta}). \tag{2.5.9}$$

This leads to Type 1 censored data (Section 2.2.1.1). The value of $E(r)$ or, equivalently, the expected information (2.5.6), provides an idea of the precision of estimation expected from a study and can be used for planning purposes, though we also need a value for θ in order to evaluate (2.5.9).

The expected information and $E(r)$ increase as n or C increases. Suppose that we want a .95 confidence interval with a relative width of about 1 when (2.5.7) is used.

This requires that $3.92r^{-1/2} \doteq 1$, or that $r \doteq 16$. By choosing n and C/θ suitably large we can make $E(r)$ or the probability that $r \geq 16$ as large as desired. If we want to make $E(r) = 16$, for example, two among an infinity of choices are (a) $n = 25$, $C/\theta = 1$, and (b) $n = 19$, $C/\theta = 2$, illustrating the trade-off between sample size and duration of study. In either case, we need to provide a value for θ in order to determine C, and θ is what we are trying to estimate! The conventional approach is to use a conservatively large value of θ.

Calculations based on expected information provide a rough idea of the number of individuals (n) and length of study (C) needed to achieve the desired precision in the estimation of θ. In more complex settings it is difficult to get much insight analytically, and a useful procedure is to simulate data sets under proposed study plans and provisional values for θ. This allows a comparison of confidence intervals based on alternatives such as (2.5.7) and (2.5.8), and displays the sampling variation inherent in the study process

In studies on lifetimes the censoring process is part of the design. The preceding example illustrated the effects of censoring and sample size in a very simple setting. Qualitatively similar effects occur in other settings and with parametric models other than the exponential distribution. The precision of nonparametric estimates also depends on the type and degree of censoring. For the product-limit estimator of $S(t)$ introduced in Chapter 3, for example, the dependence is explicit in variance estimates like (3.2.3). Other constraints on follow-up (e.g., intermittent observation that leads to interval censoring), or on the ways in which individuals are selected for study, are also part of the design and affect the information about parameters. Finally, the study design affects our ability to assess model assumptions. This is an important issue, especially when the results or conclusions from an analysis are highly model-dependent.

Analogous considerations apply in studies where hypothesis tests are a primary concern, for example, in comparisons of the efficacy of two medical treatments or of the reliability of competing industrial products. The power, or ability of tests to detect effects of a specified size, depends upon the same factors as precision of estimation. Two other aspects of study design should also be mentioned. The first concerns experiments with controllable factors or covariates. In this case, the selection of factor levels affects the information about parameters, and principles of experimental design may be used to construct economic, efficient designs. Second, adaptive or sequential plans are sometimes useful. For example, in a clinical trial to compare two treatments, we may wish to terminate the study early if it becomes obvious that one treatment is markedly superior. This topic is considered briefly in Section 4.1.4.

BIBLIOGRAPHIC NOTES

Maximum likelihood essentially dates from Fisher (1922), and contributions from many people have brought likelihood methods to their current position. Appendix C contains a summary of key theoretical results and important inferential techniques.

Sukhatme (1937), Boag (1949), Epstein and Sobel (1953), and others considered maximum likelihood in conjunction with censored data. Early discussions of asymptotic properties were given by Halperin (1952) and Bartholomew (1957, 1963) for the cases of Type 2 and Type 1 censoring, respectively. The more general concept of independent censoring and construction of the likelihood functions as described in Section 2.2.2 started with Cox (1975), with subsequent contributions by Kalbfleisch and MacKay (1978) and Kalbfleisch and Prentice (1980, Sec. 5.2). Other discussions of likelihood construction were given by Efron (1977), Williams and Lagakos (1977) and Lagakos (1979). The rigorous development of asymptotic likelihood theory under independent censoring was facilitated by the use of martingale theory (e.g., Aalen 1978a, b, Borgan 1984, Arjas 1989); Arjas and Haara (1984, 1992) discuss issues associated with observation schemes in both survival and event history analysis. Andersen et al. (1993, Chs. 2 and 3) is an important source concerning these areas.

Interval censoring was considered by Peto (1973) and by Turnbull (1976), who also discussed general forms of truncation. Huang and Wellner (1997) specify different types of interval-censoring. A discussion of the process by which inspection times for event history processes are determined is given by Grüger et al. (1991) and Farewell et al. (2002). Jewell and van der Laan (1997) and J. Sun (1997) provide historical remarks and examples of double censoring.

Truncation was considered by Turnbull (1976) and Hyde (1977). Keiding (1992) considered independent truncation mechanisms; Andersen et al. (1993, Secs. 3.3 and 3.4) give a detailed mathematical discussion of likelihood construction. For examples of left truncation in the social sciences, see Hamerle (1991) and Guo (1993), and in medicine, Cnaan and Ryan (1989).

Design issues are best considered in specific contexts. However, for an early discussion of "optimal" design in connection with maximum likelihood estimation, see Chernoff (1953).

Bayesian methods are based on the likelihoods described here and prior distributions on unknown parameters. Box and Tiao (1973), Berger (1985), Carlin and Louis (1996), and Gelman et al. (1995) discuss Bayesian inference; Berger (2000) provides many additional references. Martz and Waller (1982), Crowder et al. (1991, Ch. 6), and Meeker and Escobar (1998, Ch. 14) discuss Bayesian methods in reliability, and Ibrahim et al. (2001) deal with survival analysis. Gilks et al. (1996) discuss applications in biostatistics. It is beyond the scope of this book to describe Bayesian methods in detail, but occasional references will be made.

PROBLEMS AND SUPPLEMENTS

2.1 Consider experiments with the following two censoring mechanisms.

(a) A group of n units is observed from time 0; observation stops at the time of the rth failure or at time C, whichever occurs first.

(b) A group of n units is observed from time 0, but each time a unit fails a new unit instantly replaces it in the experiment. The experiment terminates after a preassigned time C has elapsed.

Show by direct calculation that in each case the likelihood function is of the form (2.2.3), assuming that the units have failure times which are i.i.d. with survivor function $S(t)$ and p.d.f. $f(t)$.

(Section 2.2)

2.2 Suppose that the lifetime T_i has hazard function $h_i(t)$ and that C_i is a random censoring time associated with T_i. Define

$$\lambda_i(t) = \lim_{\Delta t \to 0} \frac{Pr(t \le T_i < t + \Delta t | T_i \ge t, C_i \ge t)}{\Delta t}.$$

(a) Show that the independent censoring condition (2.2.11) is equivalent to the condition $h_i(t) = \lambda_i(t)$, assuming that we condition on fixed covariates and that there are no time-varying covariates.

(b) Suppose that there exists an unobserved covariate Z_i which affects both T_i and C_i, as follows:

$$Pr(T_i \ge t | Z_i) = \exp(-Z_i \theta t), \qquad Pr(C_i \ge s | Z_i) = \exp(-Z_i \rho s),$$

and T_i, C_i are independent, given Z_i. Assume further that Z_i has a gamma distribution (1.3.15) with mean 1 and variance ϕ^{-1}. Show that the joint survivor function for T_i, C_i is

$$Pr(T_i \ge t, C_i \ge s) = \left(1 + \frac{1}{\phi}\theta t + \frac{1}{\phi}\rho s\right)^{-\phi}.$$

(c) Obtain $h_i(t)$ and $\lambda_i(t)$ for the model in part (b), and show that $\lambda_i(t) < h_i(t)$ for finite ϕ. Thus the censoring mechanism is not independent, but it approaches independence as $\phi \to \infty$.

(Section 2.2.2)

2.3 *The effect of grouping.* Consider lifetime data that are grouped or rounded off to some degree. In particular, suppose that lifetimes from an exponential distribution that are recorded as t actually lie in the interval $(t - \Delta/2, t + \Delta/2)$. For simplicity, censoring times recorded as t will be assumed to be exactly equal to t. Consider a censored sample of n observations involving r lifetimes and $n - r$ censoring times. This gives a likelihood function

$$L(\theta) = \theta^{-r} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} t_i\right),$$

corresponding to (2.5.4) in the case where $\Delta = 0$.

(a) For $\Delta > 0$, show that the likelihood function for θ is

$$L_1(\theta) = \left(e^{\Delta/2\theta} - e^{-\Delta/2\theta}\right)^r \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} t_i\right).$$

(b) Show that the expected information based on $L_1(\theta)$ is

$$I_1(\theta) = \frac{E(r)}{\theta^2} g\left(\frac{\Delta}{\theta}\right),$$

where $g(a) = a^2 e^{-a}/(1 - e^{-a})^2$. Examine the loss of information entailed by grouping, noting that the expected information based on $L(\theta)$ is $E(r)/\theta^2$.

(Sections 2.3, 2.5)

2.4 *Loss of information from right truncation.* Consider the following observational schemes associated with a lifetime T having p.d.f. $f(t; \theta)$ and survivor function $S(t; \theta)$:

1. Type 1 censoring occurs at the prespecified time C, giving the likelihood function

$$L(\theta) = \left\{\prod_{i=1}^{r} f(t_i; \theta)\right\} S(C; \theta)^{n-r},$$

where $t_1, \ldots, t_r$ are the observed lifetimes.

2. Only the individuals with $T_i \le C$ are known about and observed, giving the likelihood function

$$L_1(\theta) = \prod_{i=1}^{r} \left\{\frac{f(t_i; \theta)}{F(C; \theta)}\right\}.$$

(a) Compare the observed and expected information about θ in cases 1 and 2.

(b) Examine the loss of information numerically as a function of C/θ when $f(t; \theta) = \theta^{-1} \exp(-t/\theta)$ is an exponential distribution.

(Sections 2.4, 2.5; Kalbfleisch and Lawless 1988b)

2.5 *Random truncation models.* Suppose that a lifetime T_i has an associated random left-truncation time U_i, as in Section 2.4.1. Let T_i have hazard function $h_i(t)$ and p.d.f. $f_i(t)$, where there are only fixed covariates present.

(a) Show that the condition (2.4.1) holds if T_i and U_i are independent, given the covariate values.

(b) Show that the independence in part (a) can be weakened by showing that (2.4.1) holds if the joint p.d.f. of T_i and U_i, given $U_i \le T_i$, is of the form $f_i(t) g_i(u)$.

(c) Extend this treatment to deal with right truncation.

(Section 2.4.1; Tsai 1990; Wellek 1990)

2.6 *Random effects and left truncation.* For delayed entry settings as portrayed in Figure 2.2, it may sometimes be the case that T_i is not independent of $U_i = \tau - X_i$, as implied in Problem 2.5. Consider the case where T_i is independent of U_i, given an unobserved random effect Z_i for individual i. Show that the p.d.f. of T_i, given $U_i = u$ and $T_i \geq u$, is

$$\int \frac{f_1(t|z)}{S_1(u|z)} g^*(z|U_i = u, T_i \geq u)\, dz,$$

where $f_1(t|z)$ and $S_1(u|z)$ are the p.d.f. and survivor function of T_i, given $Z_i = z$, and $g^*(z|U_i = u, T_i \geq u)$ is the conditional p.d.f. of Z_i, given the conditioning events. In general, this p.d.f. does not equal $f(t)/S(u)$, where $f(t)$ and $S(t)$ are the p.d.f. and survivor function for T_i. Furthermore, the p.d.f. g^* cannot in general be assumed independent of u.

(Section 2.4; Lawless and Fong 1999)

2.7 *Sampling renewal processes and left truncation.* Suppose that we wish to estimate the distribution of time between successive events in a population of renewal processes (Cox 1962; Ross 1983). If a process is intercepted at time τ, then Figure 2.2 describes the occurrence times of the two events E_{1i} and E_{2i} that bracket τ. If the renewal process is in equilibrium, then (see Cox 1962 or Ross 1983) the joint p.d.f. of $U_i = \tau - X_i$ and T_i is

$$g_i(u, t) = \frac{1}{\mu_i} f_i(t) \qquad 0 \leq u \leq t,$$

where $f_i(t)$ is the p.d.f. for the time T_i between events and $\mu_i = E(T_i)$, which is assumed to exist.

(a) Show that the marginal distribution of T_i is $t f_i(t)/\mu_i$. This is called a length-biased density. Examine its forms relative to $f_i(t)$, when $f_i(t)$ is an exponential distribution, and a Weibull distribution with shape parameter β.

(b) Show that the condition (2.4.1) for independent delayed entry holds.

(c) Consider the case where $f_i(t)$ depends on an unobservable random effect Z_i, so that $g_i(u, t)$ does as well. Show that the condition (2.4.1) does not now hold in general. Investigate the case where $f_i(t|z_i) = \lambda z_i \exp(-\lambda z_i t)$ is exponential, and Z_i has a gamma distribution (1.3.15) with mean 1 and variance ϕ^{-1}.

(Section 2.4)

CHAPTER 3

Some Nonparametric and Graphical Procedures

3.1 INTRODUCTION

Graphs and simple data summaries are important for both description and analysis of data. They are closely related to nonparametric estimates of distributional characteristics; many graphs are just plots of some estimate. This chapter introduces nonparametric estimation and procedures for portraying univariate lifetime data.

Tools such as frequency tables and histograms, empirical distribution functions, probability plots, and data density plots are familiar across different branches of statistics. For lifetime data, the presence of censoring makes it necessary to modify the standard methods. To illustrate, let us consider one of the most elementary procedures in statistics, the formation of a relative-frequency table. Suppose we have a complete (i.e., uncensored) sample of n lifetimes from some population. Divide the time axis $[0, \infty)$ into $k + 1$ intervals $I_j = [a_{j-1}, a_j)$, $j = 1, \ldots, k + 1$, where $0 = a_0 < a_1 < \cdots < a_k < a_{k+1} = \infty$; with a_k being the upper limit on observation. Let d_j be the observed number of lifetimes that lie in I_j. A frequency table is just a list of the intervals and their associated frequencies, d_j, or relative frequencies, d_j/n. A relative-frequency histogram, consisting of rectangles with bases on $[a_{j-1}, a_j)$ and areas d_j/n $(j = 1, \ldots, k)$, is often drawn to portray this. When data are censored, however, it is generally not possible to form the frequency table, because if a lifetime is censored, we do not know which interval, I_j, it lies in. As a result, we cannot determine the d_j.

Section 3.6 describes how to deal with frequency tables when data are censored; this is referred to as life table methodology. First, however, we develop methods for ungrouped data. Section 3.2 discusses nonparametric estimation of distribution, survivor, or cumulative hazard functions under right censoring. This also forms the basis for descriptive and diagnostic plots, which are presented in Section 3.3. Sections 3.4 and 3.5 deal with the estimation of hazard functions and with nonparametric estimation from some other types of incomplete data.

3.2 NONPARAMETRIC ESTIMATION OF A SURVIVOR FUNCTION AND QUANTILES

3.2.1 The Product-Limit Estimate

A useful way of portraying a random sample $t_1, \ldots, t_n$ is to graph the empirical survivor function or empirical distribution function. This also provides nonparametric estimates of the distribution under study. If there are no censored observations in a sample of size n, the empirical survivor function (ESF) is defined as

$$\hat{S}(t) = \frac{\text{Number of observations} \geq t}{n} \qquad t \geq 0. \tag{3.2.1}$$

This is a step function that decreases by $1/n$ just after each observed lifetime if all observations are distinct. More generally, if there are d lifetimes equal to t, the ESF drops by d/n just past t.

When dealing with censored lifetime data, some modification of (3.2.1) is necessary, since the number of lifetimes greater than or equal to t will not generally be known exactly. The modification described here is called the product-limit (PL) estimate of the survivor function, or the Kaplan–Meier (KM) estimate, after the authors who first discussed its properties (Kaplan and Meier, 1958). Let (t_i', δ_i), $i = 1, \ldots, n$ represent a censored random sample of lifetimes, in the notation of Section 2.2. Suppose that there are k ($k \leq n$) distinct times $t_1 < t_2 < \cdots < t_k$ at which deaths occur. The possibility of there being more than one death at t_j is allowed, and we let $d_j = \sum I(t_i' = t_j, \delta_i = 1)$ represent the number of deaths at t_j. In addition to the lifetimes $t_1, \ldots, t_k$, there are also censoring times for individuals whose lifetimes are not observed. The PL estimate of $S(t)$ is defined as

$$\hat{S}(t) = \prod_{j:t_j < t} \frac{n_j - d_j}{n_j}, \tag{3.2.2}$$

where $n_j = \sum I(t_i' \geq t_j)$ is the number of individuals at risk at t_j, that is, the number of individuals alive and uncensored just prior to t_j. If a censoring time and a lifetime are recorded as equal, we adopt the convention that the censoring time is infinitesimally larger. Thus any individuals with censoring times recorded as equal to t_j are included in the set of n_j individuals at risk at t_j, as are individuals who die at t_j. This convention is consistent with assumptions about censoring in Chapter 2. Another point about (3.2.2) concerns situations in which the largest observed time in the sample is a censoring time. In this case the PL estimate is defined only up to this last observation. The reason for this is explained later.

The estimate (3.2.2) is derived below as a nonparametric maximum likelihood estimate (m.l.e.), but intuitively it can be viewed as arising from the expression (1.2.8) for the survivor function of a discrete distribution, with the hazard function $h(t_j) = Pr(T = t_j | T \geq t_j)$ estimated by d_j/n_j. When there is no censoring, $n_1 = n$ and $n_j = n_{j-1} - d_{j-1}$ ($j = 2, \ldots, k$), and (3.2.2) reduces to the ordinary ESF (3.2.1). In both the censored and uncensored cases $\hat{S}(t)$ is a left-continuous step function that equals 1 at $t = 0$ and drops by a factor $(n_j - d_j)/n_j$ immediately after each lifetime t_j. The estimate does not change at censoring times; the effect of the censoring times is, however, felt in the values of n_j and hence in the sizes of the steps in $\hat{S}(t)$.

Before we examine the PL estimate and its properties further, let us use it in an example. Numerous software packages provide PL estimates, but for illustration its calculation is described.

Example 3.2.1. Example 1.1.7 gave remission times for two groups of leukemia patients, one given the drug 6-MP and the other a placebo. Table 3.1 outlines the calculation of the PL estimates of the survivor functions for remission time distributions associated with the two groups, and Figure 3.1 shows these on a graph. The PL estimates have $\hat{S}(0) = 1$ and have jumps just after each observed lifetime, so in the table we show the values $\hat{S}(t_j+)$. The PL estimate is easily calculated recursively, since $\hat{S}(t_1+) = (n_1 - d_1)/n_1$ and

$$\hat{S}(t_j+) = \hat{S}(t_{j-1}+)\frac{n_j - d_j}{n_j} \qquad j = 2, \ldots, k.$$

The PL estimate for the drug 6-MP group is defined only up to $t = 35$, since the last observed time for that sample is a censoring time, $C = 35$. Standard errors, described below, are also shown for each $\hat{S}(t_j+)$.

The graph is a very useful representation of the survival experience of the two groups and suggests the superiority of the drug 6-MP over the placebo in prolonging survival. Formal methods of testing and estimating differences in two or more lifetime distributions are discussed in later chapters.

Table 3.1. Computation of Two PL Estimates

Drug 6-MP					Placebo				
t_j	n_j	d_j	$\hat{S}(t_j+)$	se	t_j	n_j	d_j	$\hat{S}(t_j+)$	se
6	21	3	.857	.076	1	21	2	.905	.064
7	17	1	.807	.087	2	19	2	.810	.086
10	15	1	.753	.096	3	17	1	.762	.093
13	12	1	.690	.107	4	16	2	.667	.103
16	11	1	.627	.114	5	14	2	.571	.108
22	7	1	.538	.128	8	12	4	.381	.106
23	6	1	.448	.135	11	8	2	.286	.099
					12	6	2	.190	.086
					15	4	1	.143	.076
					17	3	1	.095	.064
					22	2	1	.048	.047
					23	1	1	.0	

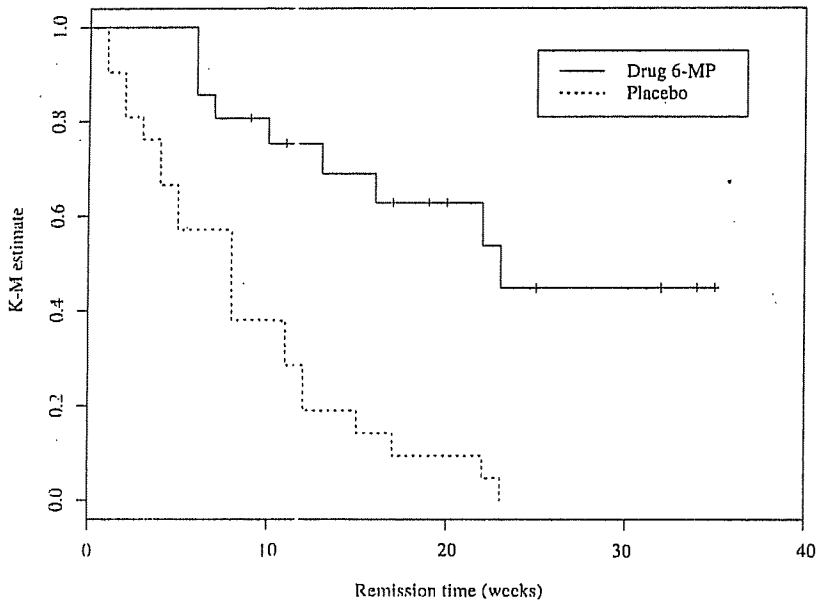

Figure 3.1. PL (Kaplan–Meier) survivor function estimates for remission duration.

3.2.1.1 *Variance Estimation*

When using PL estimates it is desirable to have an estimate of the variance of $\hat{S}(t)$. Proceeding along lines described below, we obtain the estimate

$$\widehat{\text{Var}}[\hat{S}(t)] = \hat{S}(t)^2 \sum_{j:t_j<t} \frac{d_j}{n_j(n_j - d_j)}, \tag{3.2.3}$$

which is often referred to as Greenwood's formula. It is easily shown that when there is no censoring, (3.2.3) reduces to the usual variance estimate $\hat{S}(t)[1-\hat{S}(t)]/n$. Standard errors (SE) for $\hat{S}(t)$ are given by the square root of (3.2.3). As an illustration of (3.2.3), we find an estimate of the variance of $\hat{S}(15)$ for the drug 6-MP group in Example 3.2.1 to be

$$\begin{aligned}\widehat{\text{Var}}[\hat{S}(15)] &= 0.690^2\left(\frac{3}{21(18)} + \frac{1}{17(16)} + \frac{1}{15(14)} + \frac{1}{12(11)}\right)\\ &= 0.011403,\end{aligned}$$

which gives an estimated standard deviation of 0.107. This and similar standard errors for the Placebo group provide a clearer picture of the significance of the difference in survivor functions in Figure 3.1. Confidence intervals for $S(t)$ are considered in Section 3.2.4.

3.2.1.2 *The PL Estimate as an MLE*

The PL estimate can be derived as a nonparametric m.l.e. of the survivor function, $S(t)$. This is quite straightforward in the discrete-time setting, so we consider this first.

Assume that independent lifetimes $T_1, \ldots, T_n$ have a discrete distribution with survivor function $S(t)$ and hazard function $h(t)$, where without loss of generality we take $t = 0, 1, 2, \ldots$. The key idea in the subsequent development is to consider the distribution of T through its hazard function $h(t)$, treating this as the parameter. Under the assumptions about censoring in Section 2.2.2, the observed likelihood function takes the form (2.2.12), which when $h_i(t) = h(t)$, is

$$L = \prod_{i=1}^{n}\prod_{t=0}^{\infty} h(t)^{dN_i(t)}[1-h(t)]^{Y_i(t)(1-dN_i(t))}. \tag{3.2.4}$$

Recall that with the notation used in (3.2.4), t_i represents the lifetime or censoring time for individual i, $\delta_i = I(t_i \text{ is a lifetime})$, $Y_i(t) = I(t_i \geq t)$, and $dN_i(t) = I(t_i = t, \delta_i = 1)$. We can rewrite (3.2.4) as

$$L = \prod_{t=0}^{\infty} h(t)^{d_t}[1-h(t)]^{n_t - d_t}, \tag{3.2.5}$$

where

$$d_t = \sum_{i=1}^{n} dN_i(t),\ n_t = \sum_{i=1}^{n} Y_i(t) \tag{3.2.6}$$

are the observed number of lifetimes equal to t and the number of individuals at risk (alive and uncensored) at t, respectively.

Considering the vector $\mathbf{h} = (h(0), h(1), \ldots)$ as the parameter in the lifetime distribution, we have the likelihood $L(\mathbf{h})$ from (3.2.5), and easily find that it is maximized at $\hat{h}(t) = d_t/n_t$ $(t = 0, 1, \ldots, \tau)$, where $\tau = \max(t : n_t > 0)$. The relationship (1.2.8) then gives the m.l.e. of $S(t)$ for $t = 0, 1, \ldots, \tau$ as

$$\begin{aligned}\hat{S}(t) &= \prod_{s=0}^{t-1}[1-\hat{h}(s)]\\ &= \prod_{s=0}^{t-1}\left(1 - \frac{d_s}{n_s}\right),\end{aligned} \tag{3.2.7}$$

which we recognize as identical to (3.2.2) in the discrete-time setting. When n_t equals zero, there are no terms involving $h(t), h(t+1), \ldots$ in (3.2.5), and thus there

is no information about $h(s)$ for $s > \tau$. If $d_\tau < n_\tau$ the estimate $\hat{S}(\tau+) > 0$, and the estimate is undefined beyond τ+; this happens when the largest observed t_i is a censoring time. If, however, $d_\tau = r_\tau$, then $\hat{S}(\tau+) = 0$, and since $S(t)$ is nonincreasing, the estimate of $S(t)$ is 0 for all $t > \tau$.

The variance estimate (3.2.3) can be obtained from standard maximum likelihood large-sample theory of Appendix C, if we assume $S(t) = 0$ for $t >$ some value τ. The information matrix $I(\mathbf{h})$ is easily seen to have diagonal entries $I_{rr}(\mathbf{h}) = -\partial^2 \log L/\partial h(r)^2 = n_r/\{h(r)[1 - h(r)]\}$ and off-diagonal entries equal to 0. Straightforward use of the large-sample result $\text{Asvar}(\hat{\mathbf{h}}) = I(\hat{\mathbf{h}})^{-1}$, and the asymptotic variance formula (B.2), then gives

$$\begin{aligned}\widehat{\text{Asvar}}\{\hat{S}(t)\} &= \hat{S}(t)^2 \widehat{\text{Asvar}}\{\log \hat{S}(t)\} \\ &= \hat{S}(t)^2 \sum_{s=0}^{t-1} \widehat{\text{Asvar}}\{\log[1 - \hat{h}(s)]\} \\ &= \hat{S}(t)^2 \sum_{s=0}^{t-1} \frac{\hat{h}(s)[1 - \hat{h}(s)]^{-1}}{n_s}, \end{aligned} \tag{3.2.8}$$

which is the same as (3.2.3).

The same development goes through when the data are subject to independent left truncation, as well as right censoring. In that case (2.2.12) still holds, as discussed in Section 2.4, with $Y_i(t)$ merely redefined by (2.4.3), as $Y_i(t) = I(u_i \le t \le t_i)$. However, from (1.2.18) we see that in this case we can estimate only $Pr(T \ge t | T \ge u_{\min})$, where $u_{\min} = \min(u_1, \ldots, u_n)$. To estimate the unconditional survivor function $S(t)$, we must have $u_{\min} = 0$. Left-truncated data are discussed more fully in Section 3.5.1.

Continuous time or general distributions can be handled as a limit of the discrete-time case, as in the development of (2.2.14). We now think of the cumulative hazard function $H(t)$ as the parameter to be estimated. With $dH(t)$ as the cumulative hazard function increment over $[t, t + dt)$, the likelihood (2.2.14) becomes the product integral

$$L = \prod_{(0,\infty)} dH(t)^{dN.(t)}[1 - dH(t)]^{Y.(t) - dN.(t)}, \tag{3.2.9}$$

where $dN.(t) = \sum_i dN_i(t)$ and $Y.(t) = \sum_i Y_i(t)$. If we consider (3.2.9) with respect to the space of all cumulative hazard functions $H(t)$, it is clear that (3.2.9) is maximized for a function with jumps at each distinct observed lifetime. If not, then we would have $dH(t) = 0$ when $dN.(t) > 0$ and the value of L would be zero. It follows by direct comparison with the discrete-time case that L is maximized by the function $H(t)$ with increments

$$d\hat{H}(t) = \frac{dN.(t)}{Y.(t)}, \qquad t \ge 0,\ Y.(t) > 0. \tag{3.2.10}$$

When $Y.(t)$ equals zero, $d\hat{H}(t)$ is undefined. By (1.2.16) the m.l.e. of $S(t)$ in the general case is then

$$\hat{S}(t) = \prod_{(0,t)} [1 - d\hat{H}(u)], \tag{3.2.11}$$

which is precisely (3.2.2). As in the previous discrete-time development, the estimates (3.2.10) and (3.2.11) continue to hold when there is independent left truncation with $u_{\min} = 0$, with $Y_i(t)$ defined by (2.4.3).

The derivation just given glosses over technical issues concerning the parameter space, which is a space of functions $H(t)$. It is an interesting feature of nonparametric maximum likelihood that even if we wish to consider $H(t)$ everywhere continuous, we are forced to admit functions with discontinuities in the parameter space, and find that the m.l.e. is a discrete distribution. More rigorous discussions are provided by Johansen (1978) and references cited in the Bibliographic Notes at the end of the chapter.

We gave a variance estimate (3.2.3) for the PL estimate and motivated it by using standard maximum likelihood large-sample theory in the discrete-time case to get (3.2.8). However, nonparametric estimation requires developments beyond the finite parameter theory of Appendix C, and so a rigorous treatment of the asymptotic properties of $\hat{S}(t)$ in the continuous-time case has to be pursued separately. Mathematically detailed developments are given in several sources; we outline some key ideas in Section 3.2.4 and provide references to full discussions at the end of the chapter.

3.2.2 The Nelson–Aalen Estimate

The estimate of the cumulative hazard function corresponding to (3.2.10) is given by the Riemann–Stieltjes integral (1.2.4) as

$$\begin{aligned}\hat{H}(t) &= \int_0^t d\hat{H}(u) \\ &= \int_0^t \frac{dN.(u)}{Y.(u)}, \end{aligned} \tag{3.2.12}$$

where we assume that $Y.(u) > 0$ for $0 \le u \le t$. This is sometimes called the empirical cumulative hazard function, but is more commonly known as the Nelson–Aalen (NA) estimate, having been proposed by Nelson (1969) and by Aalen in a 1972 thesis. In the notation used for the Kaplan–Meier estimate in (3.2.2),

$$\hat{H}(t) = \sum_{j:t_j \le t} \frac{d_j}{n_j}, \tag{3.2.13}$$

where $t_1, \ldots, t_k$ represent the distinct times at which failures are observed.

Plots of $\hat{H}(t)$ give useful information about the shape of the hazard function; note, for example, that $H(t)$ is linear if $h(t)$ is constant, and convex if $h(t)$ is monotonic.

The maximum likelihood development leading to (3.2.8) also provides an estimate of the asymptotic variance for $\hat{H}(t)$ as

$$\widehat{\text{Var}}[\hat{H}(t)] = \sum_{j:t_j \le t} \frac{d_j(n_j - d_j)}{n_j^3}. \tag{3.2.14}$$

An alternative variance estimate, discussed in Section 3.2.4, is

$$\widehat{\text{Var}}[\hat{H}(t)] = \sum_{j:t_j \le t} \frac{d_j}{n_j^2}. \tag{3.2.15}$$

There is little to recommend one of (3.2.14) or (3.2.15) over the other, though (3.2.15) has somewhat smaller bias in small samples. In large samples the two estimates tend to be very close.

Both $\hat{S}(t)$ and $\hat{H}(t)$ are nonparametric m.l.e.'s, and are connected by the general relationship (1.2.16) between survivor and cumulative hazard functions. Note that $\hat{S}(t)$ and $\hat{H}(t)$ are discrete and do not satisfy the relationship $H(t) = -\log S(t)$, which holds for continuous distributions. An alternative survivor function estimate $\tilde{S}(t) = \exp[-\hat{H}(t)]$ is sometimes suggested for the continuous-time case. Conversely, an alternative estimate for $H(t)$ would be $-\log \hat{S}(t)$. Most prefer $\hat{H}(t)$ and $\hat{S}(t)$, though for probability plots described in Section 3.3 the alternatives are sometimes used.

Example 3.2.2. (Example 3.2.1 continued.) Table 3.2 shows values of the Nelson–Aalen estimate $\hat{H}(t)$ at each distinct failure time, for the Placebo group in Example 3.2.1. The values (3.2.13) are easily calculated from the n_j and d_j given in Table 3.1. Standard errors, equal to the square root of (3.2.15), are also given.

Figure 3.2 shows a plot of $\hat{H}(t)$. The plot is quite close to linear, suggesting that an exponential lifetime distribution with constant hazard function $h(t)$ would be consistent with the data.

Note that $\hat{H}(t)$ in (3.2.13) is defined so it is right continuous, whereas the product limit estimate $\hat{S}(t)$ in (3.2.2) is left continuous. The latter is consistent with the definition of $S(t)$, as $Pr(T \ge t)$, and $S(t)$ is correspondingly obtained in (1.2.16) as the product integral of $dH(u)$ over the open interval $(0, t)$. Sometimes $S(t)$ is defined

Table 3.2. Nelson–Aalen Estimate for Placebo Group

t_j	$\hat{H}(t_j)$	se	t_j	$\hat{H}(t_j)$	se
1	0.095	0.067	11	1.110	0.301
2	0.201	0.100	12	1.444	0.382
3	0.259	0.116	15	1.694	0.457
4	0.384	0.146	17	2.027	0.565
5	0.527	0.178	22	2.527	0.755
8	0.860	0.244	23	3.527	1.253

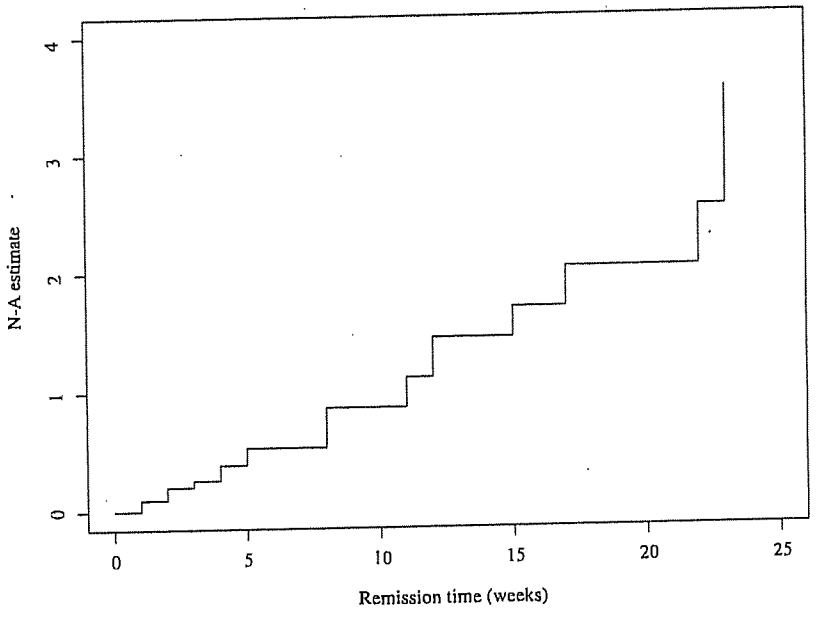

Figure 3.2. Nelson–Aalen cumulative hazard function estimate for Placebo group.

as $Pr(T > t)$; if that is the case, then the product integrals in (1.2.16) and in the estimate (3.2.11) are over the interval $(0, t]$, and the sums in (3.2.2) and (3.2.3) are over all times $t_j \le t$.

3.2.3 Interval Estimation of Survival Probabilities or Quantiles

Nonparametric methods can also be used to construct confidence intervals for lifetime distribution characteristics. In practice, survival probabilities $S(t)$ and quantiles t_p are of the most interest. The confidence intervals below are based on the same types of approximate pivotal quantities as for parametric models (see Section 2.5 and Appendix C).

3.2.3.1 Confidence Intervals for Survival Probabilities

Confidence intervals for the survivor function $S(t)$ at a specified value t can be constructed from right-censored data in a variety of ways. The most straightforward is to use the fact, discussed in Section 3.2.4, that if $\hat{S}(t)$ is the PL estimator, then $\sqrt{n}(\hat{S}(t) - S(t))$ is asymptotically normal under mild conditions. More specifically,

$$Z_1 = \frac{\hat{S}(t) - S(t)}{\hat{\sigma}_S(t)} \tag{3.2.16}$$

is approximately $N(0, 1)$, where $\hat{\sigma}_S(t)^2 = \widehat{\text{Var}}[\hat{S}(t)]$ is the Greenwood variance estimate (3.2.3). We can use Z_1 as an approximate pivotal quantity and obtain α confidence intervals by inverting probability statements of the form $Pr(a \le Z_1 \le b) = \alpha$. The choice $b = -a = z_{.5(1+\alpha)}$, where z_p is the pth quantile for the standard normal distribution, gives the approximate α confidence interval

$$\hat{S}(t) - z_{.5(1+\alpha)}\hat{\sigma}_S(t) \le S(t) \le \hat{S}(t) + z_{.5(1+\alpha)}\hat{\sigma}_S(t). \tag{3.2.17}$$

The distribution of Z_1 may not be well approximated by $N(0, 1)$ when the number of uncensored lifetimes is small or when $S(t)$ is close to 0 or 1, and (3.2.17) may even include values outside of the interval (0, 1). A procedure that gives admissible confidence intervals and coverage probabilities closer to the stated nominal value is to consider one-to-one functions $\psi(t) = g[S(t)]$, which take values on $(-\infty, \infty)$. The m.l.e. of the transformed parameter $\psi(t)$ is $\hat{\psi}(t) = g[\hat{S}(t)]$, and its asymptotic variance is estimated through the asymptotic variance formula (B4) by

$$\hat{\sigma}_{\hat{\psi}}(t)^2 = \{g'[\hat{S}(t)]\}^2\widehat{\text{Var}}[\hat{S}(t)]. \tag{3.2.18}$$

There are several choices of function $g(s)$ for which the approximate pivotal quantity

$$Z_2 = \frac{\hat{\psi}(t) - \psi(t)}{\hat{\sigma}_{\hat{\psi}}(t)} \tag{3.2.19}$$

is closer to standard normal than Z_1, and gives better performing confidence intervals. Two of these that are often used are the logit transformation $\psi(s) = \log((1 - s)/s)$ and the log-log transformation $\psi(s) = \log(-\log s)$.

Confidence intervals for $\psi(t)$ can be obtained by treating Z_2 as standard normal, and the resulting interval can then be transformed to an interval for $S(t)$. For example, with the transformation $\psi(t) = \log[-\log S(t)]$, the inverse transformation is $S(t) = \exp(-e^{\psi(t)})$ and the interval $\psi_L \le \psi(t) \le \psi_U$ transforms to

$$\exp(-e^{\psi_U}) \le S(t) \le \exp(-e^{\psi_L}). \tag{3.2.20}$$

In parametric models it is often found that confidence intervals obtained by using likelihood ratio statistics have close to nominal coverage in small samples, even when intervals obtained from Wald statistics (see Appendix C) like (3.2.16) do not. Interestingly, likelihood ratio methods can also be applied to the current nonparametric setting, as shown originally by Thomas and Grunkemeier (1975). This approach is now often referred to as an empirical likelihood procedure (Owen 2001). For the discrete-time case, one considers the ratio of the likelihood function (3.2.5) maximized unconditionally and under $H_0 : S(t) = s_0$; this is appropriate for testing H_0 versus $H_1 : S(t) \ne s_0$. From the arguments that follow (3.2.6), the overall maximized likelihood function is

$$L_1 = \prod_{j=1}^{k}(1 - \hat{p}_j)^{d_j}\hat{p}_j^{n_j - d_j},$$

where for convenience we define $\hat{p}_j = 1 - \hat{h}(t_j) = 1 - d_j/n_j$, with the $t_j (j = 1, \ldots, k)$ representing the distinct observed failure times. To consider the maximum of the likelihood under H_0, note that $S(t) = s_0$ in (3.2.5) implies that we must have $\tilde{S}(t_\ell+) = \tilde{S}(t_{\ell+1}) = s_0$, where ℓ is such that $t \in (t_\ell, t_{\ell+1}]$. This implies that to maximize the likelihood subject to $S(t) = s_0$ it is necessary to maximize

$$L = \prod_{j=1}^{k}(1 - p_j)^{d_j}p_j^{n_j - d_j},$$

subject to the restriction $p_1 \cdots p_\ell = s_0$. To do this we use a Lagrange multiplier λ and consider

$$\log L + \lambda\left(\sum_{i=1}^{\ell}\log p_i - \log s_0\right) = \sum_{j=1}^{k} d_j \log(1 - p_j) + (n_j - d_j)\log p_j + \lambda\left(\sum_{i=1}^{\ell}\log p_i - \log s_0\right).$$

Setting derivatives with respect to each of $p_1, \ldots, p_k$ equal to zero, we find the constrained m.l.e.'s under H_0 to be

$$\tilde{p}_j = 1 - \frac{d_j}{n_j + \lambda} \qquad j = 1, \ldots, \ell$$

$$\tilde{p}_j = 1 - \frac{d_j}{n_j} \qquad j = \ell + 1, \ldots, k,$$

where $\lambda = \lambda(s_0)$ satisfies

$$\prod_{i=1}^{\ell}\tilde{p}_i = \prod_{i=1}^{\ell}\left(1 - \frac{d_i}{n_i + \lambda}\right) = s_0. \tag{3.2.21}$$

The maximum of the likelihood under $H_0 : S(t) = s_0$ is thus

$$L_2 = \prod_{j=1}^{k}(1 - \tilde{p}_j)^{d_j}\tilde{p}_j^{n_j - d_j},$$

and the likelihood ratio statistic for testing H_0 is

$$\begin{aligned}\Lambda &= -2(\log L_2 - \log L_1)\\ &= -2\left[\sum_{j=1}^{k}(n_j - d_j)\log\left(\frac{\tilde{p}_j}{\hat{p}_j}\right) + d_j\log\left(\frac{1 - \tilde{p}_j}{1 - \hat{p}_j}\right)\right]\\ &= 2\sum_{j=1}^{\ell}\left[n_j\log\left(1 + \frac{\lambda}{n_j}\right) - (n_j - d_j)\log\left(1 + \frac{\lambda}{n_j - d_j}\right)\right].\end{aligned} \tag{3.2.22}$$

An α confidence interval for $S(t)$ is given by the set of all s_0 such that $\Lambda \le \chi^2_{(1),\alpha}$. This can be obtained by finding the set of all λ in (3.2.22) that make $\Lambda \le \chi^2_{(1),\alpha}$ and then obtaining the corresponding set of s_0 values from (3.2.21). The sets of λ values making $\Lambda \le \chi^2_{(1),\alpha}$ are closed intervals $[\lambda_L, \lambda_U]$ such that $\lambda_L < 0 < \lambda_U$ unless $\hat{S}(t) = 0$, in which case $0 = \lambda_L < \lambda_U$. Because s_0 in (3.2.21) is an increasing function of λ, the confidence intervals for $s = S(t)$ are thus of the form $[s_L, s_U]$, with $0 < s_L < s_U < 1$, unless $\hat{S}(t) = 0$, in which case $s_L = 0 < s_U < 1$. Note that

$$s_L = \prod_{i=1}^{\ell}\left(1 - \frac{d_i}{n_i + \lambda_L}\right) \quad \text{and} \quad s_U = \prod_{i=1}^{\ell}\left(1 - \frac{d_i}{n_i + \lambda_U}\right). \tag{3.2.23}$$

Nonparametric bootstrap methodology can also be applied to any of the (approximate) pivotal quantities considered here (see Appendix D.2). Bootstrap samples $(t_i^*, \delta_i^*), i = 1, \ldots, n$ are generated by sampling with replacement from $\{(t_i, \delta_i), i = 1, \ldots, n\}$. Each bootstrap sample produces a value for quantities such as (3.2.16) or (3.2.19), as described in Appendix D.2, and a set of B bootstrap samples can be used to estimate the distribution of the pivotal quantities. Except possibly for quite small samples, this doesn't usually improve much on the use of (3.2.19) with a normal approximation.

Example 3.2.3. Example 1.1.7 gave remission duration times (in weeks) for two groups of leukemia patients in a clinical trial. The data were discussed in Examples 3.2.1 and 3.2.2, where Kaplan–Meier and Nelson–Aalen estimates were shown for the two groups. We now obtain confidence limits for the survival function at times 10 and 20 weeks; because there is a failure at 10 weeks and a censoring time at 20 weeks in the treatment (drug 6-MP) group, to avoid ambiguity we will consider confidence intervals for $S(t+)$ for $t = 10, 20$.

Table 3.3 shows approximate .95 confidence intervals obtained by the following methods, described earlier:

Table 3.3. 0.95 Confidence Intervals for $S(10+)$, $S(20+)$

Group	Method	$S(10+)$	$S(20+)$
Placebo	$Z_1(1)$	(.17, .59)	(0, .22)
	$Z_2(2)$	(.18, .58)	(.016, .26)
	$Z_2(3)$	(.20, .60)	(.024, .31)
	LR	(.20, .59)	(.016, .27)
	Exact	(.18, .63)	(.012, .30)
Drug 6-MP	$Z_1(1)$	(.56, .94)	(.40, .85)
	$Z_2(2)$	(.50, .89)	(.37, .81)
	$Z_2(3)$	(.53, .89)	(.39, .82)
	LR	(.54, .90)	(.40, .82)

1. Formula (3.2.17) with $\alpha = .95$. To illustrate, from Example 3.2.1 we have for the Placebo group that $\hat{S}(10+) = .381$ and $\hat{\sigma}_s(10+) = .106$; then (3.2.17) with $z_{.975} = 1.96$ gives the interval (.17, .59).
2. The approximate pivotal quantity (3.2.19) with $\psi(t) = \log[-\log S(t)]$ treated as standard normal. By (3.2.18) the standard error of $\hat{\psi}(t)$ is obtained from

$$\hat{\sigma}_\psi(t)^2 = \frac{\hat{\sigma}_s(t)^2}{[\hat{S}(t)\log\hat{S}(t)]^2},$$

   and the approximate .95 confidence interval is given by $\hat{\psi}(t) \pm 1.96\hat{\sigma}_\psi(t)$. To illustrate, in the Placebo group we get $\hat{\psi}(10+) = -.0357$ and $\hat{\sigma}_\psi(10+) = .288$, giving the confidence interval $-.601 \le \psi(10+) \le -.529$. By (3.2.20), this converts to the interval $.18 \le S(10+) \le .58$.
3. The approximate pivotal (3.2.19) with $\psi(t) = \log[S(t)/(1 - S(t))]$.
4. The empirical likelihood ratio procedure based on (3.2.22) and $\chi^2_{(1),0.95} = 3.84$. For $S(10+)$ the set of λ values satisfying $\Lambda \le 3.84$ is easily found by graphing $\Lambda(\lambda)$ or by iterative calculation to be $-4.83 \le \lambda \le 10.97$. By (3.2.23) this gives the confidence interval for $S(10+)$ for the Placebo group, for example, as (0.196, 0.593).

Full results are shown in Table 3.3. A plot of the empirical likelihood ratio statistics $\Lambda(s_0)$ for $s_0 = S(10+)$ is shown for the two groups in Figure 3.3. This provides a concise picture of the information about s_0, and in particular shows confidence intervals for any nominal coverage probability.

The Placebo group has no censored observations, and exact confidence limits are also provided in this case. In general exact limits can be obtained for $S(t)$ whenever there are no censored observations by time t. (The term "exact" here means that an exact distribution is used, not that the confidence interval coverage is exactly α.) The limits are obtained by inverting the hypothesis test $H_0 : S(t) = s_0$, which is based on the fact that the number of lifetimes X exceeding t has a binomial (n, s_0) distribution under H_0. A lower α confidence limit for $S(t)$ is found as the set of all values s_0 such that $Pr(X \ge x_0; s_0) \ge 1 - \alpha$, where x_0 is the observed value of X. The desired set of values is of the form $(s_L, 1)$, and it can be shown that

$$s_L = \frac{x_0}{x_0 + (n - x_0 + 1)F_{(2(n-x_0+1),2x_0),\alpha}} \tag{3.2.24}$$

by using the relationship between the binomial and F distribution (e.g., Johnson et al. 1995, Chs. 25 and 27). Upper confidence limits can be found in a similar way; the upper α limit is

$$s_U = \frac{x_0 + 1}{(x_0 + 1) + (n - x_0)F_{(2(n-x_0),2x_0+2),1-\alpha}}. \tag{3.2.25}$$

The sample sizes are small here, and the .95 confidence intervals are correspondingly wide. Except for the interval for $S(20+)$ in the Placebo group that is based

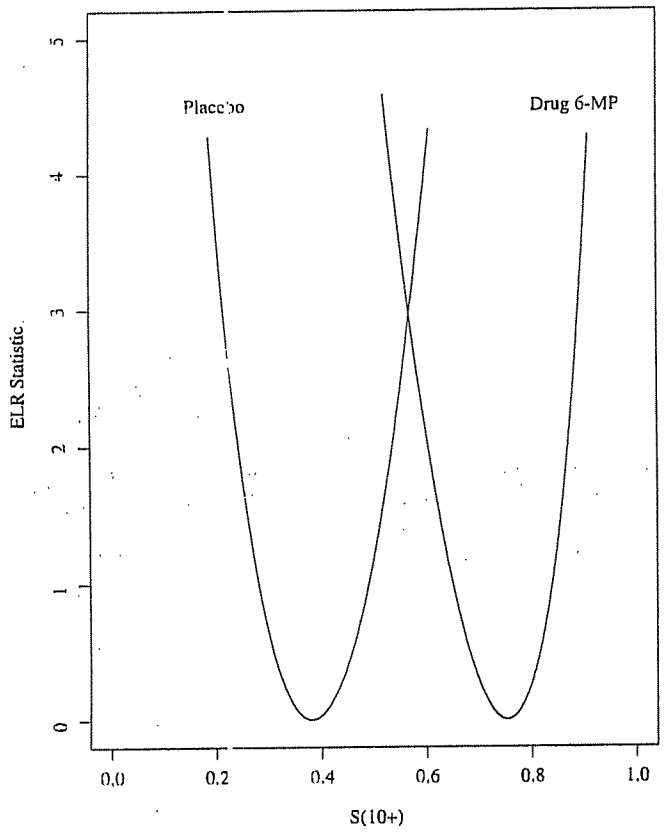

Figure 3.3. Empirical likelihood ratio statistics for $S(10+)$.

on Z_1, the various intervals agree well, and would not lead to any conflicting conclusions. The intervals based on Z_2 agree more closely with those based on the LR method than do the intervals based on Z_1.

It is sometimes useful to plot Kaplan–Meier estimates $\hat{S}(t)$ along with bands that show "pointwise" confidence intervals for all values of t. Such bands are parallel to the estimate $\hat{S}(t)$, since both the estimate and the confidence limits described in this section change values only at observed failure times. The following example illustrates this.

Example 3.2.4. The distribution $S(t)$ of time to first pulmonary exacerbations for patients in a randomized clinical trial was discussed in Example 1.1.8. There

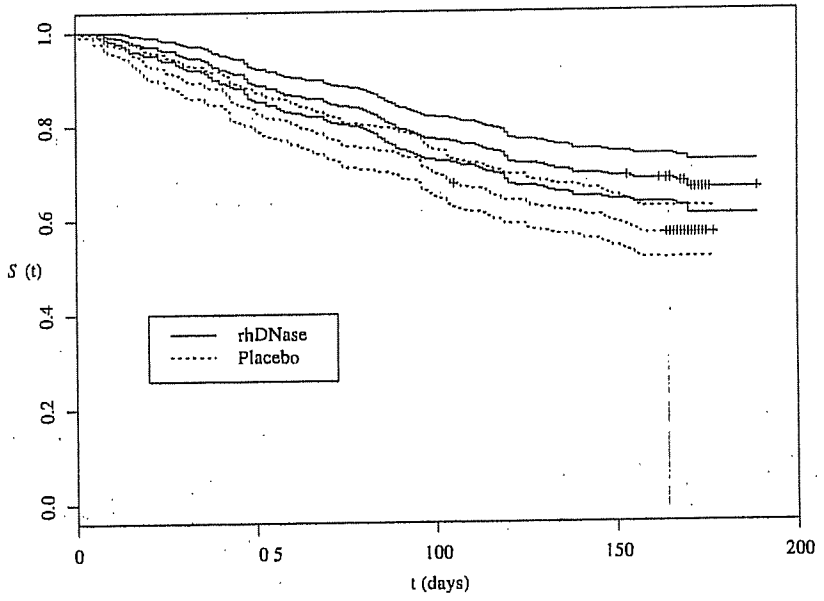

Figure 3.4. KM estimates of $S(t)$ and pointwise confidence limits for time to first exacerbation.

were two treatment groups: rhDNase and Placebo. Figure 3.4 shows Kaplan–Meier estimates for the two groups along with pointwise .95 confidence intervals for $S(t)$, obtained using Z_2 with the log-log transformation in (2) of Example 3.2.3. These estimates ignore the baseline forced expiratory volume (fev) covariate, but since the treatment assignment is random, they provide an unbiased view of the lifetime distributions in the population of patients that the study represents.

The rhDNase treatment yields a substantially higher probability $\hat{S}(t)$ of survival without an exacerbation, though there is some overlap of .95 confidence intervals with those for the Placebo group. Hypothesis tests of the equality of two survivor functions are considered in Sections 7.2 and 8.1.

Plots of Kaplan–Meier estimates implicitly show observed lifetimes, since $\hat{S}(t)$ drops at each distinct time. Software for Kaplan–Meier estimation usually provides the option of showing censoring times in the plots. This has been utilized in Figure 3.4, where the symbol + indicates censoring times.

3.2.3.2 *Confidence Intervals for Quantiles*

Estimation of the mean of a distribution is useful in many contexts, but for lifetime distributions the quantiles t_p of the distribution are usually of more interest. The median, or $t_{.50}$, is often used as a measure of location or "central tendency." Two advantages it has over the mean are that it always exists (assuming $S(\infty) < .5$), whereas the mean may not, and that it is easier to estimate when data are censored.

Nonparametric point estimates of t_p can be defined in various ways, with an essential complication being that $\hat{S}(t)$ is a step function, so that for some values of p, there is an interval of t-values satisfying $\hat{S}(t) = 1-p$. However, for most values of p there is one t-value (equal to one of the observed failure times), and it is common to take this as the point estimate $\hat{t}_p$.

Point estimates of t_p are usually of less interest than interval estimates. Approximate α confidence intervals for t_p are most easily obtained by inverting the relationship $S(t_p) = 1 - p$ between the survivor function and the quantiles of a distribution. Thus, if $t_L(\text{Data})$ is a lower confidence limit for t_p, we note that

$$Pr(t_L(\text{Data}) \leq t_p) = Pr(S(t_L(\text{Data})) \geq 1 - p).$$

Therefore, if we want a lower α confidence limit for t_p based on observed data, then we can obtain this by finding the value t_L such that $1-p$ is a lower α confidence limit for $S(t_L)$, based on the data. In other words, if $s_L(\text{data}; t)$ is a lower α confidence limit for $S(t)$, then the lower α confidence limit for t_p is obtained by finding t such that $s_L(\text{data}; t) = 1 - p$.

To illustrate the procedure, suppose we want a confidence interval for $t_{.5}$. If we base confidence intervals for $S(t)$ on the approximate pivotal quantity (3.2.16), then to get a two-sided .95 confidence interval for $t_{.50}$, we find the set of t-values satisfying $-1.96 \leq Z_1 \leq 1.96$, where

$$Z_1 = \frac{\hat{S}(t) - .5}{\hat{\sigma}_S(t)}. \tag{3.2.26}$$

Because $\hat{S}(t)$ is a step function, there will not in general be a value of t making Z_1 exactly equal to -1.96 or 1.96, so we take the failure times at which the value of Z_1 changes from being outside of $(-1.96, 1.96)$ to inside $(-1.96, 1.96)$; note that as t varies, Z_1 changes value only at the observed failure times.

There is a convenient graphical method for determining an α confidence interval for t_p. Consider the graph of $\hat{S}(t)$ along with the bands giving pointwise α confidence intervals for $S(t)$ discussed earlier; these bands are step functions parallel to $\hat{S}(t)$. To find confidence limits for t_p we simply find where the bands intersect the line $S(t) = 1 - p$; this identifies the failure times that specify the confidence interval for t_p. To do this we plot the vertical pieces of the step functions for $\hat{S}(t)$ and the bands, as shown in Figure 3.4. In the unlikely event that the line $S(t) = 1 - p$ coincides with a horizontal step of one of the confidence bands, we use the average of the failure times at either end of the step as the confidence limit for t_p.

Example 3.2.5. Consider confidence intervals for quantiles in the case of the pulmonary exacerbation time data discussed in Example 3.2.4.

As an illustration, let us obtain a two-sided .95 confidence interval for the .20 quantile, $t_{.20}$, for the rhDNase and Placebo groups. Examination of Figure 3.4 according to the graphical method described earlier (with $p = .20$) indicates that the confidence intervals for $t_{.20}$ are approximately (75, 115) days for the rhDNase population and (50, 80) days for the Placebo population. By examining the pivotal quantity (3.2.26) for different t-values, we find the exact intervals to be (73, 112) for rhDNase and (49, 79) for Placebo.

Note that neither estimate of $S(t)$ drops much below .60, so that confidence intervals for quantiles such as the median are not available.

3.2.4 Asymptotic Properties of Estimators

The PL estimate $\hat{S}(t)$ and Nelson–Aalen estimate $\hat{H}(t)$ possess desirable large-sample properties under the assumptions about the censoring process made in Section 2.2.2, among them consistency and asymptotic normality. Similar results hold under independent delayed entry. In the discrete-time setting described in Section 3.2.1, it is relatively easy to derive asymptotic results. Early treatments of the continuous-time setting (e.g., Breslow and Crowley 1974) worked from a random independent censoring model and a discretization of the time scale, using limiting arguments to get continuous-time results. Starting with Aalen (1976, 1978a), however, counting processes and martingale theory were deployed to provide elegant and more general treatments. Authoritative and very detailed accounts of the theory are given by Fleming and Harrington (1991) and Andersen et al. (1993). In this section we outline some of the main ideas; martingales and counting processes are reviewed in Appendix F.

We use terminology and notation introduced in Section 2.2.2. Let $S^0(t) = Pr(Y_i(t) > 0)$ denote the probability an individual is alive and uncensored at time t; this probability depends on the lifetime distribution and the censoring process. As earlier, $Y_\cdot(t) = \sum Y_i(t)$, and for convenience we also define $J(t) = I(Y_\cdot(t) > 0)$ with the understanding that $J(t)/Y_\cdot(t) = 0$ when $Y_\cdot(t) = 0$. The counting-process–martingale development uses the fact that, under the assumptions about the censoring process in Section 2.2.2,

$$dM_i(t) = dN_i(t) - Y_i(t)\,dH(t)$$

are martingale increments satisfying $E\{dM_i(t)|\mathcal{H}(t)\} = 0$. Looking first at the Nelson–Aalen estimator (3.2.12) and defining a process

$$H^*(t) = \int_0^t J(u)\,dH(u),$$

we see that for data based on n independent individuals,

$$\begin{aligned}\hat{H}(t) - H^*(t) &= \int_0^t \frac{J(u)}{Y_\cdot(u)} \left\{ \sum_{i=1}^n dN_i(u) - \sum_{i=1}^n Y_i(u)\,dH(u) \right\} \\ &= \int_0^t \frac{J(u)}{Y_\cdot(u)} \sum_{i=1}^n dM_i(u). \end{aligned} \tag{3.2.27}$$

In (3.2.27) we have for convenience defined $d\hat{H}(u) = 0$ when $Y_\cdot(u) = 0$.

The representation (3.2.27) immediately shows that $E\{\hat{H}(t) - H^*(t)\} = 0$, and that

$$E\{\hat{H}(t) - H(t)\} = E\{H^*(t)\} - H(t)$$
$$= -\int_0^t Pr(Y.(u) = 0)\, dH(u), \tag{3.2.28}$$

which $\to 0$ as $Pr(Y.(u) = 0) \to 0$ over $(0, t]$. Furthermore, standard martingale calculations give that

$$\mathrm{Var}\{\sqrt{n}[\hat{H}(t) - H^*(t)]\} = E\int_0^t \frac{nJ(u)}{Y.(u)}[1 - \Delta H(u)]\, dH(u), \tag{3.2.29}$$

where $\Delta H(u) = H(u) - H(u-)$. This is easily shown directly from (3.2.27) in the discrete-time setting; the general case (see Appendix F and Fleming and Harrington 1991, p. 92ff) requires some additional machinery. A key ingredient in either development is the fact that the $dM_i(u)$'s have mean 0 and are orthogonal (uncorrelated) for distinct values u, u'.

Under the assumption that $Pr(Y_i(u) > 0) > 0$ for $0 \le u \le t$, it follows from (3.2.28) and (3.2.29) that

$$\sigma^2_{NA}(t) = \mathrm{Asvar}\{\sqrt{n}[\hat{H}(t) - H(t)]\}$$
$$= \int_0^t \frac{[1 - \Delta H(u)]}{S^0(u)}\, dH(u). \tag{3.2.30}$$

Central limit theory for martingales shows that $\sqrt{n}[\hat{H}(t) - H(t)]$ also has a limiting normal distribution. Inserting the estimate $\hat{H}(u)$ in (3.2.30) and estimating $S^0(u)$ by $Y.(u)/n$, we get the variance estimate

$$\hat{\sigma}^2_{NA}(t) = n\sum_{j:t_j \le t} \frac{d_j(n_j - d_j)}{n_j^3}, \tag{3.2.31}$$

which is the same as (3.2.14). An alternative estimate is frequently used when $H(t)$ and $S(t)$ are continuous functions. In that case $\Delta H(u) = 0$ and (3.2.30) can be rewritten as

$$\sigma^2_{NA}(t) = \int_0^t \frac{dH(u)}{S^0(u)},$$

which gives the estimate in (3.2.15),

$$\hat{\sigma}^2_{NA}(t) = n\sum_{j:t_j \le t} \frac{d_j}{n_j^2}. \tag{3.2.32}$$

The Kaplan–Meier (PL) estimate $\hat{S}(t)$ can be handled via martingale theory by noting that, in general,

$$S(t) = 1 - Pr(T < t)$$
$$= 1 - \int_0^{t-} S(u)\, dH(u)$$

and

$$\hat{S}(t) = 1 - \int_0^{t-} \hat{S}(u) d\hat{H}(u).$$

This gives a representation of $\hat{S}(t) - S(t)$ as a stochastic integral of a martingale (see Appendix F), to which standard theory can be applied. This leads to a proof of asymptotic normality and a result analogous to (3.2.30),

$$\sigma^2_{KM}(t) = \mathrm{Asvar}\{\sqrt{n}[\hat{S}(t) - S(t)]\}$$
$$= S(t)^2 \int_0^{t-} \frac{[-dS(u)]}{S(u+)S^0(u)}, \tag{3.2.33}$$

where, because $S(u)$ is defined to be right continuous, $dS(u) = S(u+) - S(u)$ if $S(u)$ jumps at u. When $\hat{S}$ is inserted for S and $Y.(u)/n$ is inserted for $S^0(u)$ in (3.2.33), we get the variance estimate

$$\hat{\sigma}^2_{KM}(t) = n\hat{S}(t)^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)}. \tag{3.2.34}$$

This is the same as the Greenwood variance estimate (3.2.3).

Stronger asymptotic results can also be derived. For example, in the continuous-time case, if τ is a value such that $S^0(t) > 0$ for $0 \le t \le \tau$, then the random processes $\{\sqrt{n}[\hat{H}(t) - H(t)], 0 < t \le \tau\}$ and $\{\sqrt{n}[\hat{S}(t) - S(t)], 0 < t \le \tau\}$ converge weakly to mean zero Gaussian processes with respective covariance functions

$$\sigma_{NA}(t, t') = \int_0^{\min(t,t')} \frac{1}{S^0(u)}\, dH(u) \tag{3.2.35}$$

and

$$\sigma_{KM}(t, t') = S(t)S(t') \int_0^{\min(t,t')-} \frac{1}{S(u+)S^0(u)}[-dS(u)]. \tag{3.2.36}$$

The results concerning Gaussian limiting processes enable the construction of confidence bands for $S(t)$ or $H(t)$, and estimation of quantities that are functionals

of $S(t)$. Two such are the mean lifetime, which can be represented as $\mu = \int_0^\infty S(t)\,dt$ (see Problem 1.1, part (a)), and more generally, $\mu_\tau = E\{\min(T, \tau)\}$, which is called the mean lifetime restricted to τ. This is given by

$$\mu_\tau = \int_0^\tau S(t)\,dt \tag{3.2.37}$$

and can be estimated by replacing $S(t)$ with the PL estimate $\hat{S}(t)$. The results concerning the limiting Gaussian process for $\sqrt{n}[\hat{S}(t) - S(t)]$ allow it to be shown that $\sqrt{n}(\hat{\mu}_\tau - \mu_\tau)$ is asymptotically normal, with variance given by

$$\int_0^\tau \frac{A(t)^2 |dS(t)|}{S(t)S^0(t)}, \tag{3.2.38}$$

where $A(t) = \int_t^\tau S(u)\,du$. The derivation of this result and an estimate of variance based on it are discussed in Problem 3.6.

3.3 DESCRIPTIVE AND DIAGNOSTIC PLOTS

3.3.1 Plots Involving Survivor or Cumulative Hazard Functions

Plots of PL or Nelson–Aalen estimates provide good descriptions of univariate lifetime data. They can also be employed to assess the appropriateness of a parametric model, as we now discuss. Graphical assessments are subjective but useful; they can be supplemented with formal goodness-of-fit tests, considered in Chapter 10.

3.3.1.1 Plots of Survivor Functions

Suppose that a parametric model has survivor function $S(t; \boldsymbol{\theta})$ and distribution function $F(t; \boldsymbol{\theta})$, and let $\hat{\boldsymbol{\theta}}$ be an estimate obtained from a specific data set. If the parametric family is appropriate, then $S(t; \hat{\boldsymbol{\theta}})$ or $F(t; \hat{\boldsymbol{\theta}})$ should not differ too much from nonparametric estimates of $S(t)$ or $F(t)$. The simplest model assessment procedure is simply to plot $S(t; \hat{\boldsymbol{\theta}})$ and the PL estimate $\hat{S}(t)$ on the same graph; alternatively, the corresponding distribution functions can be plotted. The sampling variability in the two estimates must be kept in mind, and nonparametric confidence limits for $S(t)$ as described in Section 3.2.3 are often a useful addition to the plot.

Example 3.3.1. The data below were given by Thoman et al. (1969), who attributed them to tests on the endurance of deep-groove ball bearings discussed by Lieblein and Zelen (1956). Coroni (2002) has noted that they are not the same as the original data, which involved some censored observations. However, for illustrative purposes we will treat them as an uncensored sample. The observations are the number of million revolutions before failure for each of 23 ball bearings; the individual bearings were inspected periodically to determine whether "failure" had occurred, but we treat the failure times as continuous. The 23 failure times are

17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.40, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.40.

Figure 3.5 shows plots of the Kaplan–Meier estimate of the survivor function, $S(t)$, along with estimates $S(t; \hat{\boldsymbol{\theta}})$ from Weibull and log-normal distributions that were fitted to the data. With the Weibull model in the form (1.3.6), the m.l.e.'s (see Section 5.2.1) are $\hat{\lambda} = 0.0122$, $\hat{\beta} = 2.10$; the log-normal model (1.3.10) has m.l.e.'s (see Section 5.3.1) $\hat{\mu} = 4.15$, $\hat{\sigma} = 0.522$. The plots indicate good agreement between the nonparametric $\hat{S}(t)$ and both the Weibull and log-normal models; the log-normal fits the data slightly better.

The plot in Example 3.3.1 indicated how well a parametric model fitted the data. It can be supplemented by formal goodness-of-fit tests, which are described in Chapter 10. Other types of plots are also useful, especially ones that extend easily to models involving fixed or time-varying covariates. The plots that we describe now also compare nonparametric and model-based estimates of distributions, but are designed to be roughly linear when the parametric model is appropriate. These plots are less directly descriptive of the data, but emphasize systematic differences between $\hat{S}(t)$ and $S(t; \hat{\boldsymbol{\theta}})$.

3.3.1.2 Probability Plots

One important type of plot is the *P-P* (probability-probability) plot, which is essentially a plot of points $(S(t_j; \hat{\boldsymbol{\theta}}), \hat{S}(t_j))$, where $t_1 < t_2 < \cdots < t_k$ are the distinct times at which failures occur in the data. Thus the model-based and empirical survivor functions are compared at the failure times, and if the parametric model is appropriate the points should lie around a straight line with slope one. A common variation of this procedure when $S(t; \hat{\boldsymbol{\theta}})$ is continuous in t is to replace $\hat{S}(t_j)$ with

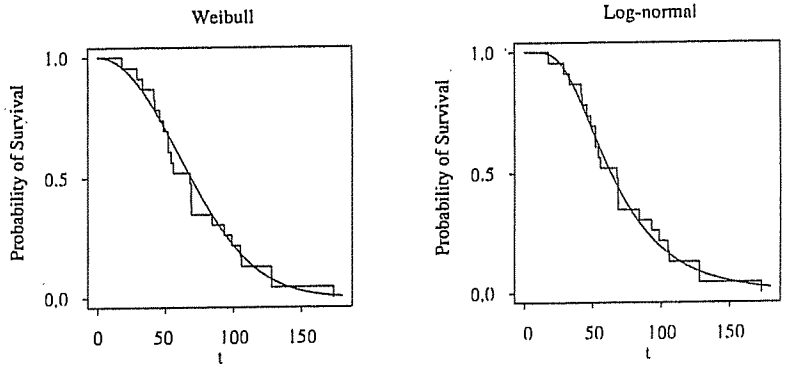

Figure 3.5. KM and parametric estimates of survival for ball-bearing data.

the value

$$\hat{S}_j^* = .5\hat{S}(t_j) + .5\hat{S}(t_j+), \tag{3.3.1}$$

since $\hat{S}(t)$ is a step function with jumps at the t_j. This is also a more convenient choice for other procedures, described below.

A very similar procedure is the Q-Q (quantile-quantile) plot of model-based versus empirical quantiles. For example, the quantile function (1.3.7) for the Weibull model is

$$t(p; \alpha, \beta) = \alpha[-\log(1-p)]^{1/\beta}.$$

A plot of the points $(t(p_j; \hat{\alpha}, \hat{\beta}), t_j)$, $j = 1, \ldots, k$, where p_j is given by $\hat{S}_j^*$ of (3.3.1), should be roughly linear if the Weibull model is appropriate.

A technique closely related to P-P and Q-Q plots is used with parametric models for which the survivor or distribution function can be "linearized." This means that some transform of $S(t; \boldsymbol{\theta})$ is a linear function of t or of some function of t, that is, $g_1[S(t; \boldsymbol{\theta})]$ is a linear function of $g_2(t)$ for some functions g_1 and g_2. The idea is then to plot $g_1(\hat{S}(t))$ versus $g_2(t)$; if the parametric family is appropriate the result should be roughly linear. This procedure has the advantage of not requiring an estimate of the parameter $\boldsymbol{\theta}$.

Suppose, for example, that the possibility of an underlying exponential distribution (1.3.2) is being considered. The survivor function satisfies

$$\log S(t) = -\lambda t, \tag{3.3.2}$$

so a plot of $\log \hat{S}(t)$ versus t should be close to a straight line through the origin if (3.3.2) is appropriate. No estimate of λ is needed for this plot. In fact, a "graphical" estimate of λ can be obtained when the plot is roughly linear by fitting a straight line through the points.

For the Weibull distribution $S(t)$ of (1.3.6) satisfies

$$\log[-\log S(t)] = \beta \log t + \beta \log \lambda.$$

Thus a plot of $\log[-\log \hat{S}(t)]$ versus $\log t$ should be roughly linear if a Weibull model is appropriate. In addition, when the plot is approximately linear, one can obtain graphical estimates of λ and β by fitting a straight line to the plot and calculating the slope and intercept: the slope is an estimate of β and the intercept on the horizontal ($\log t$) axis is an estimate of $-\log \lambda$.

The linearization procedure is applicable to models in which some transform $Y = g(T)$ of lifetime has a location-scale parameter distribution, as in Section 1.3.6. In this case Y has a survivor function of the form (assuming Y is an increasing function of T)

$$\begin{aligned} Pr(Y \geq y) &= S_0\left(\frac{y-u}{b}\right) \\ &= Pr(T \geq t) = S(t), \end{aligned}$$

where $t = g^{-1}(y)$ and $u(-\infty < u < \infty)$ and $b > 0$ are parameters. Thus

$$S_0^{-1}[S(t)] = \frac{1}{b}y - \frac{u}{b} \tag{3.3.3}$$

is a linear function of $y = g(t)$, and a plot of $S_0^{-1}[\hat{S}(t)]$ versus $g(t)$ should be roughly linear if the family of models being considered is appropriate. The Weibull and exponential distributions fall into this category. For the Weibull, $Y = \log T$ has an extreme value distribution (see Section 1.3.2) with $S_0(z) = \exp(-e^z)$, which leads to the plot of $\log[-\log \hat{S}(t)]$ versus $\log t$ previously suggested. Two approaches are possible for the exponential distribution: we can plot $\log \hat{S}(t)$ versus t, as suggested by (3.3.2) or, since the exponential is the special case of the Weibull distribution with $\beta = 1$, we can plot $\log[-\log \hat{S}(t)]$ versus $\log t$. These plots can be considered a type of Q-Q plot; note that $S_0^{-1}(1-p)$ is the pth quantile for the standard variate $(Y-u)/b$.

The result (3.3.3) holds with $y = \log t$ for all log-location-scale models (see Section 1.3.6). Besides the Weibull, this covers the log-normal and log-logistic distribution; the associated distributions of $Z = (Y-u)/b$ are the extreme value, normal, and logistic distributions. The forms of $S_0(z)$ for all three distributions are given prior to (1.3.19) in Section 1.3.6. Graphical estimates of u and b can be obtained from lines fitted to plots of $S_0^{-1}[\hat{S}(t)]$ versus $\log t$; b^{-1} is estimated by the slope and u by the y ($\log t$) intercept. Models for which $S_0(z)$ involves extra parameters, such as (1.3.20) in Section 1.3.6, can also be checked using this type of plot, provided that we first estimate the extra parameters.

Instead of plotting the step function $S_0^{-1}[\hat{S}(t)]$, we normally just plot points corresponding to the distinct observed failure times $t_1 < \cdots < t_k$, as described for P-P and Q-Q plots. When $S_0(y)$ is continuous, it is customary, instead of plotting $S_0^{-1}[\hat{S}(t_j)]$, to plot the points

$$(y_j, S_0^{-1}(\hat{S}_j^*)), \qquad j = 1, \ldots, k, \tag{3.3.4}$$

where $\hat{S}_j^*$ is given by (3.3.1) and $y_j = \log t_j$ or, more generally, $y_j = g(t_j)$.

The preceding procedures are all referred to as probability plots, although (3.3.4) is really a quantile plot. Their main use is for informal model assessment. In many cases plots indicate reasonable support for a family of models. Graphical estimates of parameters are sometimes useful for preliminary analysis or as initial values for maximum likelihood computation. A probability plot may conversely suggest a departure from an assumed model. Plots are subject to sampling variation, however, and one needs a sense of what constitutes normal variation under a given model. The examination of plots based on simulated data sets is an excellent way to develop this.

The following examples illustrate the graphical techniques.

Example 3.3.2. In Example 3.3.1 it was shown that the survivor functions for Weibull and log-normal distributions fitted to ball-bearing failure time data agreed closely with the Kaplan–Meier estimate, or empirical survivor function. An alternative way to assess visually the fit of the parametric models is via probability plots. Figure 3.6 shows plots based on (3.3.4), utilizing the fact that the Weibull and log-normal models are both log-location-scale distributions. For the Weibull model, $Y = \log T$ has an extreme value distribution for which $S_0(z) = \exp(-e^z)$ in (3.3.3). Therefore $S_0^{-1}(p) = \log(-\log p)$ and the probability plot consists of the points

$$(y_j, \log(-\log \hat{S}_j^*)) \qquad j = 1, \ldots, 23,$$

where $y_j = \log t_j$ are the log failure times and $\hat{S}_j^*$ is given by (3.3.1). On the vertical axis of the plots the values (3.3.1) are denoted by KM^*. Note that since none of the $n = 23$ failure times is censored, the Kaplan–Meier estimate (3.2.2) gives $\hat{S}(t_j+) = (n-j)/n$, so that $\hat{S}_j^* = 1 - (j - .5)/n$.

The log-normal probability plot similarly consists of the points $(y_j, \Phi^{-1}(1-\hat{S}_j^*))$, where $\Phi(z)$ is the standard normal cumulative distribution function (c.d.f.) and its inverse $\Phi^{-1}(p)$ is the corresponding quantile function. Figure 3.6 shows both plots as close to linear, though there is a slight suggestion of a bend in the Weibull plot. This corresponds to our observation in Example 3.3.1 that the log-normal model fitted the data slightly better. Simulation experiments with probability plots (see Example 3.3.4) show that this degree of nonlinearity is, however, not unusual with a correct model and a sample size of $n = 23$, and formal goodness-of-fit tests in Chapter 10 do not reject the Weibull model.

Graphical estimates of the extreme value or normal distribution parameters u and b can be obtained from the plots. A straight line drawn by eye through the normal probability plot, for example, gave a slope of approximately 1.9 and a y ($\log t$) intercept of approximately 4.2. Since (see (3.3.3)) the slope estimates b^{-1} and the intercept estimates u in the normal distribution $N(u, b^2)$ for Y, this gives graphical estimates $\tilde{u} = 4.2$ and $\tilde{b} = 0.53$. These are close to the m.l.e.'s that were given in Example 3.3.1 as $\hat{u} = 4.15$ and $\hat{b} = 0.522$.

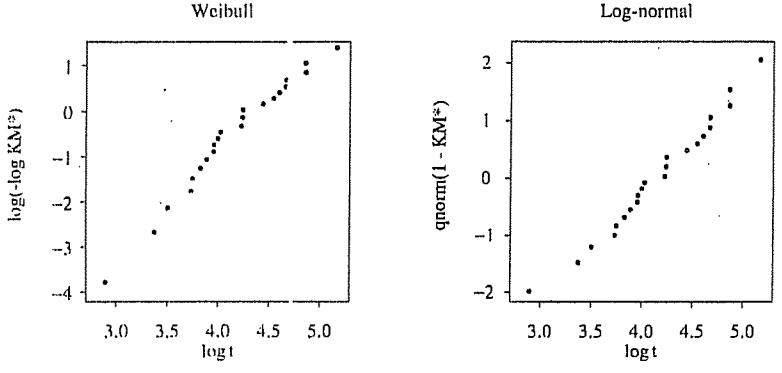

Figure 3.6. Weibull and log-normal probability plots for ball-bearing data.

Example 3.3.3. Lin et al. (1999) discussed data on patients treated for colon cancer (Moertel et al. 1990). Some of the patients later had a recurrence of the disease and may subsequently die from it. The patients took part in a randomized clinical trial in which a drug therapy (levamisole plus flourouracil) was compared with the standard treatment. There were 315 and 304 patients in the Control (standard treatment) and Therapy (drug therapy) groups, respectively. Maximum follow-up time was over 8 years.

We consider here the distribution of time T to recurrence of colon cancer, measured from the time of randomization to treatment. By the end of the study, 177 Control patients and 119 Therapy patients had experienced a recurrence. Figure 3.7 shows Kaplan–Meier estimates for the survivor functions (s.f.) $S(t)$ of T in the two treatment groups. Both a comparison of the two recurrence time distributions and estimation of the individual $S(t)$'s is of interest. It is clear that the Therapy group tends to have longer recurrence times and less recurrence; formal significance tests are considered in Example 4.5.1 and in Section 8.1. The Kaplan–Meier estimates

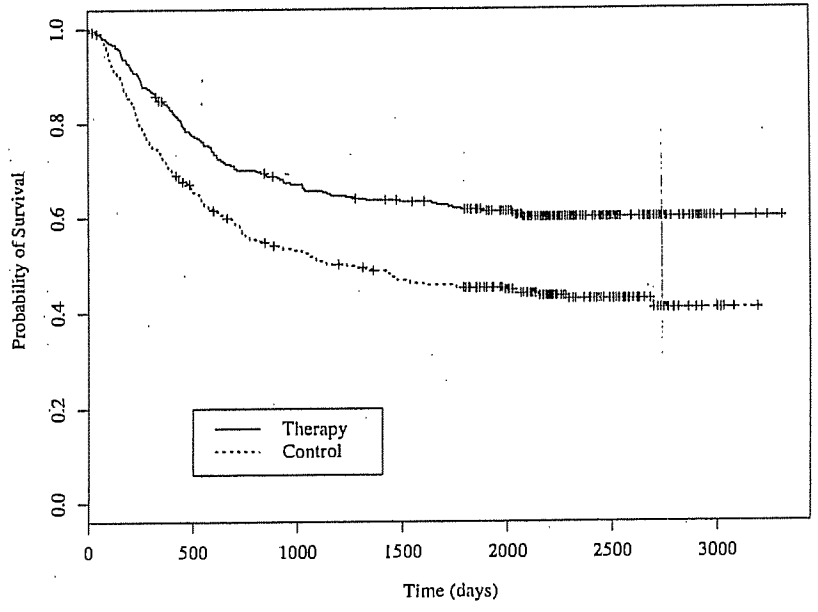

Figure 3.7. KM s.f. estimates for time to colon cancer recurrence.

suggest that the hazard function for recurrence becomes small for large t. This may indicate that some patients are cured and will not experience any disease recurrence; in that case $S(t) > 0$ for t large. If one wished to consider parametric models, then distributions such as the Weibull, log-normal, or log-logistic, for which $S(t) \to 0$ as $t \to \infty$, would then presumably be unsuitable. Figure 3.8 shows probability plots of the data for the Weibull and log-logistic distributions. For the latter the plots consist of points (3.3.4) with $y_j = \log x_j$ and $S_0^{-1}(p) = \log((1-p)/p)$. The inadequacy of the two models is confirmed. There is a convex pattern in the plots, indicating that

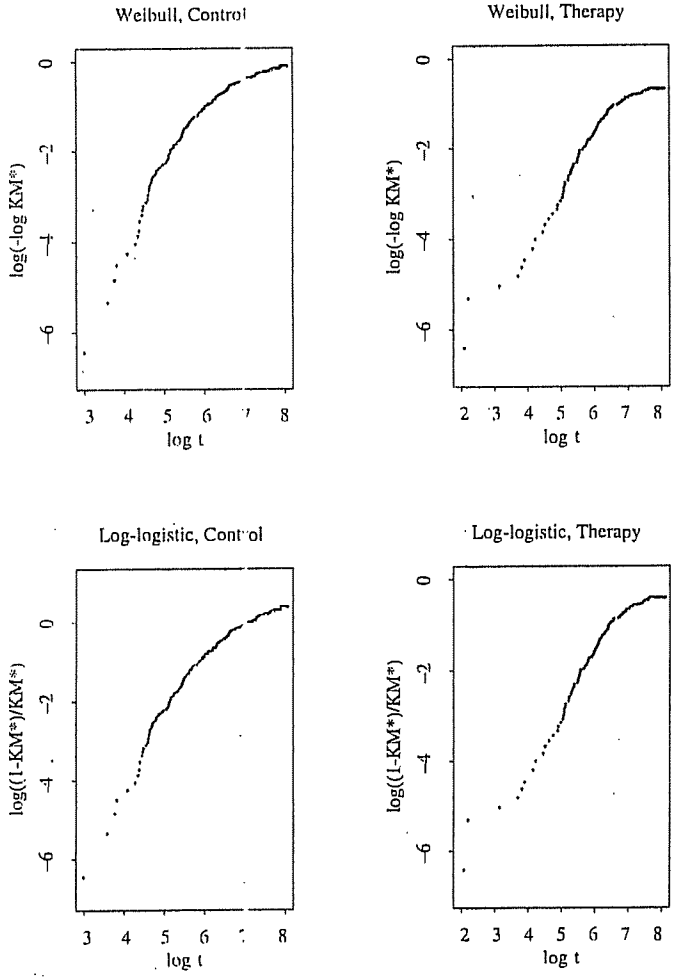

Figure 3.8. Weibull and log-logistic probability plots for cancer recurrence data.

survival times in the right tails of the two distributions are considerably larger than would be consistent with either parametric family.

In this setting, a parametric mixture model of the form (1.3.27) would perhaps fit the data, and provide an estimate of the probability of long-term nonrecurrence of disease. Such models are fitted in Example 4.5.1, and are shown to describe the data well.

Example 3.3.4. Figure 3.9 provides an indication of the variability in probability plots. It shows 15 plots, given in rows of five. Row 1 shows Weibull probability plots for five pseudorandomly generated Weibull samples of size $n = 20$; these plots consist of the points (3.3.4) with $S_0^{-1}(p) = \log(-\log p)$, as in Example 3.3.2. The five panels in row 2 show similar plots for five pseudorandom Weibull samples of size 40. The third row of panels shows Weibull probability plots for five pseudorandom samples of size 40 from a log-normal distribution.

The plots for the samples of size 40 are reasonably consistent in their patterns. Those for row 2 are quite close to linear and none suggest any evidence against the Weibull model. Those for row 3 exhibit a type of systematic curvature that suggests inadequacy of the Weibull model. However, for two of the samples (the first and third) the plots are close to linear and do not indicate any problem with the Weibull model. The plots in row 1 for the samples of size 20 are much more variable, and show that what might appear to be systematic departures from linearity are within ordinary sampling variation for the Weibull distribution.

The message in these plots is consistent with those for probability plots in many other settings. Plots based on samples of size less than 20 or so are quite variable and should be interpreted cautiously. Plots based on samples of size 40 or 50 are more reliable, but may not be very powerful in showing moderate departures from an assumed model. Censoring in the data further limits the power of such plots; the number of uncensored times is then analogous to the sample sizes in the plots of Figure 3.9.

Finally, to indicate the visual effect of the scales on the x- and y-axes in probability plots, we give in Figure 3.10 P-P plots of the two samples of size 20 represented in the first two panels of row 1 in Figure 3.9. The plots are of the points $((j-.5)/20, \hat{F}_j)$, $j = 1, \ldots, 20$, where

$$\hat{F}_j = 1 - \exp\{-(t_{(j)}/\hat{\alpha})^{\hat{\beta}}\},$$

with $t_{(1)} < \cdots < t_{(20)}$ the ordered failure times for the sample, and $\hat{\alpha}$, $\hat{\beta}$ the Weibull m.l.e. (see Section 5.2.1). The patterns in Figure 3.10 are similar to those for the first two panels in row 1 of Figure 3.9, but the compression of the tails of the distribution in the P-P plot alters the visual effect. In particular, the pattern for sample 2 appears more severe in the P-P plots than in the quantile plot of Figure 3.9. A point related to the scaling of the axes is that the pattern of variation as we move from the left to right tail of the distribution varies according to the type of plot. In particular, for P-P plots the variation is smaller in the two tails than in the middle of the distribution, whereas it is usually more stable for Q-Q plots.

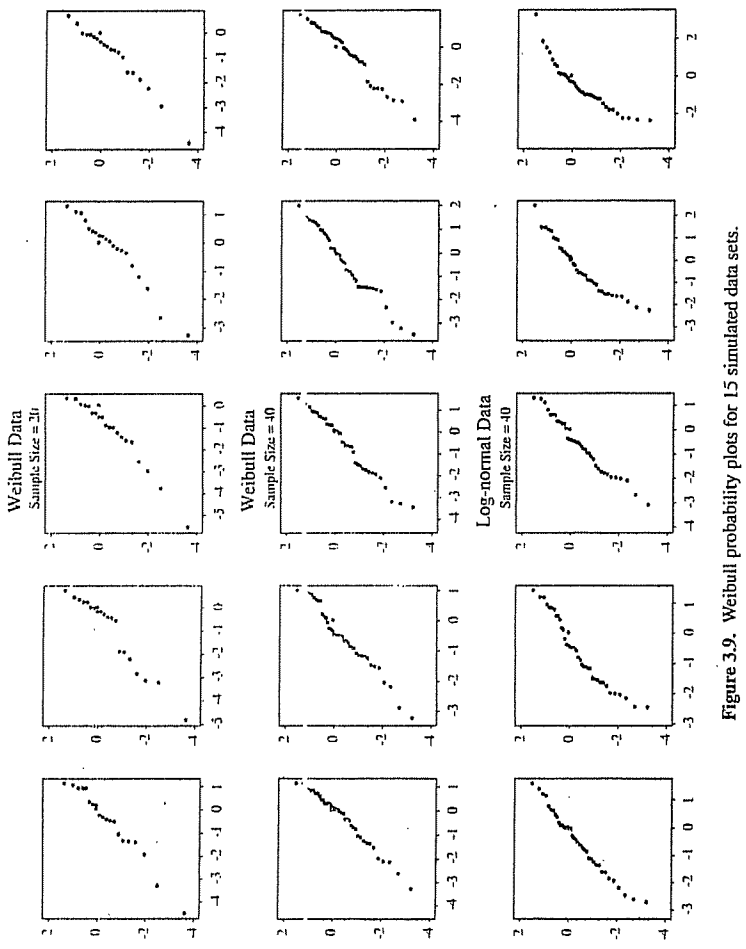

Figure 3.9. Weibull probability plots for 15 simulated data sets.

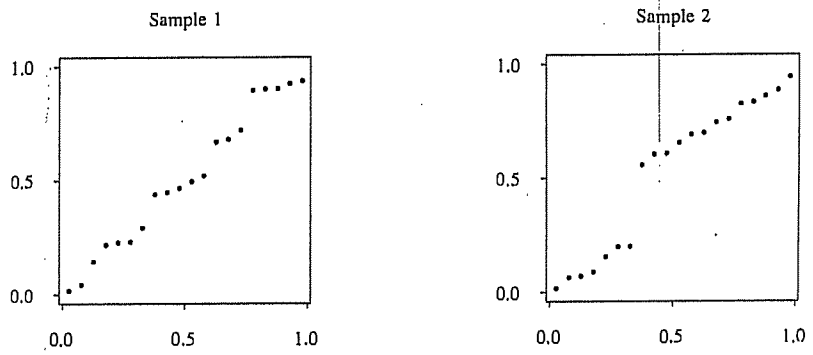

Figure 3.10. Weibull P-P plots for two simulated data sets.

3.3.1.3 *Hazard Plots*

The plotting procedures above were described in terms of survivor functions. They can equally well be described in terms of cumulative hazard functions $H(t)$, in which case they are often termed hazard plots. For the Weibull distribution, for example, $H(t) = -\log S(t)$ satisfies

$$\log H(t) = \beta \log t + \beta \log \lambda.$$

An alternative to plotting $\log[-\log \hat{S}(t)]$ versus $\log t$ would be to plot $\log \hat{H}(t)$ versus $\log t$, where $\hat{H}(t)$ is the nonparametric Nelson–Aalen estimate (3.2.13). As discussed in Section 3.2.2, $\hat{H}(t)$ does not equal $-\log \hat{S}(t)$, so the two plots differ slightly, primarily for large t. We often just plot points corresponding to the observed failure times and, analogous to (3.3.1), replace $\hat{H}(t_j)$ with the value $\hat{H}_j^* = .5\hat{H}(t_j-) + .5\hat{H}(t_j)$, bearing in mind that $\hat{H}(t)$ is right continuous.

Other plotting procedures can often be developed in ad hoc ways. For example, the linear hazard function model where $h(t) = \alpha + \beta t$ has $H(t) = \alpha t + \beta t^2/2$. Thus $t^{-1}H(t) = \alpha + \beta t/2$ is a linear function of t, so approximately linear plots of $t^{-1}\hat{H}(t)$ versus t should result if the model is reasonable.

3.3.1.4 *Discussion*

Plots comparing empirical and model-based survivor or distribution functions provide excellent displays and allow visual checks on models. They complement more formal methods of estimation and testing that are discussed throughout the book.

In cases where probability plots indicate a departure from some parametric model (manifested through nonlinearity in the plot) it is usually possible to see whether the lack of fit is primarily in the tails of the distribution or whether it is in the overall shape. In the latter case, a plot of either the Nelson–Aalen estimate $\hat{H}(t)$ or of $-\log \hat{S}(t)$ versus t is often useful in showing the shape of hazard function needed to

model the data. Nonparametric estimates of $h(t)$, described in Section 3.4, can also be considered.

The use of simulation is recommended for acquiring a sense of the variability in plotting procedures, thus avoiding overinterpretation of features seen in plots. Simulation bands or "envelopes" (Atkinson 1985) are useful with probability plots for a given data set. An awareness of the differing variability across values of t or $\log t$ is also important; variance estimates derived from those for the Kaplan–Meier estimate $\hat{S}(t)$ (see (3.2.3)) provide guidance, as do pointwise nonparametric confidence limits for $S(t)$ or functions of $S(t)$. Simultaneous confidence bands for $S(t)$ can also be constructed (e.g., Nair 1981, 1984), but they are rather complex and, given the informal nature of visual assessment, do not offer much practical advantage for model checking over the pointwise limits.

Finally, we have graphed estimates and presented probability plots in terms of survivor functions throughout this section. In some applications we may wish to graph estimates of distribution functions, $\hat{F}(t) = 1 - \hat{S}(t)$; this is of course immediate, and since $\mathrm{Var}[\hat{F}(t)] = \mathrm{Var}[\hat{S}(t)]$, all of the methods previously considered extend easily to estimation of $F(t)$.

3.3.2 Classic Probability Plots

Historically, probability plotting procedures for uncensored or Type 2 censored univariate data were developed in considerable detail and used extensively for analysis in times when computational power was limited (e.g., Barnett 1975; Nelson 1982). We give a brief description of this classic methodology.

Probability plots in their most common classic form are used with location-scale parameter models. Suppose that X is a random variable with distribution function of the form $F_0[(x-u)/b]$, where b is a scale parameter and u a location parameter ($b > 0$, $-\infty < u < \infty$). Let $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ be the ordered observations in a random sample of size n from the distribution of X. A classic probability plot is a plot of the $x_{(i)}$ against quantities $m_i = F_0^{-1}(a_i)$, where a_i is an expected value related to $F_0[(x_{(i)} - u)/b]$. If the stated model is reasonable, the plot of the points $(x_{(i)}, m_i)$ should be roughly linear. In fact, the points should lie fairly near the line $x = u + bm$, and thus estimates of u and b can be obtained from the plot.

The a_i are referred to as the plotting positions. The two most popular choices are $a_i = (i - .5)/n$ and $a_i = i/(n+1)$. The former is motivated by the fact that the empirical distribution function changes from $(i-1)/n$ to i/n at $x_{(i)}$, so that one can think of $x_{(i)}$ as corresponding to something between the $(i-1)/n$ and i/n quantiles. Taking $(i - .5)/n$, which is midway between these two values, and then equating $x_{(i)}$ and the $(i-.5)/n$ quantile of the distribution, we get $F_0[(x_{(i)} - u)/b] = (i - .5)/n = a_i$. The second choice mentioned, $a_i = i/(n+1)$, is motivated by the fact that $E\{F_0[(X_{(i)} - u)/b]\} = i/(n+1)$; this follows from the fact that the variables $F_0[(X_i - u)/b]$ are Uniform(0, 1). Still another choice is $a_i = E\{[(X_{(i)} - u)/b]\}$, provided that these quantities are available for the distribution in question. For the extreme value distribution these were given by White (1969), and for the normal distribution by Sarhan and Greenberg (1962), among others.

To facilitate probability plots, special probability graph papers were constructed for the more common distributions. These graph papers had a scale based on values of $F_0^{-1}(a)$, but were labeled with an a-scale, so that to effectively plot the points $[x_{(i)}, F_0^{-1}(a_i)]$, one needed only plot the points $(x_{(i)}, a_i)$. This saved the trouble of computing $F_0^{-1}(a)$. Probability papers for the extreme value and normal distributions were particularly useful in early life distribution work; see Nelson (1982). Modern computer graphics can now of course generate these plots.

The classic probability plots are described in terms of the distribution function $F_0(z)$ rather than the s.f. $S_0(z) = 1 - F_0(z)$. However, it is easy to see that when the data are uncensored or Type 2 censored, the probability plot of the points (3.3.4) is precisely the same as a classic probability plot with positions $a_i = (i - .5)/n$. To see this note that in (3.3.4) we then have $x_j = x_{(j)}$ and, as remarked in Example 3.3.2, $\hat{S}_j^* = 1 - (j - .5)/n$. Since $S_0^{-1}(1-p) = F_0^{-1}(p)$ for $0 < p < 1$, the stated result holds.

Historically, there was also considerable discussion about the pros and cons of different plotting positions (e.g., Barnett 1975), some of it directed at the estimation of u and b. There is no generally superior choice of position, and except for very small samples, the choice has little discernible effect on diagnostic plots. The most common practice is to use the positions $a_i = (i - .5)/n$, which corresponds to using (3.3.1) for general probability plots. We use these positions for plots throughout the book.

3.4 ESTIMATION OF HAZARD OR DENSITY FUNCTIONS

3.4.1 General Remarks

Plots of the Kaplan–Meier and Nelson–Aalen estimates of $S(t)$ and $H(t)$ provide some indication of the shapes of the p.d.f. $f(t) = -S'(t)$ and hazard function $h(t) = H'(t)$. It is sometimes desirable to plot nonparametric estimates $\tilde{f}(t)$ and $\tilde{h}(t)$, which give more direct impressions of the density and hazard functions. A comprehensive treatment of this area is beyond the scope of the book; some general discussion and an illustrative example is provided here. Precise nonparametric estimation of density and hazard functions is inherently difficult, since they represent rates of change in probabilities. Nonparametric estimates give an impression of the shape of $f(t)$ or $h(t)$, but it is usually unwise to infer too much about local curvature. The estimates of densities or hazard functions are generally based on smoothing, and various approaches have been considered. These include kernel density estimation (e.g., Tanner and Wong 1983; Ramlau-Hansen 1983), penalized nonparametric maximum likelihood (e.g., O'Sullivan 1988; Green and Silverman 1994), local likelihood (e.g., Hastie and Tibshirani 1990; Loader 1999), and adaptive regression splines fitted by maximum likelihood (e.g., Kooperberg and Stone 1992; Rosenberg 1995). Additional references to methodology and to software are provided in the Bibliographic Notes and Computational Notes at the end of the chapter.

These and other approaches effectively take local averages (perhaps weighted) of observations, discrete estimates, or functions thereof. This involves choices that affect the degree of smoothing. For example, kernel-density estimates of a hazard function $h(t)$ based on a censored random sample are often of the form

$$\tilde{h}(t) = \frac{1}{b}\sum_{j=1}^{k} w\left(\frac{t - t_j^*}{b}\right)\hat{h}(t_j^*), \tag{3.4.1}$$

where, in the notation of Section 3.2, $t_1^* < \cdots < t_k^*$ are the distinct observed failure times in the sample, $\hat{h}(t_j^*) = d_j/n_j$ is the increment in the Nelson–Aalen estimate (3.2.13) at t_j^*, $b > 0$ is a specified bandwidth or window parameter, and $w(u)$ is a fully specified p.d.f. that equals 0 outside the interval $[-1, 1]$. (This estimate needs a little modification for t outside the range $(b, t_k^* - b)$, but for simplicity of discussion we ignore this.) The estimate (3.4.1) at t is a weighted average of values $\hat{h}(t_j^*)$ for t_j^* such that $|t - t_j^*| \le b$. The smoothness of $\tilde{h}(t)$ depends on the shape of $w(u)$ and on b; the bandwidth b is more important, with larger values incorporating more observations t_j^* in the local average at t, thereby giving a smoother $\tilde{h}(t)$.

The amount of smoothing greatly affects estimates $\tilde{f}(t)$ or $\tilde{h}(t)$ and the visual impression created by a plot, but the degree of smoothing to use for a given data set is rather arbitrary. A low degree of smoothing tends to yield "bumpy" estimates with considerable variation in local curvature, and such estimates are usually implausible. On the other hand, a high degree of smoothing, giving a smoother estimate, may miss interesting features suggested by the data. Software implementations often offer automatic selection of bandwidth or analogous smoothing parameters, but it is usually better to base decisions about smoothing on subjective notions about the shape of the underlying $f(t)$ or $h(t)$, and to examine estimates with varying degrees of smoothness.

There are other difficulties with nonparametric estimation of $f(t)$ or $h(t)$. Estimation is imprecise in the tails of the distribution, and estimates of $S(t)$ or $H(t)$ implied by $\tilde{f}(t)$ or $\tilde{h}(t)$ may not agree well with the Kaplan–Meier or Nelson–Aalen estimates. In finite samples, estimators $\tilde{f}(t)$ or $\tilde{h}(t)$ can be quite biased, since they in effect estimate some weighted average of $f(t)$ or $h(t)$. For example, with b fixed (3.4.1) does not estimate $h(t)$ consistently for large samples, but rather

$$h^*(t) = \frac{1}{b}\int_0^\infty w\left(\frac{t-u}{b}\right)h(u)\,du.$$

Finally, in order to adapt to the available data, it may be necessary to vary the degree of smoothing for different t.

The next section considers some simple procedures for the estimation of hazard or density functions, and an illustrative example.

3.4.2 Some Simple Procedures and an Example

Nonparametric estimation of $f(t)$ or $h(t)$ requires a reasonably large number of failure times. A simple and effective approach is to break the time axis into intervals, estimate the hazard or density function at the midpoint of each (finite) interval, and then smooth the estimates. Such procedures are sometimes criticized because of the need to choose intervals and because estimates can sometimes be sensitive to the choice. However, experience suggests that such methods are effective for practical sample sizes, and one can examine estimates for different choices of intervals.

Consider the standard censored random sample of lifetimes, (t_i, δ_i), $i = 1, \ldots, n$. Let $a_0 = 0 < a_1 < \cdots < a_{k+1} = \infty$ partition the time axis, and denote intervals $I_j = (a_{j-1}, a_j]$ for $j = 1, \ldots, k+1$. Let d_j and w_j be the number of failure times and censoring times, respectively, that fall into I_j, and let n_j be the number of t_i (i.e., failure or censoring times) that exceed a_{j-1}. For $j = 1, \ldots, k$ let $t_{mj} = .5(a_{j-1}+a_j)$ denote the midpoint of I_j and $\Delta_j = a_j - a_{j-1}$ denote the interval widths. For some purposes, such as the consideration of frequency properties of estimates, the intervals should in principle be specified independently of the data, but for purposes of data description and visual examination there is little harm in choosing intervals with an eye on the observed data. The number of intervals can depend on the number of failure times, but it is preferable for each interval to contain at least several times.

In what follows, we focus on estimation of the hazard function $h(t)$ for the underlying distribution. There are various ways in which one can estimate $h(t_{mj})$ for $j = 1, \ldots, k$: we mention three. The first is via

$$\hat{h}(t_{mj}) = \frac{\hat{H}_{NA}(a_j) - \hat{H}_{NA}(a_{j-1})}{\Delta_j}, \qquad j = 1, \ldots, k, \tag{3.4.2}$$

where $\hat{H}_{NA}(t)$ is the Nelson–Aalen estimate (3.2.13). The estimate (3.4.2) is simple and effective when $h(t)$ is close to linear over $(a_{j-1}, a_j]$. A second estimate, which is obtained as an m.l.e. under the assumption that $h(t)$ is constant over each interval $j = 1, \ldots, k$ is (see Problem 3.8)

$$\hat{h}(t_{mj}) = \frac{d_j}{\sum_{t_i \in I_j}(t_i - a_{j-1}) + n_{j+1}\Delta_j}, \qquad j = 1, \ldots, k. \tag{3.4.3}$$

Finally, in some settings we may know the n_j, d_j, and w_j values but not the exact failure or censoring times; this is the life table setting discussed in Section 3.6, where estimation of parameters $q_j = Pr(T \le a_j | T > a_{j-1})$ is considered. A good estimate when censoring times tend to be uniformly distributed across I_j is $\hat{q}_j = d_j/(n_j - .5w_j)$, and Problem 3.15 indicates that a reasonable estimate of $h(t_{mj})$ when $h(t)$ does not vary too much over I_j is

$$\hat{h}(t_{mj}) = \frac{-\log(1 - \hat{q}_j)}{\Delta_j}, \qquad j = 1, \ldots, k. \tag{3.4.4}$$

A plot of the points $(t_{mj}, \hat{h}(t_{mj}))$ gives a rough idea of the features of $h(t)$. If desired, standard errors or confidence limits can be shown at each t_{mj}; standard errors are readily obtained for each of (3.4.2)–(3.4.4), or for functions such as $\log \hat{h}(t_{mj})$, which are preferable for constructing confidence intervals. For example, for (3.4.2) it follows from the development of the variance estimate (3.2.32) for the Nelson–Aalen estimator that a variance estimate is

$$\widehat{\text{Var}}[\hat{h}(t_{mj})] = \frac{1}{\Delta_j^2} \sum_{t_i \in I_j} \frac{\delta_i}{n(t_i)^2}, \tag{3.4.5}$$

where $n(t_i)$ denotes the number of individuals at risk (i.e., alive and uncensored) just prior to t_i. The $\hat{h}(t_{mj})$ may be smoothed using some type of scatterplot smoother (e.g., Hastie and Tibshirani 1990, Ch. 2) or a nonparametric regression fit applied to the points $(t_{mj}, \hat{h}(t_{mj}))$; this allows a smooth estimate $\tilde{h}(t)$ to be obtained.

Estimates of $f(t)$ can be obtained from estimates of $h(t)$ and $S(t)$, via $f(t) = h(t)S(t)$. Note that a smooth estimate $\tilde{h}(t)$ produces a corresponding smooth estimate $\tilde{S}(t) = \exp[-\int_0^t \tilde{h}(u)\,du]$, though as we discuss below, these estimates are in some cases not very good. It is also possible to estimate $f(t)$ directly. A reasonable estimate of $f(t_{mj})$ when $f(t)$ is roughly linear over I_j is

$$\hat{f}(t_{mj}) = \frac{\hat{S}_{KM}(a_{j-1}) - \hat{S}_{KM}(a_j)}{\Delta_j}, \qquad j = 1, \ldots, k$$

and an ad hoc approach to obtaining a smooth estimate $\tilde{f}(t)$ is to smooth these values and then rescale the estimate so it integrates to one.

Example 3.4.1. The data below show the number of cycles to failure for a group of 60 electrical appliances in a life test. The failure times have been ordered for convenience.

14	34	59	61	69	80	123	142	165	210
381	464	479	556	574	839	917	969	991	1064
1088	1091	1174	1270	1275	1355	1397	1477	1578	1649
1702	1893	1932	2001	2161	2292	2326	2337	2628	2785
2811	2886	2993	3122	3248	3715	3790	3857	3912	4100
4106	4116	4315	4510	4584	5267	5299	5583	6065	9701

There are a substantial number of small failure times, and the data suggest that the hazard function $h(t)$ may be relatively high for small times. We will investigate this by considering nonparametric estimates of $h(t)$.

Figure 3.11 shows a Nelson–Aalen plot of the data. An initial steep increase in $\hat{H}_{NA}(t)$ is suggested; but given its scale and short duration, this does not stand out dramatically. We investigate $h(t)$ by grouping the data as shown in Table 3.4, and estimating the hazard function at the midpoints of the intervals represented in the table. The table shows the values of n_j, d_j, Δ_j, t_{mj}, and the estimate $\hat{h}(t_{mj})$ given by (3.4.2). The intervals were selected so as to have 5–10 failures in each (except for

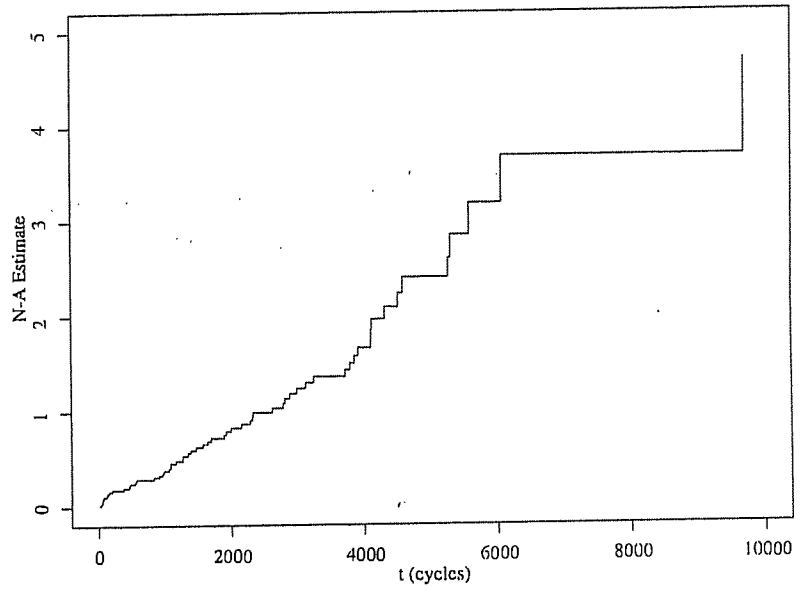

Figure 3.11. Nelson–Aalen c.h.f. estimate for appliance failure times.

Table 3.4. Grouped Appliance Failure Data and Hazard Function Estimates

I_j	n_j	d_j	Δ_j	t_{mj}	$1{,}000\,\hat{h}(t_{mj})$	$1{,}000\,se(\hat{h}(t_{mj}))$
0–100	60	6	100	50	1.044	0.430
100–500	54	7	400	300	0.344	0.131
500–1,000	47	6	500	750	0.270	0.112
1,000–1,500	41	9	500	1,250	0.489	0.166
1,500–2,000	32	5	500	1,750	0.334	0.152
2,000–3,000	27	10	1,000	2,500	0.452	0.148
3,000–4,000	17	6	1,000	3,500	0.420	0.179
4,000–5,000	11	6	1,000	4,500	0.737	0.330
5,000–6,000	5	3	1,000	5,500	0.783	0.548
6,000–10,000	2	2	4,000	8,000	0.375	—

the last two shown); the final interval (10,000, ∞) has no failures and is not shown. Standard errors for the $\hat{h}(t_{mj})$'s are also given, and we note they are rather large; standard errors for estimates (3.4.3) or (3.4.4) are slightly smaller than these. The discussion of shape for $h(t)$ that follows is therefore necessarily tentative.

Figure 3.12 show the hazard values $\hat{h}(t_{mj})$ and some smooth estimates $\tilde{h}(t)$ obtained from the points $(t_{mj}, \hat{h}(t_{mj}))$. The first is a cubic smoothing spline (Green and Silverman 1994), fitted using weights based on the variance estimates for the log

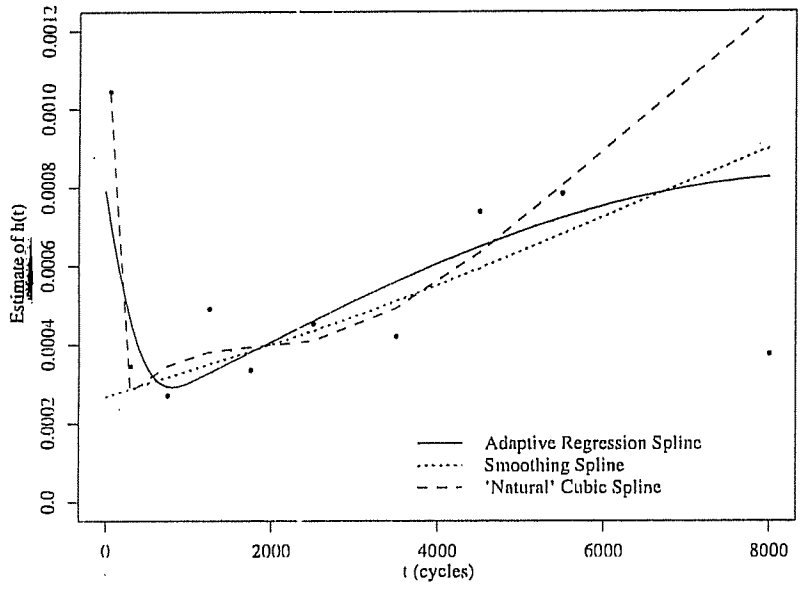

Figure 3.12. Smoo h estimates of $h(t)$ for appliance failure time.

$\hat{h}(t_{mj})$'s; these variance estimates are given by (3.4.5) divided by $\hat{h}(t_{mj})^2$. The S-Plus 2000 function smooth.spline with automatic selection of smoothing parameter was used to obtain the estimate; other scatterplot smoothers tend to give results similar to this. Such procedures are not able to accommodate the large estimate $\hat{h}(t_{m1})$ at the first interval's midpoint. A procedure that is better able to capture this behavior is to fit a cubic regression spline. This is a function $\tilde{h}(t)$ that is smooth and piecewise cubic between cut points, which are called knots; such models are discussed in Section 4.2.3. Figure 3.12 shows a "natural" cubic spline with internal knots at $t = 100$, 300 and 2500, and boundary knots at $t = 50$ and 5500; the natural spline is piecewise cubic everywhere between the two boundary knots and linear outside them. The estimate portrayed in Figure 3.12 was obtained very simply by a weighted least-square fit to the data (y_j, x_j) with $y_j = \hat{h}(t_{mj})$, $x_j = t_{mj}$ and the same weights used for the smoothing spline fit. This procedure approximates the estimates $\hat{h}(t_{mj})$ considerably better.

The regression spline is essentially a parametric model, and the knot positions used here were chosen so as to capture features in the $\hat{h}(t_{mj})$'s shown in Figure 3.12. Adaptive regression spline procedures can also handle this quite well. Figure 3.12 includes a third estimate $\tilde{h}(t)$ obtained by using the logspline software of Kooperberg and Stone (1992), by way of illustration. This approach uses a selection procedure for the number of knots, and utilizes the ungrouped data. It agrees rather well with the regression spline estimate obtained from the grouped data. Kernel smoothing

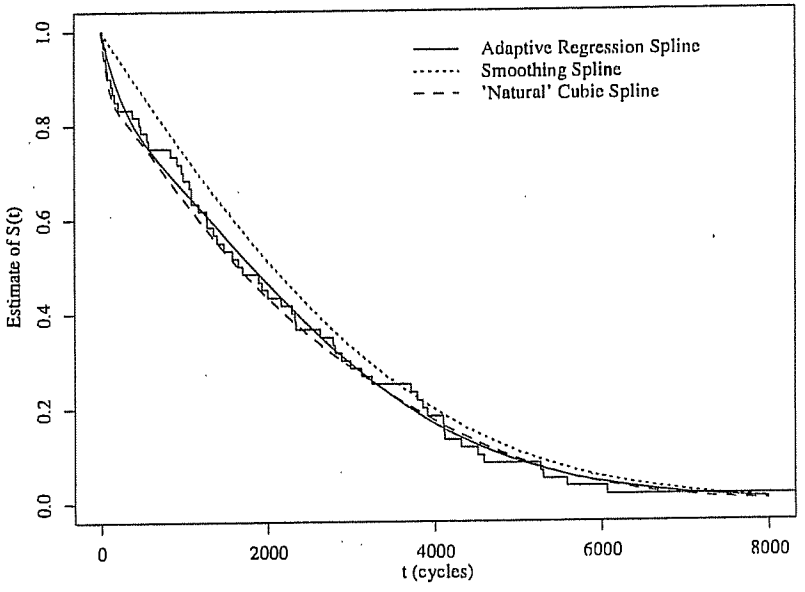

Figure 3.13. Smooth estimates and KM estimates of $S(t)$ for appliance failure time.

methods based on the ungrouped data (e.g., Tanner and Wong 1983) can also be applied.

Estimates $\tilde{h}(t)$ that do not mimic the $\hat{h}(t_{mj})$ values fairly closely can yield estimates

$$\tilde{S}(t) = \exp\left[-\int_0^t \tilde{h}(u)\,du\right]$$

of the s.f. that do not agree well with the Kaplan–Meier estimate. Figure 3.13 shows the Kaplan–Meier estimate $\hat{S}(t)$ and smooth estimates $\tilde{S}(t)$ derived from the estimates of $h(t)$ in Figure 3.12. The smooth estimates mostly lie within pointwise nonparametric confidence bands for $S(t)$ obtained from the Kaplan–Meier estimate, but the regression spline estimates are considerably closer to $\hat{S}(t)$.

We conclude by mentioning that in Example 4.4.2 a mixture of two Weibull distributions is fitted to the data in this example.

3.5 METHODS FOR TRUNCATED AND INTERVAL CENSORED DATA

As described in Sections 2.3 and 2.4, lifetimes T_i are often subject to left or right truncation, and to interval censoring. This section discusses nonparametric estimates of $S(t)$ in these cases.

3.5.1 Left-Truncated Data

Settings in which left-truncated lifetimes arise were described in Section 2.4. Following the notation there, we assume that for each of n independent individuals a triple (u_i, t_i, δ_i) is observed, where u_i is the left truncation time, $t_i \geq u_i$ is a lifetime or censoring time, and $\delta_i = 1\,(0)$ if t_i is a lifetime (censoring time). It is assumed further that truncation is "independent," so that (2.4.1) holds. Our objective is nonparametric estimation of $S(t) = Pr(T_i \geq t)$, but as will become apparent, it is possible to estimate only

$$\begin{aligned} \frac{S(t)}{S(u_{\min})} &= Pr(T \geq t | T \geq u_{\min}), \\ &= S_L(t; u_{\min}) \qquad t \geq u_{\min}, \end{aligned} \tag{3.5.1}$$

where $u_{\min} = \min(u_1, \ldots, u_n)$.

The likelihood function based on (u_i, t_i, δ_i), $i = 1, \ldots, n$ is given by (2.4.2). As noted in Section 3.2.1, estimation of the hazard function in the discrete-time setting is essentially the same as when there is no truncation. The likelihood can be written in the same form (3.2.4) if we define $n_t = \sum I(u_i \leq t \leq t_i)$, and the m.l.e. of the hazard function at time t is then $\hat{h}(t) = d_t/n_t$, provided $n_t > 0$. There is, however, an important difference with the untruncated case. Since $n_t = 0$ for $t < u_{\min}$, (3.2.4) has no information about $h(t)$ for $t < u_{\min}$. Consequently, we cannot estimate $S(t)$ unless $u_{\min} = 0$ (see (1.2.16)), but only $S_L(t; u_{\min})$ of (3.5.1). By (1.2.18) the estimate is

$$\hat{S}_L(t; u_{\min}) = \prod_{\ell=u_{\min}}^{t-1} [1 - \hat{h}(\ell)]. \tag{3.5.2}$$

In the continuous-time setting we get the estimate

$$\hat{S}_L(t; u_{\min}) = \prod_{j: t_j^* < t} \left(1 - \frac{d_j}{n_j}\right), \tag{3.5.3}$$

where $t_1^* < t_2^* < \cdots < t_k^*$ are the distinct observed failures times, $d_j = \sum I(t_i = t_j^*, \delta_i = 1)$, and $n_j = \sum I(u_i \leq t_j^* \leq t_i)$. The Greenwood-type variance estimate

$$\widehat{\mathrm{Var}}\{\hat{S}_L(t; u_{\min})\} = \hat{S}_L(t; u_{\min})^2 \sum_{j: t_j^* < t} \frac{d_j I(n_j > d_j)}{n_j(n_j - d_j)} \tag{3.5.4}$$

is obtained by the same approach as in Section 3.2.1. For reasons discussed in the following, it is necessary to include the factor $I(n_j > d_j)$ in (3.5.4) and to define $0/0$ as 0.

Other issues arise with left-truncated data. The n_j are not monotone decreasing as they are for untruncated data, and it is possible to have times t between $u_{\min}$ and $\max(t_i)$ for which $n_t = \sum I(u_i \leq t \leq t_i) = 0$. Strictly speaking, the likelihood function has no information about $h(t)$ or $dH(t)$ at such points, but it is conventional to ignore this and define $d\hat{H}(t) = 0$, thus allowing estimation of $S_L(t; u_{\min})$. It is also possible to have $d_j = n_j$ for any $j = 1, \ldots, k$; with untruncated data this can occur only at t_k^*, the largest failure time. But if $d_j = n_j$, then $d\hat{H}(t_j^*) = 1$, and in (3.5.3) we have $\hat{S}_L(t; u_{\min}) = 0$ for all $t > t_j^*$. If this occurs for t_j^* small (it is even possible to have $d_1 = n_1$), then this estimate is usually quite implausible, and is not very useful.

This pathological behavior of (3.5.3) is not uncommon when few u_i are close to 0. Often the best option is to recognize that the data have little information about $H(t)$ close to $t = 0$, and to select a new left truncation time $u^* > u_{\min}$, retaining only individuals for whom $u_i \geq u^*$. We can then estimate $dH(t)$ for $t \geq u^*$ and $S_L(t; u^*)$ nonparametrically; under the independent truncation assumption, this selection of individuals does not introduce any bias.

Other estimates than (3.5.3) are sometimes suggested. One approach is to adopt a flexible parametric model such as a regression spline for $h(t)$, or to smooth the estimates $d\hat{H}(t_j^*) = d_j/n_j$ as discussed in Section 3.4. Another is to estimate $S_L(t; u_{\min})$ through the cumulative hazard function. Defining

$$H_L(t; u_{\min}) = \int_{u_{\min}}^{t} dH(u), \tag{3.5.5}$$

we have from the maximum likelihood development the Nelson–Aalen estimate

$$\hat{H}_L(t; u_{\min}) = \sum_{j: t_j^* \leq t} \frac{d_j}{n_j}, \tag{3.5.6}$$

with associated variance estimate

$$\widehat{\mathrm{Var}}\{\hat{H}_L(t; u_{\min})\} = \sum_{j: t_j^* \leq t} \frac{d_j}{n_j^2}. \tag{3.5.7}$$

An alternative estimate for $S_L(t; u_{\min})$ in the continuous-time case may be based on the fact that $S_L(t; u_{\min}) = \exp[-H_L(t; u_{\min})]$ for continuous models, suggesting

$$\tilde{S}_L(t; u_{\min}) = \exp[-\hat{H}_L(t; u_{\min})]. \tag{3.5.8}$$

This estimate does not drop to zero when $d_j = n_j$. Other alternative estimators that have been proposed compensate for a lack of information through additional assumptions. Neither they nor (3.5.8) are satisfactory in some situations, however, and it is best to adopt the more severe lower limit $u^* > u_{\min}$ discussed earlier.

Confidence intervals for $S_I(t; u_{\min})$ or $S_L(t; u^*)$ can be obtained by using any of the approaches in Section 3.2, merely replacing $S(t)$ and related estimates by $S_L(t; u_{\min})$ and related estimates.

Probability plots to check parametric models cannot readily be linearized in the case of truncated data. However, nonparametric and parametric estimates of $S_L(t; u_{\min})$ or $H_L(t; u_{\min})$ are easily compared by plotting them on the same graph. It is also possible to base probability plots on suitably defined residuals. Both approaches are considered in the following example.

Example 3.5.1. Example 2.4.2 presented a set of data on the lifetimes of the brake pads on 98 automobiles. The lifetimes were actually estimated, and were subject to left truncation because of the way the automobiles in the sample were selected. We ignore the effect of estimation of lifetimes in this example. Table 2.1 shows the (estimated) lifetimes t_i and truncation times u_i (in thousands of km driven) for the 98 vehicles. Since $u_{\min} = 7.0$ we can estimate $S_L(t; 7.0)$ nonparametrically. Figure 3.14 shows the estimate (3.5.3) along with a parametric estimate based on a log-normal distribution (1.3.10) for the lifetimes T_i; the m.l.e.'s for this model are $\hat{\mu} = 4.109$, $\hat{\sigma} = 0.421$. It is reasonable to assume that $S(7.0)$ is effectively one (i.e., no brake pads wear out before 7000 km driven), so $\hat{S}_L(t; 7.0)$ can be assumed to estimate $S(t)$ here. Figure 3.14 is labeled to reflect this.

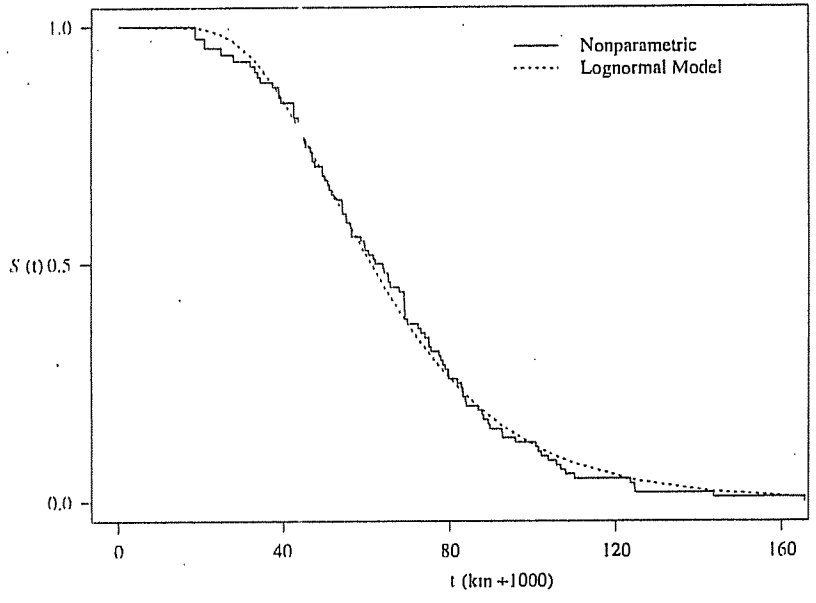

Figure 3.14. Nonparametric and log-normal estimates of $S(t)$ for brake-pad life.

The log-normal model apparently fits the data well, but let us consider a probability plot anyway, to illustrate how left truncation can be handled. The simplest approach is to employ a P-P plot by defining uniform residuals for a parametric model as

$$e_i = \frac{S(t_i; \hat{\theta})}{S(u_i; \hat{\theta})} \qquad i = 1, \ldots, n.$$

Since $S(T_i; \theta)/S(u_i; \theta)$ is Uniform (0, 1), given u_i, if the model $S(t; \theta)$ is correct, we plot as a model check the points $(e_{(i)}, (i - .5)/n)$, where $e_{(i)}$ is the ith smallest among $(e_1, \ldots, e_n)$. The $e_{(i)}$ can be thought of as theoretical approximate $U(0, 1)$ quantiles and the values $(i - .5)/n$ as sample uniform quantiles. Figure 3.15 shows the P-P plot with $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ for the log-normal model; there is no evidence to contradict the model.

Example 3.5.2. To illustrate the pathological behavior that can occur for a nonparametric estimate of $S(t)$ based on truncated data, let us consider the data in Table 3.5. These consist of left truncation times, u_i, and lifetimes, t_i, for 18 brake pads similar to those in Example 3.5.1. In this case, the smallest lifetime is $t_1 = 24.5$, with corresponding $u_1 = 19.6$. However, since all of the other truncation times u_i exceed 24.5, there is only this single unit at risk at $t = 24.5$, so in (3.5.3) we have

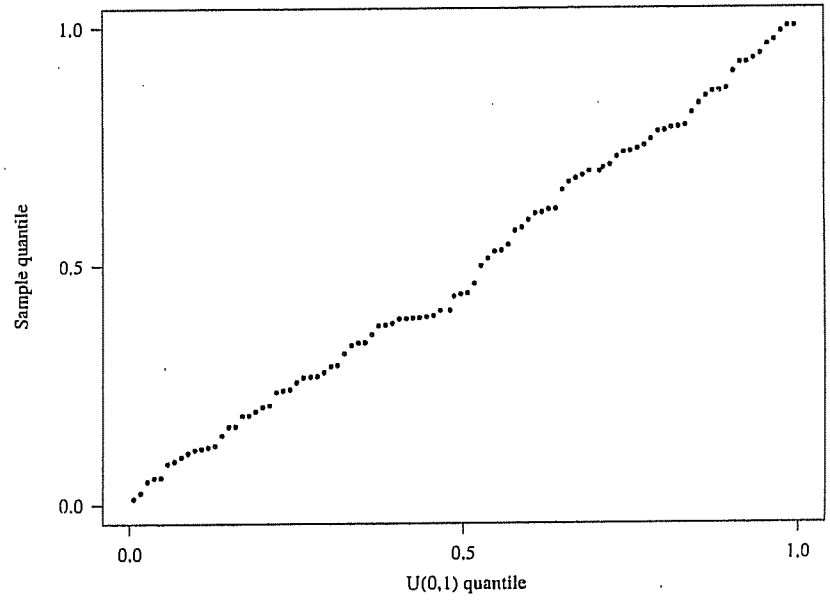

Figure 3.15. Log-normal P-P plot for truncated brake-pad life data.

Table 3.5. Brake-Pad Life (t) and Odometer Readings (u) for 18 Cars

u^a	t	u	t	u	t
19.6	24.5	44.6	46.9	26.2	59.3
30.7	53.8	48.8	61.4	47.8	97.8
37.8	58.4	45.9	64.2	48.2	86.2
34.4	89.5	28.2	45.6	27.9	40.8
37.4	75.6	48.3	81.9	33.7	93.8
41.1	56.0	42.5	86.3	37.6	77.7

[a]Units are 1000 km.

$d_1 = n_1 = 1$, and the estimate of $S_L(t; 19.6)$ is

$$\hat{S}_L(t; 19.6) = \begin{cases} 1 & 19.6 \le t \le 24.5 \\ 0 & t > 24.5 \end{cases}.$$

This is an unacceptable estimate. An alternative ad hoc approach, previously discussed, is to drop the first observation in Table 3.5 and to use the remaining 17 to estimate $S_L(t; 26.2)$. The estimate (3.5.3) for this is sensible, and if we were willing to estimate $S(26.2)$ from other information (its value is certainly close to one), then the unconditional s.f. $S(t)$ can also be estimated. Note that in this setting the value $t_1 = 24.5$ is unusually low, and though we may not wish to consider it an extreme outlier, estimating $S(t)$ with and without it included makes good sense.

3.5.2 Right-Truncated Data

Lifetimes T_i may also be subject to right truncation, as described in Example 2.4.4. In this case v_i is a truncation time such that individual i is observed only if $T_i \le v_i$. Data on n individuals consist of independent pairs (t_i, v_i), where $t_i \le v_i$ is the observed lifetime; little additional complication is created if we allow some lifetimes to be left-censored, but this occurs rarely in practice, so we disregard it for now. We also assume that truncation is independent, so that $Pr(T \le t | v, T \le v) = F(t)/F(v)$, where $F(t) = Pr(T \le t)$ is the marginal distribution function of T.

With right-truncated data it is simplest to work with the distribution function of T. Our objective is to estimate $F(t)$ nonparametrically, but it is clear that all we can estimate is

$$\frac{F(t)}{F(v_{\max})} =: Pr(T \le t | T \le v_{\max}), \qquad (3.5.9)$$
$$=: F_R(t; v_{\max}) \qquad t \le v_{\max},$$

where $v_{\max} = \max(v_1, \ldots, v_n)$. This is analogous to our being able to estimate only (3.5.1) in the case of left truncation. In fact, there is a close connection with the development in Section 3.5.1, because if we reverse the time axis, then left truncation

becomes right-truncation and right censoring becomes left censoring. It follows by analogy with (3.5.3) that the nonparametric Kaplan–Meier estimate of $F_R(t; v_{\max})$ should be

$$\hat{F}_R(t; v_{\max}) = \prod_{j: t_j^* > t} \left(1 - \frac{d_j}{n_j}\right), \qquad (3.5.10)$$

where $t_1^* < \cdots < t_k^*$ are the distinct observed failure times, $d_j = \sum I(t_i = t_j^*)$ and $n_j = \sum I(t_i \le t_j^* \le v_i)$. The Greenwood variance estimate for $\hat{F}_R(t; v_{\max})$ is, by analogy with (3.5.4),

$$\widehat{\mathrm{Var}}\{\hat{F}_R(t; v_{\max})\} = \hat{F}_R(t; v_{\max})^2 \sum_{j: t_j^* > t} \frac{d_j I(n_j > d_j)}{n_j(n_j - d_j)}, \qquad (3.5.11)$$

where we interpret $0/0$ as 0.

A direct derivation of (3.5.10) and (3.5.11) along the lines of the discrete-time development of Section 3.2 is instructive. We define the "reverse time hazard" function $h_{RT}(t) = f(t)/F(t)$, where $t = 0, 1, \ldots$, and note that for any (t, v) with $0 \le t < v$,

$$\frac{F(t)}{F(v)} = \prod_{\ell=t+1}^{v} [1 - h_{RT}(\ell)] \qquad (3.5.12)$$

$$\frac{f(t)}{F(v)} = h_{RT}(t) \prod_{\ell=t+1}^{v} [1 - h_{RT}(\ell)]. \qquad (3.5.13)$$

The likelihood function from n independent pairs (t_i, v_i) is then

$$\begin{aligned} L(\mathbf{h}_{RT}) &= \prod_{i=1}^{n} \frac{f(t_i)}{F(v_i)} \\ &= \prod_{i=1}^{n} h_{RT}(t_i) \prod_{\ell=t_i+1}^{v_i} [1 - h_{RT}(\ell)] \\ &= \prod_{t=1}^{v_{\max}} h_{RT}(t)^{d_t} [1 - h_{RT}(t)]^{n_t - d_t}, \end{aligned} \qquad (3.5.14)$$

where $d_t = \sum I(t_i = t)$, $n_t = \sum I(t_i \le t \le v_i)$, and $\mathbf{h}_{RT} = (h_{RT}(1), \ldots, h_{RT}(v_{\max}))$. Maximization of (3.5.14) gives $\hat{h}_{RT}(t) = d_t/n_t$, provided $n_t > 0$, and so (3.5.10) follows from (3.5.12). The variance estimate (3.5.11) follows from maximum likelihood large-sample theory, exactly as in Section 3.2. Rigorous treatments of the estimates in the continuous-time setting can also be given.

The same remarks concerning t-values where $n_t = 0$ hold here as in Section 3.5.1. Also, the same pathological behavior as in the case of left truncation can appear in (3.5.10). In particular, if at t_j^* we have $d_j = n_j$, then $\hat{F}_R(t; v_{\max}) = 0$ for all $t < t_j^*$. This phenomenon tends to occur when the v_i are sparse near $v_{\max}$, and one of the approaches discussed for left-truncated data should then be considered. In particular, since the data are uninformative about $dH_{RT}(t)$ near $v_{\max}$, it may be wisest to adopt a smaller right-truncation limit v^*, and to focus on $F(t)/F(v^*) = F_R(t; v^*)$.

In terms of hazard functions, we are able to estimate nonparametrically

$$H_R(t; v_{\max}) = \int_t^{v_{\max}} dH_{RT}(u),$$

the Nelson–Aalen estimate being

$$\hat{H}_R(t; v_{\max}) = \sum_{j: t_j^* > t} \frac{d_j}{n_j}. \tag{3.5.15}$$

For model checking, plots of $\hat{F}_R(t; v_{\max})$ or $\hat{H}_R(t; v_{\max})$ along with parametric estimates are useful.

Example 3.5.3. Kalbfleisch and Lawless (1989) discussed data on the induction or latency time for the Acquired Immune Deficiency Syndrome, or AIDS. A diagnosis of AIDS follows infection with the human immunodeficiency virus, or HIV. Brookmeyer and Gail (1994) provide considerable medical and statistical background on HIV and AIDS but, briefly, AIDS is a condition in humans attributable to a breakdown of the body's immune system, and the HIV is a virus that is believed to cause AIDS. The time between infection with the HIV and the diagnosis of AIDS in an individual is called the induction time. It is highly variable, and can exceed 20 years; some infected individuals may indeed never be diagnosed with AIDS.

The definition of AIDS has changed since the first cases in North America were diagnosed in the early 1980s, and treatments have been developed that can delay the onset of AIDS. However, considerable effort was expended in the 1980s on estimating the induction-time distribution for various types of individuals. This was made difficult by the fact that for most persons infected with the HIV, the time of infection was unknown. Kalbfleisch and Lawless (1989) considered data that were obtained from persons diagnosed with AIDS and whose HIV infection came from a blood transfusion on a known date. Therefore an induction time t_i and an infection time x_i were available for each individual. The data were derived from AIDS cases diagnosed prior to July 1, 1986, however, so this created a right-truncation time v_i for each individual.

We set up notation for the data as follows. We take as the time origin January 1, 1978, which is assumed to be the earliest an HIV infection via blood transfusion could occur in North America. Let τ represent the time (say in days) from January 1, 1978 to July 1, 1986, let x_i represent the time of HIV infection for individual i, and

let t_i represent the AIDS induction time. Then, because only persons diagnosed with AIDS by time τ are included in the data set, it must be the case that $t_i \le v_i$, where $v_i = \tau - x_i$, so the induction times are right truncated, as noted in Example 2.4.4.

Kalbfleisch and Lawless (1989) presented data on 295 individuals whose AIDS cases were reported to the Centers for Disease Control in Atlanta, Georgia. Here we consider a subset of 124 persons who were infected by December 31, 1985 and whose ages were between 5 and 59 years at the time of HIV infection. The data are given in Appendix G. Times were recorded in months, ignoring for convenience the fact that months vary in length, and infections were assumed to occur at the midpoints of months. In the notation of (3.5.9) and (3.5.10), $v_{\max} = 99.5$ months and the observed induction times $t_1^* < \cdots < t_k^*$ range from 4 months to 89 months. Figure 3.16 shows the nonparametric estimate (3.5.10) of the distribution function for induction time T, conditional on $T \le 99.5$ months. Confidence limits are not shown, but we note that standard errors for $\hat{F}_R(t; 99.5)$ based on (3.5.11) are rather large, except for small t; for $t \ge 42$ months, the standard errors exceed .10.

The unconditional induction probability $F(t)$ is inestimable from the truncated data, but an estimate could be obtained for $t \le 99.5$ months if there were an estimate from other sources of $F(99.5)$. A parametric model fitted to the data here would also provide an estimate of $F(t)$. This is considered in Example 4.3.3, where it is found that the truncation so limits the information about the parameters that precise estimation is impossible.

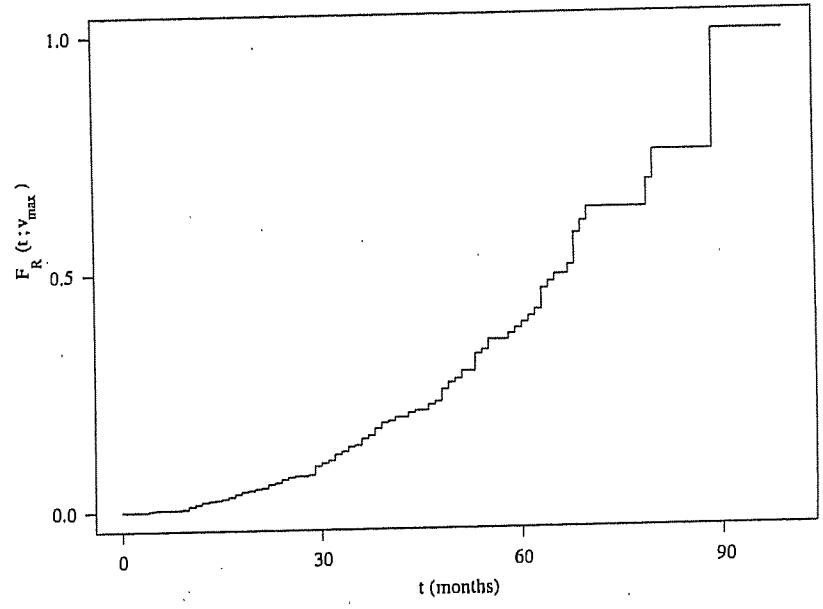

Figure 3.16. Estimate of conditional c.d.f. for AIDS induction time.

3.5.3 Interval-Censored Data

In Section 2.3 we discussed interval-censored data in which the responses are independent pairs (u_i, v_i), $i = 1, \ldots, n$, it being known that $u_i < T_i \leq v_i$. (An exactly known lifetime is given by taking $u_i = v_i -$.) When the individual lifetimes T_i are identically distributed with c.d.f. $F(t)$, the likelihood function (2.3.1) becomes

$$L = \prod_{i=1}^{n} [F(v_i) - F(u_i)], \tag{3.5.16}$$

This depends on $F(t)$ only through values at the observation times (u_i, v_i). To consider this in a slightly different form, let $0 = s_0 < s_1 < \cdots < s_{k-1} < s_k = \infty$ denote the distinct values in the set $\{0, \infty, u_i, v_i : i = 1, \ldots, n\}$, with the convention that an exact observation t is regarded as $(t-, t]$. Let $p_j = F(s_j) - F(s_{j-1})$ and define $\alpha_{ij} = I\{(s_{j-1}, s_j] \subseteq (u_i, v_i]\}$. Then (3.5.16) can be rewritten as

$$L(\mathbf{p}) = \prod_{i=1}^{n} \left[\sum_{j=1}^{k} \alpha_{ij} p_j \right], \tag{3.5.17}$$

and to obtain $\hat{F}$ we must apparently maximize $L(\mathbf{p})$ subject to the constraints $p_j \geq 0$ and $\sum p_j = 1$. Some of the $\hat{p}_j$ in the maximizer of (3.5.17) may equal 0. In fact, it can be shown (Turnbull 1976) that $\hat{p}_j$ must be 0 if s_{j-1} and s_j do not correspond to some u_i and $v_{i'}$, respectively. However, other $\hat{p}_j$ may also be zero. Various algorithms for maximizing (3.5.16) or (3.5.17) have been proposed. References are given in the Bibliographic Notes at the end of the chapter, and the Computational Notes discuss software.

Note that the maximizer of (3.5.17) assigns probabilities $\hat{p}_j$ only to intervals $(s_{j-1}, s_j]$. Thus, if $\hat{p}_j > 0$, the distribution of the probability mass between s_{j-1} and s_j is unspecified. It is customary when plotting and summarizing the estimates $\hat{F}(t)$ or $\hat{S}(t)$ to show them as constant over intervals where $\hat{p}_j = 0$, and unspecified over intervals $(s_{j-1}, s_j]$ for which $\hat{p}_j > 0$, except at s_j. Some software packages, however, extend horizontal pieces of $\hat{S}(t)$ to produce a step function.

Example 3.5.4. Consider a simple artificial example in which there are five observations: $(0, 4]$, $(3, 6]$, $(8, 10]$, $(9, \infty]$, and $[7, 7]$. Rewriting the final, exact observation as $(7-, 7]$ we obtain the intervals $(s_{j-1}, s_j]$ for $j = 1, \ldots, 9$, where $(s_0, s_1, \ldots, s_9) = (0, 3, 4, 6, 7-, 7, 8, 9, 10, \infty)$. The likelihood function (3.5.17) is thus

$$L(p_1, \ldots, p_9) = (p_1 + p_2)(p_2 + p_3)(p_7 + p_8)(p_8 + p_9)p_5,$$

but by the result that $\hat{p}_j = 0$, unless s_{j-1} is some left endpoint u_i and s_j is some right endpoint $v_{i'}$, we have that only $\hat{p}_2$, $\hat{p}_5$, $\hat{p}_8$ may be nonzero. Maximizing $L = p_2^2 p_5 p_8^2$ subject to $p_2 + p_5 + p_8 = 1$, we obtain $\hat{p}_2 = \hat{p}_8 = .4$, $\hat{p}_5 = .2$.

Table 3.6. Nonparametric Estimate from Interval-Censored Data

Interval	$\hat{S}(t)$
(0, 3]	1.0
(3, 4]	Indeterminate
(4, 7]	0.6
(7, 9]	0.4
(9, 10]	Indeterminate
(10, ∞]	0.0

The estimate $\hat{S}(t)$ is shown in Table 3.6 and Figure 3.17. Since the estimate assigns probability only to the intervals (3, 4], [7, 7] and (9, 10], $\hat{S}(t)$'s value is indeterminate except at the endpoints of those intervals.

Asymptotic theory and confidence-interval estimation are problematic in general, though standard results can be obtained if we assume that the set of interval endpoints $s_1, \ldots, s_k$ is finite and fixed as $n \to \infty$. Variance estimates and confidence intervals can then be based on the inverse of the observed information matrix obtained from (3.5.17). When confidence intervals are important an alternative approach is to

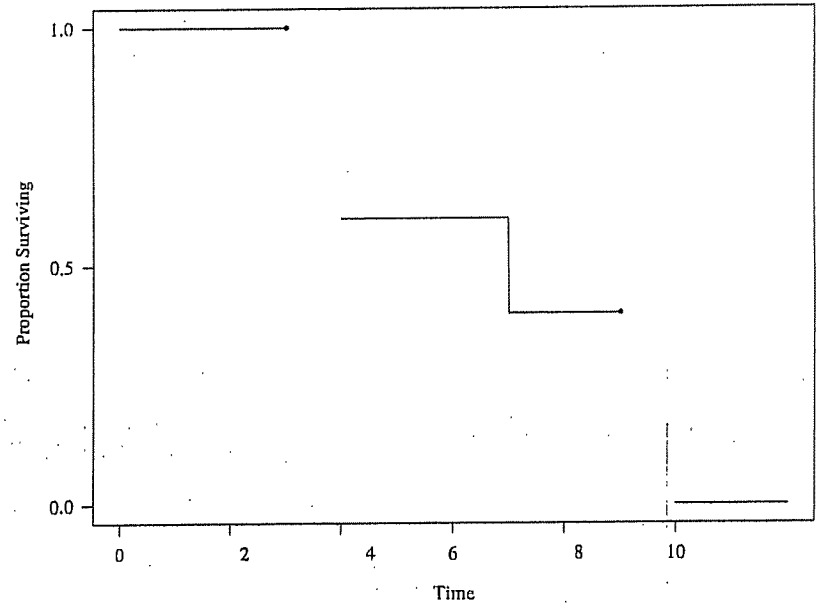

Figure 3.17. Nonparametric estimate of $S(t)$ from interval-censored data.

abandon nonparametric estimation and use a flexible parametric model, for which standard maximum likelihood methods apply. Parametric models are considered in Chapters 4 and 5.

Current status data, introduced in Example 3.2.3, involve the most severe form of interval censoring. The interval within which T_i is known to lie is either $(0, C_i]$ or (C_i, ∞); such observations arise when each individual i is seen only at time C_i, at which point it is determined whether failure has already occurred (i.e., $T_i \le C_i$) or not ($T_i > C_i$). By the preceding arguments, the nonparametric m.l.e. of $F(t)$ or $S(t)$ is defined at $t = 0$, $t = \infty$, and the distinct points $s_1 < s_2 < \cdots < s_{k-1}$, which are contained in $\{C_1, \ldots, C_n\}$. The estimate in other words assigns probabilities $\hat{p}_j$ to the intervals $(s_{j-1}, s_j]$, $j = 1, \ldots, k$, where $s_0 = 0$ and $s_k = \infty$. Interestingly, there is a closed form for $\hat{F}(t)$ in this case (e.g., Huang and Wellner 1997, p. 127): for $j = 1, \ldots, k-1$

$$\hat{F}(s_j) = \max_{u \le j} \min_{v \ge j} \left(\sum_{\ell=u}^{v} d_\ell / \sum_{\ell=u}^{v} n_\ell \right), \tag{3.5.18}$$

where $d_\ell = \sum I(T_i \le C_i, C_i = s_\ell)$ and $n_\ell = \sum I(C_i = s_\ell)$.

Example 3.5.5. Nelson (1982) and Meeker and Escobar (1998) give data from a study on the time T to the initiation (appearance) of cracks in metal turbine wheels. The data in Table 3.7 are a slightly modified version of the Nelson data, in which each of 432 wheels was examined once to determine whether a crack had yet appeared. Thus, for example, 53 wheels were inspected at time 10 (1000 hours), and four had $T_i \le 10$, and 49 had $T_i > 10$.

Table 3.8 shows the nonparametric m.l.e. for $F(t)$, obtained from (3.5.18) or, equivalently, maximization of (3.5.17). The values of $\hat{F}(t)$ are shown at the

Table 3.7. Current Status Data on Time to Crack Initiation

Inspection Time[a]	Number of Wheels Cracked	Number of Wheels Not Cracked
4	0	39
10	4	49
14	2	31
18	7	66
22	5	25
26	9	30
30	9	33
34	6	7
38	22	12
42	21	19
46	21	15

[a] In 100 hour units.

Table 3.8. Nonparametric Estimate of $F(t)$ for Wheel-Crack Initiation

Interval $(s_{j-1}, s_j]$	$\hat{p}_j$	$\hat{F}(s_j)$	$se(\hat{F}(s_j))$
(0, 4]	.000	0.000	.000
(4, 10]	.070	.070	.027
(10, 14]	.000	.070	.027
(14, 18]	.026	.096	.034
(18, 22]	.071	.167	.068
(22, 26]	.056	.222	.046
(26, 30]	.000	.222	.046
(30, 34]	.239	.462	.138
(34, 38]	.120	.581	.057
(38, 42]	.000	.581	.057
(42, 46]	.002	.583	.082
(46, ∞]	.417	1.000	.000

values s_j, which represent the distinct endpoints for the intervals of the form $(0, C_i]$ or (C_i, ∞) within which values of T_i lie; the estimate is indeterminate between successive s_j. Note that four of the $\hat{p}_j$ equal zero. Standard errors are also shown for each $\hat{F}(s_j)$; these are obtained from the asymptotic covariance matrix of the interval probabilities $\hat{p}_j$ based on the likelihood (3.5.17), with p_1, p_3, p_7, and p_{10} set equal

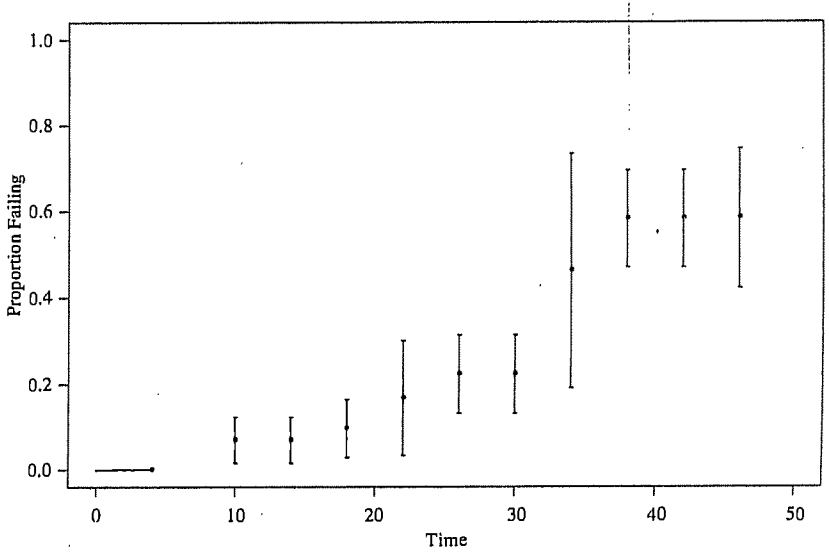

Figure 3.18. Nonparametric estimates of c.d.f. and .95 confidence limits for time to crack initiation.

to 0. The asymptotic covariance matrix is determined from standard likelihood theory as $I(\hat{\mathbf{p}})^{-1}$, where $I(\mathbf{p}) = (-\partial^2 \log L/\partial \mathbf{p} \partial \mathbf{p}')$. This approach is ad hoc and these standard errors and approximate .95 confidence intervals $\hat{F}(s_j) \pm 1.96 se(\hat{F}(s_j))$ for $F(s_j)$ should be treated cautiously. Figure 3.18 shows the estimates $\hat{F}(s_j)$ and approximate .95 confidence intervals; the confidence intervals are rather wide.

In some settings where the lifetime T represents the time between two events, it may happen that the times of both events are censored. This was referred to as double censoring in Section 2.3.2, where it was noted that the likelihood function (2.3.3) depends not just on the distribution of T, but on the distribution of the "initial" event times. Nonparametric estimation of $F(t)$ is possible in this setting (e.g., J. Sun 1997) and in more complicated settings that also involve truncation. The information about $F(t)$ is generally very limited, however, and fully nonparametric estimates often have few points of increase and are highly variable. A generally preferable approach is to use a weakly parametric model such as described in Section 3.4. The piecewise exponential (Section 1.3.8) is easy to use when estimates of $F(t)$ are the main priority (e.g., Lindsey and Ryan 1998). Joly et al. (1998) consider spline-based methods.

3.6 LIFE TABLES

The life table is one of the oldest and most widely used methods of portraying lifetime data. The cohort life table discussed here has been employed at least since the beginning of the twentieth century; population life tables used by demographers and actuaries have been around considerably longer.

The life table is primarily a device for portraying lifetime data for a sample or cohort of individuals, in which lifetimes and censoring times are grouped into intervals. In the situations considered in this book the cohort is usually assumed to be a random sample from some population, and then the life table also provides estimates of the lifetime distribution in the population. The life table is more complex than ordinary interval censoring, since for some individuals it is known that their lifetimes were censored, but the censoring time is known only to lie in some interval.

There is a standard type of life table methodology, which we will describe first. This is followed by some underlying theory and discussion of other approaches.

3.6.1 Standard Life Table Methods

Suppose the time axis is divided into $k+1$ intervals $I_j = [a_{j-1}, a_j)$, $j = 1, \ldots, k+1$, with $a_0 = 0$, $a_k = L$, and $a_{l+1} = \infty$, where L is an upper limit on observation. The definition of the I_j as open on the right and closed on the left is customary in life table methodology. For each member of a random sample of n individuals from some population, suppose that one observes either a lifetime T or a censoring time C. The data are, however, grouped so that it is only known in which intervals particular individuals died or were censored, and not the exact lifetimes and censoring times.

The data therefore consist of the numbers of lifetimes and censoring times falling into each of the $k+1$ intervals. In the case of the last interval, I_{k+1}, it can be considered that only lifetimes are in the interval, since all individuals not dead by time L must die sometime in I_{k+1}. We now define the following quantities:

N_j = Number of individuals at risk (i.e., alive and not censored) at time a_{j-1}

D_j = Number of deaths in (i.e., number of lifetimes observed to fall into) $I_j = [a_{j-1}, a_j)$

W_j = Number of withdrawals in (i.e., number of censoring times observed to fall into) $I_j = [a_{j-1}, a_j)$.

The terms "at risk," "deaths," and "withdrawals" are commonly used with life tables, though sometimes other terms are used, such as number of censoring times in I_j instead of number of withdrawals in I_j. The number of individuals known to be alive at the start of I_j is N_j, and thus $N_1 = n$ and

$$N_j = N_{j-1} - D_{j-1} - W_{j-1} \qquad j = 2, \ldots, k+1.$$

Let the distribution of lifetimes for the population under study have survivor function $S(t)$, and define the following quantities:

$$\begin{aligned} P_j &= S(a_j) \\ p_j &= Pr(\text{an individual survives beyond } I_j | \text{ they survive beyond } I_{j-1}) \\ &= \frac{P_j}{P_{j-1}} \\ q_j &= 1 - p_j \\ &= Pr(\text{an individual dies in } I_j | \text{ they survive beyond } I_{j-1}). \end{aligned} \tag{3.6.1}$$

In (3.6.1) j ranges over $1, \ldots, k+1$, with $P_0 = 1$. Note also that $P_{k+1} = 0$, $q_{k+1} = 1$, and that

$$P_j = p_1 p_2 \cdots p_j \qquad j = 1, \ldots, k+1. \tag{3.6.2}$$

This result, which is analogous to (1.2.8) for discrete lifetime distributions, gives the probability of surviving past I_j as the product of conditional probabilities of surviving past intervals up to I_j, and is the basis for life table methodology. Life table analysis involves estimating the q_j and p_j of (3.6.1), and then via (3.6.2) the P_j. It is a precursor of, and the inspiration for, the product-limit estimator of Section 3.2. The term life table refers to the format in which the estimates are portrayed.

The standard procedure is as follows: if a particular interval I_j has no withdrawals in it (i.e., $W_j = 0$), then a sensible estimate of q_j is $\hat{q}_j = D_j/N_j$, since q_j is the con-

ditional probability of an individual dying in I_j, given that they are alive at the start of I_j. If, however, the interval has $W_j > 0$ withdrawals, D_j/N_j might be expected to underestimate q_j, since some of the individuals censored in I_j might have died before the end of I_j, had they not been censored first. It is therefore desirable to make some adjustment for the censored individuals. The most commonly used procedure is to estimate q_j by the so-called standard life table estimate, which is

$$\hat{q}_j = \frac{D_j}{N_j - W_j/2} = \frac{D_j}{N_j'}. \tag{3.6.3}$$

The expression (3.6.3) assumes that $N_j > 0$; when $N_j = 0$, there is no information about q_i for $i \geq j$. The denominator $N_j' = N_j - .5W_j$ can be thought of as an effective number of individuals at risk for the interval I_j; this supposes that, in a sense, a withdrawn individual is at risk for half the interval. This adjustment is somewhat arbitrary, but sensible in many situations. Its appropriateness depends on the failure and censoring-time process, of course. In some instances other estimates of q_j may be preferable. For example, if all withdrawals in I_j occurred right at the end of I_j, the estimate $\hat{q}_j = D_j/N_j$ would be appropriate, whereas if all withdrawals occurred at the beginning of I_j, $\hat{q}_j = D_j/(N_j - W_j)$ would be appropriate. Still other estimates are useful on certain occasions; we return to this point in Section 3.6.2. The present section focuses on the standard life table estimate.

Once estimates $\hat{q}_j$ and $\hat{p}_j = 1 - \hat{q}_j$ have been calculated, P_j can, by virtue of (3.6.2), be estimated by

$$\hat{P}_j = \hat{p}_1 \cdots \hat{p}_j \qquad j = 1, \ldots, k+1.$$

The life table itself is a display of the data and the estimates $\hat{q}_j$ and $\hat{P}_j$. It generally includes columns giving, for each interval, the values of N_j, D_j, W_j, $\hat{q}_j$, and $\hat{P}_j$. Additional columns are sometimes included, giving quantities such as N_j', $\hat{p}_j$, and, occasionally, estimates of other characteristics of the underlying distribution. The general format is exemplified in Table 3.9.

Example 3.6.1. Berkson and Gage (1950) gave data describing the survival times from surgery of a group of 374 patients who underwent operations in connection with a type of malignant disease. From these data the life table given in Table 3.9 has been formed. The intervals I_j used in the table were chosen for convenience.

The estimates $\hat{q}_j$, $\hat{p}_j$, and $\hat{P}_j$ are subject to sampling variation. Under suitable assumptions it is possible to derive estimates of their variances. Details related to this are discussed in Section 3.6.2; here we present the commonly used estimates, which were suggested by Greenwood (1926). In this case $\hat{q}_j\hat{p}_j/N_j'$ estimates the variance of $\hat{q}_j$ (or $\hat{p}_j$), the $\hat{p}_j$ are asymptotically uncorrelated, and an approximation to the variance of $\hat{P}_j = \hat{p}_1 \cdots \hat{p}_j$ is then derived using Theorem B1 (Appendix B), giving

Table 3.9. Life Table Computed from Data in Berkson and Gage (1950)

Interval (I_j) in Years	D_j	W_j	N_j	N_j'	$\hat{q}_j$	$\hat{p}_j$	$\hat{P}_j$
[0, 1)	90	0	374	374.	0.241	0.759	.759
[1, 2)	76	0	284	284.	0.268	0.732	.556
[2, 3)	51	0	208	208.	0.245	0.755	.420
[3, 4)	25	12	157	151.	0.166	0.834	.350
[4, 5)	20	5	120	117.5	0.170	0.830	.291
[5, 6)	7	9	95	90.5	0.077	0.923	.268
[6, 7)	4	9	79	74.5	0.054	0.946	.254
[7, 8)	1	3	66	64.5	0.016	0.984	.250
[8, 9)	3	5	62	59.5	0.050	0.950	.237
[9, 10)	2	5	54	51.5	0.039	0.961	.228
[10, ∞)	47	0	47	47.	1.000	0.000	0

$$\widehat{\mathrm{Var}}(\hat{P}_j) = \hat{P}_j^2 \sum_{i=1}^{j} \frac{\hat{q}_i}{N_i'\hat{p}_i}. \tag{3.6.4}$$

In Example 3.6.1, estimates of quantities such as 5-year survival probabilities would be of interest. In the example this is $\hat{P}_5 = .291$; the estimated variance of $\hat{P}_5$ given by (3.6.4) is $.024^2$.

The estimates $\hat{P}_j$ and variance estimates (3.6.4) have the same form as the Kaplan–Meier estimate (3.2.2) and the variance estimate (3.2.3), with N_i' taking the place of n_i. Indeed, the Kaplan–Meier (or PL) estimate and (3.2.3) were first obtained by considering the life table estimates $\hat{P}_j$ in the limit where k becomes large and interval lengths approach 0. Confidence intervals for $P_j = S(a_j)$ can be based on the procedures described for $S(t)$ in Section 3.2.3, the simplest approach being to treat $Z_1 = (\hat{P}_j - P_j)/\widehat{\mathrm{Var}}(\hat{P}_j)^{1/2}$ as a standard normal pivotal quantity.

It can be noted that (3.6.4) gives the usual binomial estimate for the variance of $\hat{P}_j$ in the case in which there is no censoring, just as (3.2.3) does. To see this, note that when there are no withdrawals, $N_i' = N_i = n\hat{P}_{i-1}$ and $N_i\hat{p}_i = n\hat{p}_i\hat{P}_{i-1} = n\hat{P}_i$. Therefore (3.6.4) equals

$$\hat{P}_j^2 \sum_{i=1}^{j} \frac{\hat{q}_i}{n\hat{P}_i} = \frac{\hat{P}_j^2}{n} \sum_{i=1}^{j} \frac{1-\hat{p}_i}{\hat{P}_i} = \frac{\hat{P}_j^2}{n} \sum_{i=1}^{j} \left(\frac{1}{\hat{P}_i} - \frac{1}{\hat{P}_{i-1}} \right)$$
$$= \frac{\hat{P}_j\left(1-\hat{P}_j\right)}{n} \tag{3.6.5}$$

which is the usual estimate of $\mathrm{Var}(\hat{P}_j) = P_j(1-P_j)/n$.

In forming a life table, there is no need to make the intervals of equal length, though it may be convenient to do so. The number of intervals used depends on the amount of data available and on the aims of the analysis. Certain statistical properties of the estimates are enhanced when the number of intervals is fairly large, as described in the next section. On the other hand, if an easily comprehended summary of the data is wanted, it may be sensible to have as few as 8–10 intervals.

Life tables are primarily used with large bodies of data or when lifetimes and censoring times are available only in grouped form. If exact times are available, then even with large data sets it is often preferable to summarize data or estimate survival probabilities using PL estimates. An important point, discussed in the next section, is that the validity of life table methods depends on individuals who are censored in an interval having the same lifetime distribution over the interval as those who are not. This is related to earlier discussions concerning interval censoring in Section 2.3.1, and is a concern in some settings.

Let us now consider some underlying theory for life tables.

3.6.2 Theory for Life Table Methodology

The theoretical analysis of life methodology requires assumptions about the censoring process. The situation is analogous to that for interval censoring, discussed in Section 2.3.1. With life tables, however, the intervals are usually fixed, but censoring times occurring within intervals is the norm. This makes the censoring process nonignorable, except in very special circumstances.

Let us start by considering the special case where all censoring or withdrawals occur at the ends of intervals I_j; this also includes the case where there are no withdrawals at all. The observed data from a cohort of n individuals consist of $D_1, W_1, \ldots, D_k, W_k$, with $D_{k+1} = N_{k+1} = n - D_1 - W_1 - \cdots - W_k$. This is formally the same as the setup for right censoring in a discrete-time model, discussed in Section 2.2.2, and as in (2.2.9) we have the decomposition of $Pr(D_1, W_1, \ldots, D_k, W_k)$ as

$$\prod_{j=1}^{k} Pr(D_j|\mathcal{H}(j))Pr(W_j|D_j, \mathcal{H}(j)), \tag{3.6.6}$$

where $\mathcal{H}(j) = (D_1, W_1, \ldots, D_{j-1}, W_{j-1})$, with $\mathcal{H}(1)$ empty. To proceed further we assume that for $j = 1, \ldots, k$

$$Pr(D_j|\mathcal{H}(j)) \sim \text{Binomial}(N_j, q_j). \tag{3.6.7}$$

This is analogous to assumption (2.3.2) plus independence of individuals in the case of interval censoring, and it says that the distribution of D_j, conditional on the deaths and withdrawals up to a_{j-1}, is precisely as if there were no censoring process. The assumption (3.6.7) thus says that the censoring process is independent and hence ignorable. It is not reasonable when withdrawals can occur anywhere in

$I_j = (a_{j-1}, a_j]$, as we discuss below, but is reasonable under the special scheme being considered here.

Maximum likelihood proceeds as for discrete-time models in Section 2.2.2. If the terms $Pr(W_j|D_j, \mathcal{H}(j))$ do not include information about the parameters $q_1, \ldots, q_k$, then they can be dropped from the likelihood given by (3.6.6) and (3.6.7) to give

$$L(q_1, \ldots, q_k) = \prod_{j=1}^{k} \binom{N_j}{D_j} q_j^{D_j}(1-q_j)^{N_j - D_j}. \tag{3.6.8}$$

The m.l.e.'s are easily found to be $\hat{q}_j = D_j/N_j$, and the information matrix $I(\mathbf{q})$ is diagonal with entries

$$I(\mathbf{q})_{jj} = \frac{D_j}{q_j^2} + \frac{(N_j - D_j)}{(1-q_j)^2}, \qquad j = 1, \ldots, k.$$

The expected information matrix can be obtained by using conditional expectation,

$$\begin{aligned} \mathcal{I}(\mathbf{q})_{jj} &= E_{N_j} E\{I(\mathbf{q})_{jj}|N_j\} \\ &= \frac{E(N_j)}{q_j(1-q_j)}. \end{aligned} \tag{3.6.9}$$

Either $I(\hat{\mathbf{q}})^{-1}$ or $\mathcal{I}(\hat{\mathbf{q}})^{-1}$ with $E(N_j)$ estimated by N_j gives an estimate of the asymptotic covariate matrix for $\hat{\mathbf{q}}$ as diagonal with entries

$$\widehat{\text{Asvar}}(\hat{q}_j) = \frac{\hat{q}_j(1-\hat{q}_j)}{N_j}. \tag{3.6.10}$$

Since $P_j = (1-q_1)\cdots(1-q_j)$, we get

$$\widehat{\text{Asvar}}(\hat{P}_j) = \hat{P}_j^2 \sum_{i=1}^{j} \frac{\hat{q}_i}{N_i(1-\hat{q}_i)} \tag{3.6.11}$$

by straightforward application of the formula (B2) in Appendix B. This is the same as (3.6.4) in this case.

Asymptotics assume that $E(N_j) \to \infty$ as $n \to \infty$ for $j = 1, \ldots, k$. Some small-sample calculations are possible, given the withdrawal mechanism; we note only that $E(\hat{q}_j) = q_j$ if $Pr(N_j > 0) = 1$, which follows from the fact that $E(\hat{q}_j|N_j) = q_j$ if $N_j > 0$, and that the $\hat{q}_j$ are uncorrelated. The latter follows from the fact that if $j > \ell$,

$$E[(\hat{q}_\ell - q_\ell)(\hat{q}_j - q_j)] = E[(\hat{q}_\ell - q_\ell)E\{(\hat{q}_j - q_j)|\mathcal{H}(j)\}] = 0.$$

Consider now the general case, where withdrawals may occur throughout the intervals I_j. In this case, withdrawals operate as a competing risk for death, as

described in Section 1.5, and it is not possible to avoid modeling the withdrawals process. The standard life table methods of Section 3.6.1 were developed heuristically, so it is of interest to examine their properties under plausible models. In addition, modeling may be used to suggest alternative procedures.

Assume that the underlying distribution of lifetime, T, is continuous with s.f., $S(t)$, and hazard function, $h(t)$. If an individual is alive and uncensored at a_{j-1}, we assume that the hazard function for death that operates over $(a_{j-1}, a_j]$ is $h(t)$. This assumption is unverifiable using only data D_j, W_j $(j = 1, \ldots, k)$ and must be based on background knowledge about the censoring process. Even with this assumption, (3.6.7) does not hold, however, since an individual may be censored in I_j before they can be observed to die. We can view this as a competing failure-modes problem (Section 1.5), where there is a mode-specific hazard function, $h_W(t)$, for withdrawal, as in (1.5.2). We make the additional assumption that the death and censoring processes operate independently over I_j, given survival to a_{j-1}, so that $h(t)$ is the hazard function for mode death.

Breslow and Crowley (1974) investigated the standard life table methodology under such a random censorship model. They considered the special case that each individual $i = 1, \ldots, n$ has a random censoring or withdrawal time, C_i, with s.f., $G(t)$, with $T_1, \ldots, T_n, C_1, \ldots, C_n$ mutually independent. We will outline their results.

Let π_j^D be the probability an individual is observed to die in I_j and let π_j^W be the probability they are observed to be withdrawn. For example,

$$\begin{aligned}\pi_1^D &= Pr(\text{an individual dies in } I_1 \text{ and is observed to do so}) \\ &= Pr(T_i \le a_1, T_i \le C_i) \\ &= \int_0^{a_1} G(x)|dS(x)|.\end{aligned}$$

We find in general that for $j = 1, \ldots, k$

$$\pi_j^D = \int_{a_{j-1}}^{a_j} G(x)|dS(x)|$$

$$\pi_j^W = \int_{a_{j-1}}^{a_j} S(x)|dG(x)|,$$

where for generality we use Riemann–Stieltjes integrals. Since $\mathbf{D} = (D_1, W_1, \ldots, D_k, W_k)'$ is multinomial, it follows that as $n \to \infty$, the distribution of $n^{-1/2}(\mathbf{D} - n\boldsymbol{\pi})$ converges to a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\sum = \text{diag}(\pi_1^D, \pi_1^W, \ldots, \pi_{k+1}^N) - \boldsymbol{\pi}\boldsymbol{\pi}'$ (e.g., Bishop et al., 1975, p. 470). The standard life table estimates $\hat{q}_j = D_j/(N_j - .5W_j)$ are smooth functions, and hence the distribution of $\sqrt{n}(\hat{\mathbf{q}} - \mathbf{q}^*)$ also converges to a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\sum_q$, say, where $\hat{\mathbf{q}} = (\hat{q}_1, \ldots, \hat{q}_k)$ and $\mathbf{q}^* = (q_1^*, \ldots, q_k^*)$ is the probability limit for $\hat{\mathbf{q}}$ in large samples. Since

$$\hat{q}_j = \frac{D_j/n}{N_j/n - W_j/2n},$$

it follows that

$$q_j^* = \frac{\pi_j^D}{\pi_j^N - \pi_j^W/2},$$

where $\pi_j^N = E(N_j/n) = G(a_{j-1})S(a_{j-1})$. Thus

$$q_j^* = \left(\int_{a_{j-1}}^{a_j} G(x)|dS(x)|\right) \Big/ \left(G(a_{j-1})S(a_{j-1}) - \frac{1}{2}\int_{a_{j-1}}^{a_j} S(x)|dG(x)|\right). \tag{3.6.12}$$

In general q_j^* does not equal

$$q_j = \frac{S(a_{j-1}) - S(a_j)}{S(a_{j-1})},$$

so the standard life table estimate (3.6.3) is not a consistent estimate of q_j and $\hat{P}_j = \hat{p}_1 \cdots \hat{p}_j$ is not a consistent estimate of $P_j = p_1 \cdots p_j$. An important practical question is whether the asymptotic bias in the estimates is sufficiently small to render this inconsistency relatively harmless. It appears that this is in fact the case in many situations, as we shall see momentarily.

The entries in the asymptotic covariance matrix $\sum_q$ of $\hat{\mathbf{q}}$ can be determined by a straightforward application of the multivariate delta formula (B5). Using this (also see Problem 3.16), we find that $\sum_q$ is a diagonal matrix, and hence $\hat{q}_j$ and $\hat{q}_\ell$ $(j \ne \ell)$ are asymptotically uncorrelated, just as in the model with censoring only at interval endpoints. The asymptotic variance of $\sqrt{n}(\hat{q}_j - q_j^*)$ turns out to be

$$\text{Asvar}\left[\sqrt{n}(\hat{q}_j - q_j^*)\right] = \frac{q_j^* - q_j^{*2}[(\pi_j^N - \pi_j^W/4)/(\pi_j^N - \pi_j^W/2)]}{\pi_j^N - \pi_j^W/2}. \tag{3.6.13}$$

The life table estimate of $\text{Var}(\hat{q}_j)$ used in (3.6.4) is

$$\widehat{\text{Var}}(\hat{q}_j) = \frac{\hat{q}_j - \hat{q}_j^2}{N_j'}, \tag{3.6.14}$$

and is based on the heuristic replacement of N_j with N_j' in (3.6.10). If q_j and q_j^* are not too different, this tends to overestimate the true variance somewhat, since N_j'/n converges in probability to the denominator of (3.6.13), and the term in square brackets in (3.6.13) is less than unity. If q_j is small, the second terms in (3.6.13) and

(3.6.14) are small relative to the first and the agreement between the two formulas is improved.

The limiting distribution of the $\sqrt{n}(\hat{P}_j - P_j)$ is multivariate normal, with means, variances, and covariances that can be determined by (B5). Let $\hat{\mathbf{P}} = (\hat{P}_1, \ldots, \hat{P}_k)$ and $\mathbf{P}^* = (P_1^*, \ldots, P_k^*)$, where $F_j^* = p_1^* \cdots p_j^*$; the limiting distribution of $\sqrt{n}(\hat{\mathbf{P}} - \mathbf{P}^*)$ is multivariate normal with mean $\mathbf{0}$ and a covariance matrix whose (j, ℓ) term is, for $j \le \ell$,

$$P_j^* P_\ell^* \sum_{i=1}^{j} \frac{\mathrm{Var}[\sqrt{n}(\hat{q}_i - q_i^*)]}{(1 - q_i^*)^2}. \tag{3.6.15}$$

Putting $j = \ell$, using (3.6.14), and replacing P_j^* and q_i^* with $\hat{P}_j$ and $\hat{q}_i$, we get from (3.6.15) the estimate

$$\widehat{\mathrm{Var}}(\hat{P}_j) = \hat{P}_j^2 \sum_{i=1}^{j} \frac{\hat{q}_i - \hat{q}_i^2}{(1 - \hat{q}_i)^2 N_i'}$$

$$= \hat{P}_j^2 \sum_{i=1}^{j} \frac{\hat{q}_i}{N_i' \hat{p}_i},$$

which is Greenwood's formula (3.6.4).

Broadly speaking, the standard life table estimates are acceptable under random independent censorship provided that censoring is fairly evenly distributed across individual intervals and not too heavy. It helps if the intervals are not too wide. It is nevertheless wise to remember that estimates of survival probabilities are generally biased, as is the variance estimate (3.6.4). Alternative variance estimates can be based on (3.6.13), with $\hat{q}_j$, N_j/n, and W_j/n estimating q_j^*, π_j^N, and π_j^W, respectively.

The guidelines suggested by the random-censorship model given earlier should also be reasonable under broader conditions in which censoring within an interval, I_j, is independent of lifetimes, conditional on being alive and uncensored at the start of I_j. If censoring and failure are not more or less uniform over the interval I_j, an estimator other than (3.6.3) may be preferred; examples are when withdrawals occur mainly at the beginning or the end of an interval. If we have some idea of the shapes of the hazard functions for death and censoring in I_j, then this may provide some guidance, following essentially the same analysis as for the random censoring model. Chapter 9 on competing risks contains further information.

BIBLIOGRAPHIC NOTES

The PL estimate $\hat{S}(t)$ of a survivor function from right-censored data was considered in the actuarial literature in the early 1900s, but the modern treatment began with Kaplan and Meier (1958). They obtained $\hat{S}(t)$ as a nonparametric m.l.e. and gave several fundamental results. Large-sample properties were considered by Efron

(1967), Breslow and Crowley (1974), Meier (1975), Peterson (1977), and Winter et al. (1978), and discussions of $\hat{S}(t)$ as an m.l.e. were given by Aalen and Johansen (1978), Johansen (1978), and Scholz (1980).

The Nelson–Aalen estimate $\hat{H}(t)$ was introduced independently by Nelson (1969) and Altschuler (1970), the latter for the competing risks setting. Aalen (1976, 1978a,b) introduced the estimate for intensity functions in Markov counting process models and began the modern study of $\hat{H}(t)$ and $\hat{S}(t)$ by martingale methods. Watson and Leadbetter (1964a,b) considered an earlier version of "hazard analysis."

The early work on $\hat{S}(t)$ and $\hat{H}(t)$ provided methods of confidence interval estimation based on asymptotic standard errors; methods based on (3.2.19) have been used for a long time, and empirical studies have been carried out by Link (1984), Bie et al. (1987), Klein (1991), and others. The idea of using the likelihood ratio statistic (3.2.22) to obtain confidence intervals for $S(t)$ is due to Thomas and Grunkemeier (1975); see also Matthews (1988). Brookmeyer and Crowley (1982) discussed the method of obtaining confidence intervals for quantiles given in Section 3.2.3. Nair (1984) and Hollander and Pena (1989) review the construction of confidence bands for $S(t)$. Estimation of mean lifetimes from censored data is considered by Yang (1977), Susarla and van Ryzin (1980), and Gill (1983). The use of the nonparametric bootstrap for right-censored data was introduced by Efron (1981), and this provides an alternative way to obtain confidence intervals; Davison and Hinkley (1997, Ch. 3) and Strawderman and Wells (1997) review this area. Bayesian nonparametric estimation of survivor and cumulative hazard functions has been considered by Susarla and van Ryzin (1976), Kalbfleisch (1978), Ferguson and Phadia (1979), and Hjort (1990b).

There is by now a large literature on theory associated with the Kaplan–Meier and Nelson–Aalen estimators. Detailed reviews and references are provided by Fleming and Harrington (1991) and Andersen et al. (1993).

Probability plots of censored lifetime data were once widely used for estimation as well as for description and model checking; Barnett (1975), Nelson (1982) and D'Agostino and Stephens (1986, Ch. 2, p. 11) discuss classic probability plots. Cox (1978) and Cleveland (1985) consider some general aspects of graphical methods. Gentleman and Crowley (1991) discuss various aspects of graphics with censored data.

Nonparametric estimation of hazard functions using smoothing methodology was introduced by Watson and Leadbetter (1964a,b) and Rice and Rosenblatt (1976). They used kernel smoothers, and this work was subsequently extended to handle censoring by Ramlau-Hansen (1983), Tanner and Wong (1983, 1984, 1987) and Yandell (1983). Andersen et al. (1993, pp. 324–326) survey more recent work on kernel methods. Penalized maximum likelihood methods featured in Section 3.4 were introduced by Anderson and Senthilselvan (1980); see also O'Sullivan (1988). Regression spline methods are considered by Kooperberg and Stone (1992), Abrahamowicz et al. (1992), and Rosenberg (1995). Efron (1988) and Muller et al. (1997) look at smooth hazard estimates from grouped data. Bacchetti (1990) and Tutz and Pritscher (1996) use penalized likelihood and kernel estimation, respectively, for estimation of discrete-time hazard functions.

Kaplan and Meier (1958) discussed the extension of the PL estimate to handle left truncation, and Lynden-Bell (1971) discussed right-truncation. Theoretical properties were studied subsequently by Woodroofe (1985), Wang et al. (1986), and Keiding and Gill (1990). Tsai (1990) and Kalbfleisch and Lawless (1991) discussed tests for the independent truncation assumption. Efron and Petrosian (1999) consider data that are simultaneously left and right truncated. Nonparametric estimation of the survival function from interval-censored data was considered by Peto (1973) and Turnbull (1976), who considered both truncation and interval censoring (see also Frydman 1994). Groeneboom and Wellner (1992), Gentleman and Geyer (1994), Böhning et al. (1996), and Wellner and Zhan (1997) discuss computational procedures and properties of the nonparametric estimator. Huang and Wellner (1997) review the area, and Lindsey and Ryan (1998) discuss practical matters, including regression modeling. The special case of current-status data has received considerable additional attention. Jewell and Shiboski (1990), Diamond and McDonald (1992), Keiding (1991), and Groeneboom and Wellner (1992) discuss examples and numerous results; Jewell and van der Laan (1997) review this topic. Problems involving double censoring have been considered by J. Sun (1995, 1997), Jewell and van der Laan (1997), and others. Alternatives to strict nonparametric estimation are often attractive with interval censoring. Approaches such as those mentioned in Section 3.4 have been considered by Tanner and Wong (1987), Bacchetti (1990), Keiding et al. (1996), Joly et al. (1998), Lindsey and Ryan (1998), Kooperberg and Clarkson (1997), Betensky et al. (1999) and Duchesne and Stafford (2002).

In some applications censoring times may be missing for some or all individuals whose lifetimes are censored, in which case there may also be supplementary follow-up of certain individuals. Suzuki (1985ab, 1995), Kalbfleisch and Lawless (1988ab), Hu and Lawless (1996), and Hu et al. (1998) discuss this area. See also Problem 3.11.

The life table methods of Section 3.6.1 date from the 1600s, with considerable theoretical development in the 1900s (see, e.g., Namboodiri and Suchindran 1987). The variance estimate (3.6.4) for estimators of survival was given by Greenwood (1926). Theoretical studies include those of Littell (1952), Kuzma (1967), Breslow and Crowley (1974), and Drolette (1975). Alternative estimators, designed to reflect specific types of withdrawal patterns within life table intervals have been considered by Sacher (1956), Elveback (1958), Chiang (1960a,b), and many others. Comparisons of methods (e.g., Elandt-Johnson 1977, Johnson 1977) suggest there is little difference among the various estimators unless withdrawals are rather heavy. Elandt-Johnson and Johnson (1980) provide a detailed treatment and references on life table methodology.

COMPUTATIONAL NOTES

Several packages provide inferences and plots based on the Kaplan–Meier and Nelson–Aalen estimates; in S-Plus see function survfit and in SAS, the LIFETEST procedure. S-Plus function kaplanMeier handles interval-censored data, though the S-Plus 2000 version occasionally returns a point that is not the m.l.e.; this occurs with the data in Example 3.5.5. Plots of $\hat{S}(t)$ are also portrayed as step functions, without an indication that the estimate is undefined over certain intervals. Left truncation is not handled by the S-Plus 2000 functions mentioned, but function coxph can be coerced to handle this. Therneau and Grambsch (2000, Ch. 2 and Appendix A) and Venables and Ripley (1999, Ch. 12) provide useful information on S-Plus capabilities; the former also discusses survival methods in SAS. Collett (1994, Ch. 11) provides a review of software for survival analysis.

Software for obtaining smooth estimates of hazard or density functions from censored data is available on Web sites such as statlib. In addition, S-Plus has a variety of smoothers and spline functions that can be adapted as in Section 3.4.

PROBLEMS AND SUPPLEMENTS

3.1 The data below are remission times, in weeks, for a group of 30 patients with leukemia who received similar treatment. Asterisks denote censoring times.

1, 1, 2, 4, 4, 6, 6, 6, 7, 8, 9, 9, 10, 12, 13, 14, 18, 19, 24, 26, 29, 31*, 42, 45*, 50*, 57, 60, 71*, 85*, 91.

(a) Obtain and plot the Kaplan–Meier estimate $\hat{S}(t)$ of the survivor function for remission time.

(b) Obtain approximate .95 confidence intervals for the median remission time and for the probability that remission lasts over 26 weeks.

(c) Plot $\log(-\log \hat{S}(t))$ and $\log \hat{H}(t)$ on the same graph, where $\hat{H}(t)$ is the Nelson–Aalen estimate. Is there much difference?

(Section 3.2)

3.2 The data below show survival times (in months) of patients with Hodgkin's disease who were treated with nitrogen mustards (Bartolucci and Dickey, 1977). Group A patients received little or no prior therapy, whereas Group B patients received heavy prior therapy. Starred observations are censoring times.

Group A 1.25, 1.41, 4.98, 5.25, 5.38, 6.92, 8.89, 10.98, 11.18, 13.11, 13.21, 16.33, 19.77, 21.08, 21.84*, 22.07, 31.38*, 32.62*, 37.18*, 42.92.

Group B 1.05, 2.92, 3.61, 4.20, 4.49, 6.72, 7.31, 9.08, 9.11, 14.49*, 16.85, 18.82*, 26.59*, 30.26*, 41.34*.

(a) Obtain and compare Kaplan–Meier estimates for the two groups. Does there appear to be a difference in the 1-year survival probability for the two types of patients? Give confidence limits for $S(1)$ and for the median survival time $t_{.50}$ for each group.

(b) Use plots of the Nelson–Aalen estimate $\hat{H}(t)$ to examine and compare the two life distributions.

(c) Do any parametric models whereby one might compare the two distributions suggest themselves?

(Sections 3.2, 3.3)

3.3 *Precision of nonparametric estimation.* Consider the random censorship model and the asymptotic variance formula (3.2.33). Use this to examine the asymptotic variance $v_n(t)^2$ of $\sqrt{n}[(\hat{S}(t) - S(t))/S(t)]$ when $S(t) = \exp(-t)$ and the censoring time C is uniform on $(1, 3)$. Plot $v_n(t)$ vs. t for $0 < t < 3$.

(Section 3.2)

3.4 Consider the data given in Example 1.1.5 of Chapter 1 concerning the failure times of electrical insulation specimens subjected to a constant voltage stress. Make Weibull probability plots of the data for the experiments run at 28, 30, and 32 kV, respectively.

(a) Does the suggestion of a Weibull failure time distribution for each situation with the shape parameters, but not the scale parameters, having the same value in the three cases seem plausible?

(b) Compute graphical estimates of parameters in the Weibull models and compare the estimated s.f. from these with the empirical s.f.'s (PL estimates).

(Section 3.3)

3.5 Pike (1966) gave results of a laboratory experiment in which 19 female rats were painted with the carcinogen DMBA. The number of days T until the appearance of a carcinoma was of interest, and the data gave the following times (asterisks denote censoring times):

143, 164, 188, 188, 190, 192, 206, 209, 213, 216, 220, 227, 230, 234, 246, 265, 304, 216^*, 244^*

(a) It was thought that carcinomas could not appear before some threshold time $\gamma > 0$, so a Weibull model for $T' = T - \gamma$ was considered. Give two Weibull probability plots, using (1) the raw data (t-values), and (2) the values $t' = t - 100$. Is there any strong indication that $T - 100$ is closer to Weibull-distributed than is T?

(b) Obtain a nonparametric .95 confidence interval for the median time to carcinoma, $t_{.50}$. Comment on the advantages and disadvantages of this estimate over one based on a Weibull model.

(Sections 3.2, 3.3)

3.6 *Mean lifetime.* Consider the mean lifetime μ_τ restricted to τ, defined by (3.2.37) in Section 3.2.4, and the associated estimate $\hat{\mu}_\tau$.

(a) Using the results in Section 3.2.4, show heuristically that $\sqrt{n}(\hat{\mu}_\tau - \mu_\tau)$ is asymptotically normal with variance given by (3.2.38).

(b) Use the result of part (a) to motivate the variance estimate

$$\widehat{\mathrm{Var}}(\hat{\mu}_\tau) = \sum_{j:t_j \le \tau} \frac{\hat{A}_j^2 d_j}{n_j(n_j - d_j)},$$

where

$$\hat{A}_j = (t_{j+1} - t_j)\hat{S}(t_{j+1}) + (t_{j+2} - t_{j+1})\hat{S}(t_{j+2}) + \cdots + (\tau - t_m)\hat{S}(\tau),$$

with t_m being the largest observed lifetime less than or equal to τ.

(c) In the special case in which there is no censoring possible, let $\tau \to \infty$ and show that $\hat{\mu} = \hat{\mu}_\infty$ reduces to $\bar{t} = \sum t_i/n$ and that (3.2.38) reduces to $\sigma^2 = \mathrm{Var}(t_i)$, where it is assumed that $\mathrm{Var}(t_i)$ exists.

(Section 3.2; Kaplan and Meier, 1958)

3.7 *Mean residual lifetime.* Recall the definition of the mean residual life function $m(t)$, given in Problem 1.1. Its estimation is of interest in areas such as demography and insurance.

(a) Show that if $E(T) = m(0)$ exists, then it equals $\int_0^\infty S(t)\,dt$, and that $m(t)$ exists for $t > 0$ if $m(0)$ exists.

(b) Assume that $S(t) = 0$ for $t \ge M$, where M is some upper limit on lifetime. Describe a procedure for nonparametric estimation of $m(t)$, for a specified value $t > 0$. Use the results of Problem 3.6 and Section 3.2.4 to provide confidence intervals. For convenience, assume that if the largest time in your sample is a censoring time, then it is always greater than M.

(Section 3.2)

3.8 *Piecewise exponential models.* Consider the piecewise exponential model specified by (1.3.25) in Section 1.3.8.

(a) Write down the likelihood function for the parameters $\lambda_1, \ldots, \lambda_m$, based on a sample of failure times and censoring times obtained under an independent censoring mechanism. Estimate $h(t)$ and $S(t)$, showing that the estimate of λ_m is given by (3.4.3). What does the estimate of $S(t)$ tend to as m increases and the $|a_j - a_{j-1}|$ in (1.3.25) become small?

(b) Apply this to the data in Problem 3.1 using intervals of length 5 in (1.3.25), except with the last interval being $[80, \infty)$. Compare $\hat{S}(t)$ graphically with the Kaplan–Meier estimate. Repeat the procedure using intervals of length 1 up to $t = 80$.

(Sections 3.2, 3.4)

3.9 Let $X_{(i)}$ be the ith-order statistic in a random sample of size n from a continuous distribution with p.d.f. $f(x)$ and let x_p denote the pth quantile of the distribution $(0 < p < 1)$. Let $n \to \infty$ and $i \to \infty$ in such a way that $i/n \to p$.

Show that in large samples $X_{(i)}$ can be considered to be approximately normally distributed with mean x_p and variance

$$p(1-p)/\left[nf(x_p)^2\right].$$

What are the implications of this with regard to variation in classic probability plots?

(Section 3.3, Appendix B)

3.10 For the pulmonary exacerbation data of Example 3.2.4, apply the methods of Section 3.4 to obtain nonparametric estimates of the hazard functions for the two treatment groups, ignoring the fev covariate. See Appendix G to obtain the data.

(Section 3.4)

3.11 *Missing censoring times.* In some applications the censoring times C_i for censored units are missing. That is, we observe $T_i = t_i$ if $\delta_i = 1$, but $t_i (= C_i)$ is missing if $\delta_i = 0$. Suppose that the independent random censorship model of Section 2.2.1b holds, so that the likelihood function takes the form

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} Pr(T_i > C_i)^{1-\delta_i}. \tag{3.7.1}$$

(a) Assuming that the s.f. $G(c)$ for the C_i is known, consider the discrete-time framework where T and C can take on values $1, 2, 3, \ldots$. Let $c_{\max} = \sup\{c : G(c) > 0\}$ be finite. Show that (3.7.1) is maximized for $\{f(t), t = 1, \ldots, c_{\max} : f(t) \geq 0\}$ by

$$\tilde{f}(t) = \frac{d_t}{nG(t)}, \tag{3.7.2}$$

where $d_t = \sum I(t_i = t, \delta_i = 1)$. This also maximizes (3.7.1) subject to $\sum_{t=1}^{c_{\max}} f(t) \leq 1$, if $\sum_{t=1}^{c_{\max}} \tilde{f}(t) \leq 1$.

(b) Motivate (3.7.2) as a moment estimator by considering $E(d_t)$. Obtain the variance of $\tilde{F}(t) = \tilde{f}(1) + \cdots + \tilde{f}(t)$ by noting that $(d_1, \ldots, d_t)$ follows a multinomial distribution.

(Section 3.2, Hu et al. 1998)

3.12 For the left-truncated data in Table 3.5, estimate and plot the conditional survivor function $S(t|T \geq 26.2)$ by dropping the first pair $(u_1, t_1) = (19.6, 24.5)$ from the data.

(Section 3.5.1)

3.13 Finkelstein (1986) and Lindsey and Ryan (1998) discussed interval-censored data from a study of patients with breast cancer. The response variable of interest was the time, T, to cosmetic deterioration of the breast, and whether there

Table 3.10. Interval Censored Times to Cosmetic Deterioration

Radiotherapy			Radiotherapy and Chemotherapy		
(45, ∞]	(25, 37]	(37, ∞]	(8,12]	(0, 5]	(30, 34]
(6, 10]	(46, ∞]	(0, 5]	(0, 22]	(5, 8]	(13, ∞]
(0, 7]	(26, 40]	(18, ∞]	(24, 31]	(12, 20]	(10, 17]
(46, ∞]	(46, ∞]	(24, ∞]	(17, 27]	(11, ∞]	(8, 21]
(46, ∞]	(27, 34]	(36, ∞]	(17, 23]	(33, 40]	(4, 9]
(7, 16]	(36, 44]	(5, 11]	(24, 30]	(31, ∞]	(11, ∞]
(17, ∞]	(46, ∞]	(19, 35]	(16, 24]	(13, 39]	(14, 19]
(7, 14]	(36, 48]	(17, 25]	(13, ∞]	(19, 32]	(4, 8]
(37, 44]	(37, ∞]	(24, ∞]	(11, 13]	(34, ∞]	(34, ∞]
(0,8]	(40, ∞]	(32, ∞]	(16, 20]	(13, ∞]	(30, 36]
(4,11]	(17, 25]	(33, ∞]	(18, 25]	(16, 24]	(18, 24]
(15, ∞]	(46, ∞]	(19, 26]	(17, 26]	(35, ∞]	(16,60]
(11, 15]	(11, 18]	(37, ∞]	(32, ∞]	(15, 22]	(35, 39]
(22, ∞]	(38, ∞]	(34, ∞]	(23, ∞]	(11, 17]	(21, ∞]
(46, ∞]	(5, 12]	(36, ∞]	(44, 48]	(22, 32]	(11, 20]
(46, ∞]		(14, 17]	(10, 35]	(48, ∞]	

was a difference in the distribution of T for women who received radiation therapy alone versus a combination of radiation and chemotherapy. The data are shown in Table 3.10 for the two groups.

Obtain and compare nonparametric estimates of the survivor function $S(t)$ for each group.

(Section 3.5.3)

3.14 The following data are survival times for 121 breast cancer patients treated over the period 1929–1938, quoted in Boag (1949). Times are in months, and asterisks denote censoring times.

0.3	0.3*	4.0*	5.0	5.6	6.2	6.3	6.6	6.8
7.4*	7.5	8.4	8.4	10.3	11.0	11.8	12.2	12.3
13.5	14.4	14.4	14.8	15.5*	15.7	16.2	16.3	16.5
16.8	17.2	17.3	17.5	17.9	19.8	20.4	20.9	21.0
21.0	21.1	23.0	23.4*	23.6	24.0	24.0	27.9	28.2
29.1	30	31	31	32	35	35	37*	37*
37*	38	38*	38*	39*	39*	40	40*	40*
41	41	41*	42	43*	43*	43*	44	45*
45*	46*	46*	47*	48	49*	51	51	51*
52	54	55*	56	57*	58*	59*	60	60*
60*	61*	62*	65*	65*	67*	67*	68*	69*
78	80	83*	88*	89	90	93*	96*	103*
105*	109*	109*	111*	115*	117*	125*	126	127*
129*	129*	139*	154*					

(a) Calculate the Kaplan–Meier estimate of the survivor function. Estimate 1- and 5-year survival probabilities and give a standard error for these estimates.

(b) Group the data into a life table with 1-year intervals. Compare the 1- and 5-year survival probability estimates with those obtained in part (a).

(c) In the data given by Boag the individuals with censored survival times are actually known to fall into one of three groups:

(i) Individuals free from signs or symptoms of breast cancer, but who died from some other cause.

(ii) Individuals free from signs or symptoms of breast cancer and still alive at the time the data were collected.

(iii) Individuals still alive at the time the data were collected, but who were suffering a persistence or recurrence of the cancer that was unlikely to yield to further treatment.

How might you take this information into account in analyzing the data?

(Section 3.2, 3.6)

3.15 Sometimes it is desired to estimate the hazard function from life table data. With the notation of Section 3.6.1, let $t_{mj} = (a_{j-1} + a_j)/2$ be the midpoint of the jth interval and $\Delta_j = a_j - a_{j-1}$ the width ($j = 1, \ldots, k$). Two estimates that have been suggested for $h(t_{mj})$ are

$$\hat{h}_1(t_{mj}) = \frac{2\hat{\ell}_j}{\Delta_j(1 + \hat{p}_j)} \qquad \text{(Kimball 1960; Gehan 1969)}$$

$$\hat{h}_2(t_{mj}) = \frac{-\log \hat{p}_j}{\Delta_j} \qquad \text{(Sacher 1956).}$$

(a) Motivate these choices of estimates. Compare the estimates by expanding them in powers of $\hat{q}_j$.

(b) Give variance estimates for the two estimates.

(c) Suggest estimates and associated variance estimates for the density function $f(t_{mj})$ at $t = t_{mj}$.

(Section 3.6)

3.16 Using (B2) and (B5) of Appendix B, derive expressions (3.6.13) for the asymptotic variance of the $\sqrt{n}(\hat{q}_j - q_j^*)$ and show that they are asymptotically uncorrelated.

(Section 3.6)

3.17 Examine the asymptotic bias in the standard life table estimate of q_j by considering the random-censoring model leading to (3.6.12) when the lifetime distribution is exponential and the censoring time distribution is (1) exponential and (2) uniform over $[a_{j-1}, a_j)$, respectively.

(Section 3.6)

3.18 Thompson (1977) suggested a pseudolikelihood function for the q_j and p_j for the case of life table data. With the notation of Section 3.6, this is

$$L(\mathbf{q}) = \prod_{j=1}^{k} q_j^{D_j}(1 - q_j)^{N_j - D_j - .5W_j}. \tag{3.7.3}$$

(a) Show that maximization of $L(\mathbf{q})$ gives the standard life table estimates (3.6.3).

(b) Show that $L(\mathbf{q})$ can be written in the form

$$\prod_{j=1}^{k+1}[S(a_{j-1}) - S(a_j)]^{D_j} S(a_j)^{.5W_j + .5W_{j+1}}, \tag{3.7.4}$$

where $W_{k+1} = W_{k+2} = 0$.

(c) Compare (3.7.4) and the likelihood function based on an independent random censorship model as in Section 3.6.2, where the survivor function for censoring times is linear over $(0, a_k)$, with the censoring survival probability $G(a_k)$ much larger than $S(a_k)$.

(Section 3.6)

CHAPTER 4

Inference Procedures for Parametric Models

Likelihood methods for lifetime data were introduced in Chapter 2, and some procedures are summarized in Appendix C. This chapter provides detailed illustrations of the methodology, while dealing with several important lifetime distributions and different types of data. In most cases exact distribution theory for testing and estimation is not available, and we resort to approximations, based mainly on maximum likelihood large-sample theory.

The exponential distribution occupies an important historical position in lifetime distribution work, and Section 4.1 is devoted to it. Exact distributional results can be obtained for certain tests and estimation procedures, and these are presented along with large-sample methods. Section 4.2 provides shorter treatments of the gamma, inverse Gaussian, and other models. Sections 4.3 to 4.5 consider more complex settings involving interval censoring, threshold parameters, and mixture models. In addition to giving procedures for specific models, we show how parametric likelihood methods can be applied generally.

The most widely used parametric lifetime distribution models are those of log-location-scale type (Section 1.3.6). Likelihood methods for log-location-scale models, including the Weibull, log-normal and log-logistic distributions, are discussed in Chapter 5.

4.1 INFERENCE PROCEDURES FOR EXPONENTIAL DISTRIBUTIONS

The exponential distribution was the first lifetime model for which statistical methods were extensively developed. The existence of exact tests and confidence intervals for certain types of life test experiments was a major factor in the popularity of the model. It is recognized, however, that the applicability of the exponential distribution is limited to settings where the hazard function is close to constant, and that procedures based on the exponential tend to be nonrobust. It is thus important that the adequacy of the model in any setting be checked. The following three subsections discuss estimation and hypothesis tests, and Section 4.1.4 considers the planning of studies. Goodness-of-fit tests are discussed in Chapter 10.

4.1.1 Methods Based on Large Sample Theory

We consider the exponential distribution in the form (1.3.3) with density function $f(t;\theta) = \theta^{-1}\exp(-t/\theta)$. Under right-censoring processes that satisfy the conditions of Section 2.2.2, the log-likelihood function obtained from the general likelihood (2.2.3) is

$$\ell(\theta) = -r\log\theta - \frac{1}{\theta}\sum_{i=1}^{n} t_i, \tag{4.1.1}$$

where $r = \sum \delta_i$ is the number of uncensored lifetimes and t_i is a lifetime or censoring time. The likelihood equation $d\ell/d\theta = 0$ gives

$$\hat{\theta} = \sum_{i=1}^{n} t_i / r, \tag{4.1.2}$$

assuming $r > 0$. If $r = 0$, the log-likelihood $\ell(\theta)$ is bounded but monotone increasing as $\theta \to \infty$, so does not yield a finite maximum likelihood estimate (m.l.e.). In general both $\sum t_i$ and r are random variables, and the exact distributions of $\hat{\theta}$ and other quantities considered in the following paragraphs are mathematically intractable.

Maximum likelihood methodology is easily applied. As discussed in Example 2.5.1, the observed information $-d^2\ell/d\theta^2$ is

$$I(t) = \frac{-r}{\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^{n} t_i \tag{4.1.3}$$

and $I(\hat{\theta}) = r/\hat{\theta}^2$. Several procedures can be used to make inferences about θ (see Appendix C). The most straightforward is based on the asymptotic normal approximation

$$Z = \frac{\hat{\theta} - \theta}{I(\hat{\theta})^{-1/2}} \sim N(0, 1), \tag{4.1.4}$$

taking Z as a pivotal quantity. For example, this gives $\hat{\theta} \pm 1.96 I(\hat{\theta})^{-1/2}$ as an approximate .95 confidence interval for θ. Asymptotic likelihood theory indicates that $I(\hat{\theta})$ in (4.1.4) can be replaced with $I(\theta)$, with the expected information $\mathcal{I}(\theta)$ given by (2.5.6), or with $\mathcal{I}(\hat{\theta})$. As discussed in Example 2.5.1, calculation of $\mathcal{I}(\theta)$ requires an explicit model for the censoring process, and can be complicated. The use of $I(\theta)$ or $\mathcal{I}(\theta)$ in (4.1.4) also makes the inversion of intervals such as $-1.96 \le Z \le 1.96$ to get confidence intervals for θ more complicated, so in practice (4.1.4) is typically used.

The approximation (4.1.4) is not very accurate in small samples. This is associated with the fact that $\ell(\theta)$ tends to be asymmetric and not closely approximable by a quadratic when the number of uncensored observations is small. Alternative procedures that are more accurate are available; we mention two approaches that can be recommended.

Method 2: Sprott (1973) and others have shown that if the parameterization $\phi = \theta^{-1/3}$ is used, the log-likelihood $\ell_1(\phi) = \ell(\phi^{-3})$ is typically close to quadratic, and the approximation

$$Z = \frac{\hat{\phi} - \phi}{I_1(\hat{\phi})^{-1/2}} \sim N(0, 1) \tag{4.1.5}$$

is quite accurate, even for small samples. It is easily found that $I_1(\hat{\phi}) = 9r/\hat{\phi}^2$. Confidence intervals or tests for ϕ can be based on (4.1.5), and converted to intervals or tests for θ as desired.

Method 3: The likelihood ratio statistic

$$\Lambda(\theta) = 2\ell(\hat{\theta}) - 2\ell(\theta) \tag{4.1.6}$$

is approximately $\chi^2_{(1)}$ in large-samples when θ is the true parameter value, and can be used as an approximate pivotal quantity for testing or estimation of θ. Approximate two-sided α confidence intervals are obtained as the set of θ values for which $\Lambda(\theta) \le \chi^2_{(1),\alpha}$. One-sided intervals or tests are usually obtained by treating

$$Z = [\text{sign}(\hat{\theta} - \theta)]\Lambda(\theta)^{1/2}$$

as $N(0, 1)$, as described in Appendix C. Refinements that improve the accuracy of these methods are available, but do not make much difference except in very small samples.

Tests or confidence intervals for monotone functions of θ such as $S(t;\theta) = \exp(-t/\theta)$ or $h(t;\theta) = \theta^{-1}$ are easily obtained. For example, if Data $= \{(t_i, \delta_i), i = 1, \ldots, n\}$ and

$$L(\text{Data}) \le \theta \le U(\text{Data})$$

is (approximately) an α confidence interval for θ, then

$$\exp[-t_0/L(\text{Data})] \le S(t;\theta) \le \exp[-t_0/U(\text{Data})]$$

is (approximately) an α confidence interval for $S(t;\theta)$.

Example 4.1.1. (Example 1.1.6 revisited). The data in Example 1.1.6 concerned the lifetimes of 10 pieces of equipment. The observation scheme gives the following values for t_i and δ_i $(i = 1, \ldots, 10)$:

t_i :	2	72	51	60	33	27	14	24	4	21
δ_i :	1	0	1	0	1	1	1	1	1	0

We find $r = 7$, $\hat{\theta} = 44.0$ (days), $I(\hat{\theta}) = 7/44.0^2$, and $I(\hat{\theta})^{-1/2} = 16.6$. Using (4.1.4), we obtain a two-sided .95 confidence interval for θ as $\hat{\theta} \pm 1.96 I(\hat{\theta})^{-1/2}$, or $11.5 \le \theta \le 76.5$.

The sample size is small and it is illustrative to compare this result with ones based on alternative methods 2 and 3, as follows.

2. We find $\hat{\phi} = \hat{\theta}^{-1/3} = 0.2833$, $I_1(\hat{\phi}) = 9r/\hat{\phi}^2 = 784.96$, and $I_1(\hat{\phi})^{-1/2} = 0.0357$. An approximate .95 confidence interval for ϕ is $\hat{\phi} \pm 1.96 I_1(\hat{\phi})^{-1/2}$, which gives $.2133 \le \phi \le .3533$. Converting this to an interval for $\theta = \phi^{-3}$, we get $22.7 \le \theta \le 103.0$.
3. The likelihood ratio statistic (4.1.6) reduces here to

$$\Lambda(\theta) = 2r\{(\hat{\theta}/\theta) - 1 - \log(\hat{\theta}/\theta)\}.$$

Since $Pr\{\chi^2_{(1)} \le 3.84\} = .95$, a two-sided approximate .95 confidence interval is found, as the set of θ values giving $\Lambda(\theta) \le 3.84$. This yields $22.8 \le \theta \le 102.5$.

Methods 2 and 3 agree well, but the interval based on (4.1.4) is rather different. Figure 4.1 shows confidence intervals, and the degree of agreement, for the three methods. Two-sided confidence intervals for each method are based on finding θ values that satisfy something of the form $W(\theta) \le \chi^2_{(1),\alpha}$. The $W(\theta)$'s are, respectively,

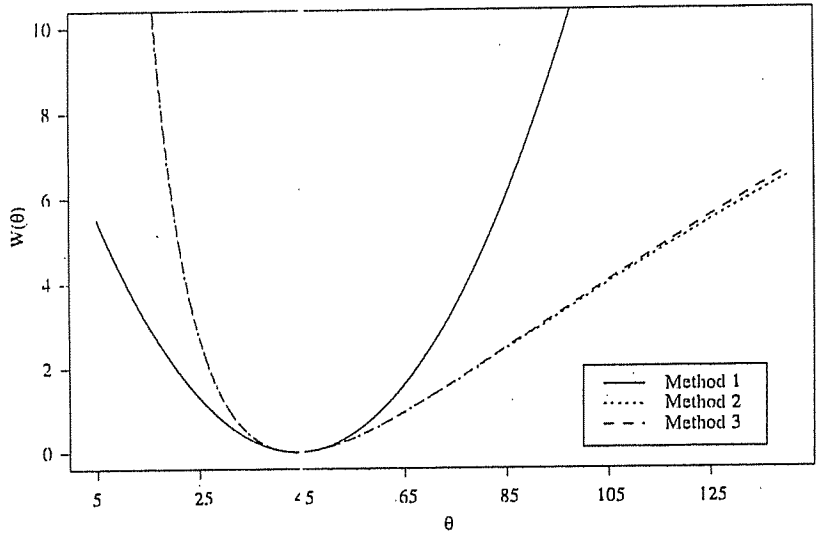

Figure 4.1. Likelihood ratio statistic $W_3(\theta)$ and approximations $W_1(\theta)$, $W_2(\theta)$.

$$W_1(\theta) = (\hat{\theta} - \theta)^2 I(\hat{\theta}) \quad \text{for (4.1.4)}$$
$$W_2(\theta) = (\hat{\theta}^{-1/3} - \theta^{-1/3})^2 I_1(\hat{\theta}^{-1/3}) \quad \text{for (4.1.5)}$$
$$W_3(\theta) = 2\ell(\hat{\theta}) - 2\ell(\theta) \quad \text{for (4.1.6).}$$

Figure 4.1 shows $W_j(\theta)$ for $j = 1, 2, 3$. The likelihood ratio statistic $W_3(\theta)$ is very asymmetric; the quadratic $W_1(\theta)$ does not approximate it well, but $W_2(\theta)$ does. As a result, confidence intervals based on methods 2 and 3 will agree closely, but differ from those based on (4.1.4).

Plots like Figure 4.1 are useful for comparing alternative large-sample methods. They also show confidence intervals with any given confidence coefficient; these are sometimes referred to as confidence distributions. For example, two-sided .90, .95, and .99 intervals consist of θ satisfying $W(\theta) \le 2.706$, $W(\theta) \le 3.841$, and $W(\theta) \le 6.635$, respectively.

Example 4.1.2. A Small Simulation Study. To indicate the extent to which the large-sample methods give appropriate coverage properties in small samples, some results of a small simulation study are presented in Table 4.1. Confidence intervals were obtained using (4.1.4), (4.1.5), and (4.1.6) in several censored data settings; results are shown for lower α confidence limits on θ. Two sample sizes $n = 10$ and 20, and three single Type 1 censoring patterns are represented. In each case all individuals had the same censoring time C, with C selected to give values $Q' = \exp(-C/\theta)$ of .10, .25, and .50. Thus, Q' is the effective censoring fraction, since each lifetime has a probability Q' of being censored.

There were 2000 samples generated for each $(Q'; n)$ combination. The simulations used $\theta = 1$, but the results shown are valid whatever the value of θ, since censoring times are chosen to be fixed multiples of θ. Table 4.1 shows the observed coverage proportions for confidence intervals with nominal coverage probabilities .90 and .95. The methods based on (4.1.5) and (4.1.6) give close to the nominal coverage, especially for $n = 20$, but (4.1.4) produces confidence limits that are too low, yielding coverage probabilities that are too high.

Table 4.1. Proportion of the Time (out of 2000 trials) That Approximate One-Sided .90, .95 Confidence Intervals Contained θ

		$Q' = .10$		$Q' = .25$		$Q' = .50$	
n	Method	.90	.95	.90	.95	.90	.95
10	(4.1.4)	.928	1.000	.996	1.000	1.000	1.000
	(4.1.5)	.913	.960	.908	.951	.893	.950
	(4.1.6)	.910	.960	.905	.944	.891	.945
20	(4.1.4)	.952	.993	.961	.995	.981	1.000
	(4.1.5)	.908	.951	.903	.951	.892	.949
	(4.1.6)	.907	.950	.901	.949	.893	.949

The approximate methods based on (4.1.5) and (4.1.6) are satisfactory for the situations represented in the table. Simulations for problems in which censoring times are unequal gave similarly good results. When the methods are used to obtain upper confidence limits, coverage probabilities are not quite as close to the nominal values as for the lower limits, but are still broadly acceptable. In many applications lower confidence limits are called for, so it is important that these turn out to be particularly accurate.

4.1.2 Exact Methods for Certain Life Tests

For certain special types of life tests it is possible to obtain exact confidence intervals and hypothesis tests; these are described briefly.

4.1.2.1 Type 2 Censored Test Plans

In a Type 2 censored plan (Section 2.2.1.3) the life test is terminated when the rth failure occurs, where r ($1 \le r \le n$) is prespecified. The data consist of the r smallest order statistics $t_{(1)} < \cdots < t_{(r)}$, and under an exponential model the joint distribution of $t_{(1)}, \ldots, t_{(r)}$ is, from (2.2.6),

$$\frac{n!}{(n-r)!}\left\{\prod_{i=1}^{r}\frac{1}{\theta}e^{-t_{(i)}/\theta}\right\}\left\{e^{-t_{(r)}/\theta}\right\}^{n-r}. \tag{4.1.7}$$

The log-likelihood function is of the same form as (4.1.1), with $\sum_{i=1}^{n} t_i$ replaced by

$$T = \sum_{i=1}^{r} t_{(i)} + (n-r)t_{(r)},$$

and $\hat{\theta} = T/r$. In life tests T is sometimes called the "total time on test" statistic, since it is the total of the observed lifetimes or censoring times across the n test units. As we now show, the distribution of T is easily found. Make the change of variables

$$\begin{aligned} W_1 &= nt_{(1)} \\ W_i &= (n-i+1)(t_{(i)} - t_{(i-1)}) \qquad i = 2, \ldots, r. \end{aligned} \tag{4.1.8}$$

Since

$$T = \sum_{i=1}^{r} t_{(i)} + (n-r)t_{(r)} = \sum_{i=1}^{r} W_i$$

and the Jacobian is

$$\frac{\partial(W_1, \ldots, W_r)}{\partial(t_{(1)}, \ldots, t_{(r)})} = \frac{n!}{(n-r)!},$$

the joint probability density function (p.d.f.) of $W_1, \ldots, W_r$ is found from (4.1.7) to be

$$\frac{1}{\theta^r}\exp\left(-\sum_{i=1}^{r}\frac{w_i}{\theta}\right) \qquad w_i > 0.$$

We have proved the following result:

THEOREM 4.1.1. Let $t_{(1)}, \ldots, t_{(r)}$ be the first r ordered observations of a random sample of size n from the exponential distribution (1.3.3). Then the quantities $W_1, \ldots, W_r$ given by (4.1.8) are independent and identically distributed, also with p.d.f. (1.3.3).

Since $T = \sum_{i=1}^{r} W_i$, we also immediately have from Section 1.3.5 the following result.

COROLLARY 4.1.1. Under the conditions of Theorem 4.1.1,

$$T = \sum_{i=1}^{r} t_{(i)} + (n-r)t_{(r)}$$

has a distribution given by $2T/\theta \sim \chi^2_{(2r)}$.

Tests and confidence intervals for θ are easily obtained using the pivotal quantity $2T/\theta$. For example, to obtain a two-sided $1-\alpha$ confidence interval for θ, we take

$$Pr\left(\chi^2_{(2r),\alpha/2} \le \frac{2T}{\theta} \le \chi^2_{(2r),1-\alpha/2}\right) = 1-\alpha,$$

where $\chi^2_{(2r),p}$ is the pth quantile of $\chi^2_{(2r)}$. Then

$$\frac{2T}{\chi^2_{(2r),1-\alpha/2}} \le \theta \le \frac{2T}{\chi^2_{(2r),\alpha/2}}$$

is the $1-\alpha$ confidence interval for θ.

Example 4.1.3. The first 8 observations in a random sample of 12 lifetimes from an assumed exponential distribution are, in hours

31, 58, 157, 185, 300, 470, 497, 673.

Hence $n = 12$, $r = 8$, and $T = 5063$. The m.l.e. for θ is $\hat{\theta} = 5063/8 = 632.9$ hours. To obtain, for example, a two-sided .95 confidence interval for θ, we find from the χ^2 distribution that $Pr(6.91 \le 2T/\theta \le 28.8) = Pr(6.91 \le \chi^2_{(16)} \le 28.8) = .95$, which gives $(2T/28.8, 2T/6.91)$ as a .95 confidence interval for θ. For

the sample observed, $T = 5063$, and the realized .95 confidence interval for θ is therefore (351.6, 1465.4).

4.1.2.2 *Testing with Replacement*

When life tests of equipment are conducted using a physical layout that can accommodate some maximum number of items, it may be feasible to replace failed items immediately with new ones. This is referred to as testing with replacement, and the test is typically terminated after a fixed length of time (Type 1) or a fixed number of failures (Type 2).

For either Type 1 or Type 2 testing with replacement, the censoring mechanism satisfies the conditions of Section 2.2.2, and the log-likelihood function under an exponential lifetime distribution for the units is given by (4.1.1). However, direct derivations of $L(\theta)$ indicate that exact inference procedures are available. The key point is that the observed failure times in the experiment are the times of occurrence of events in a homogeneous Poisson process with intensity n/θ (e.g., Cox and Lewis 1966, Ch. 2). With Type 1 testing the total time on test, $T = \sum t_i$, is fixed at the value nL_0, where L_0 is the length of the test, and the number of failures $r = \sum \delta_i$ has a Poisson distribution with mean equal to nL_0/θ. The Poisson probability function gives the likelihood

$$L(\theta) \propto e^{-nL_0/\theta} \frac{(nL_0/\theta)^r}{r!}$$
$$\propto \frac{1}{\theta^r} e^{-T/\theta},$$

which is of the general form arising from (2.2.3). Confidence intervals or tests for θ can in this case be obtained using standard methods for the Poisson distribution (e.g., Cox and Lewis 1966, Ch. 2). Since r, or equivalently, $\hat{\theta} = T/r$ is a minimal sufficient statistic for θ, we can formulate procedures in terms of $\hat{\theta}$ if we wish.

With Type 2 testing the number of failures, r, is fixed and the duration, L_0, of the test is a random variable. Because L_0 is the time to the rth event in a Poisson process with intensity n/θ, we have $nL_0/\theta \sim \text{Ga}(r)$ or, equivalently, $2nL_0/\theta = 2r\hat{\theta}/\theta \sim \chi^2_{(2r)}$ (Cox and Lewis 1966). The gamma p.d.f. (1.3.19) produces a log-likelihood of the form (4.1.1), L_0 (or $\hat{\theta}$) is a minimal sufficient statistic, and confidence intervals or tests can be based on the pivotal quantity $2r\hat{\theta}/\theta$. The procedures are formally the same as for the case of Type 2 testing without replacement (i.e., the case of ordinary Type 2 censoring), discussed earlier.

4.1.3 Comparison of Distributions

The comparison of two or more lifetime distributions is often an important goal. When the distributions are all exponential, this amounts to a comparison of their means. Tests and confidence intervals for comparison are considered in this section, based on independent samples from the distributions in question.

4.1.3.1 *Likelihood Ratio Tests*

The standard m-sample problem is to test equality of m distributions, which here is equivalent to testing the hypothesis

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_m.$$

Based on independent censored samples from the m distributions, the combined likelihood function is

$$L(\theta_1, \ldots, \theta_m) = \prod_{i=1}^{m} \frac{1}{\theta_i^{r_i}} e^{-T_i/\theta_i}, \tag{4.1.9}$$

where the data from distribution i ($i = 1, \ldots, m$) consist of $\{(t_{ij}, \delta_{ij}), j = 1, \ldots, n_i\}$, and

$$r_i = \sum_{j=1}^{n_i} \delta_{ij}, \qquad T_i = \sum_{j=1}^{n_i} t_{ij}$$

are the observed number of failures and total time on test (or at risk) for the ith distribution.

The likelihood ratio statistic for testing H_0 is

$$\Lambda = 2\ell(\hat{\theta}_1, \ldots, \hat{\theta}_m) - 2\ell(\tilde{\theta}_1, \ldots, \tilde{\theta}_m), \tag{4.1.10}$$

where $\ell(\theta_1, \ldots, \theta_m) = \log L(\theta_1, \ldots, \theta_m)$ and the $\hat{\theta}_i$ and $\tilde{\theta}_i$ are the unrestricted m.l.e.'s and the m.l.e.'s under the hypothesis H_0, respectively. The unrestricted m.l.e.'s are $\hat{\theta}_i = T_i/r_i$, from previous results. Under H_0 the θ_i are equal, and it is easily seen that (4.1.9) is maximized at $(\tilde{\theta}_1, \ldots, \tilde{\theta}_m) = (\tilde{\theta}, \ldots, \tilde{\theta})$, where

$$\tilde{\theta} = \sum_{i=1}^{m} T_i \Big/ \sum_{i=1}^{m} r_i. \tag{4.1.11}$$

These results give (4.1.10) as

$$\Lambda = \left(2\sum_{i=1}^{m} r_i\right) \log \tilde{\theta} - 2\sum_{i=1}^{m} r_i \log \hat{\theta}_i. \tag{4.1.12}$$

Asymptotically Λ has a $\chi^2_{(m-1)}$ distribution if H_0 is true, and the χ^2 approximation is suitable for computing significance levels if the r_i are not too small.

In the case where the censoring is of Type 2 for each of the m samples, a refinement to the χ^2 approximation improves its accuracy. The idea is to treat $\Lambda_1 = C\Lambda$ as $\chi^2_{(m-1)}$, where

$$C^{-1} = 1 + \frac{1}{6(m-1)} \left(\sum_{i=1}^{m} r_i^{-1} - r^{-1}\right), \tag{4.1.13}$$

and $r = \sum r_i$. This was first suggested by Bartlett (1937) in connection with tests for variance estimates. Chao and Glaser (1978) and Dyer and Keating (1980) provide additional results and references for the case of Type 2 censoring.

Example 4.1.4. As a numerical illustration, suppose that four independent samples of size 10 each had 7 failures, and gave m.l.e.'s under the exponential model as $\hat{\theta}_1 = 106$, $\hat{\theta}_2 = 80$, $\hat{\theta}_3 = 140$, $\hat{\theta}_4 = 158$. To test $H_0 : \theta_1 = \theta_2 = \theta_3 = \theta_4$ using (4.1.12), we find $\tilde{\theta} = \sum r_i \hat{\theta}_i / 28 = 121$, and an observed value of $\Lambda = 1.87$. The significance level (p-value) for the test is, using the χ^2 approximation,

$$Pr(\chi^2_{(3)} \geq 1.87) = .60,$$

thus providing no evidence against H_0. If the censoring in each sample were of Type 2, the refined approximation using (4.1.13) would give $C = 0.971$ and the p-value $Pr[\chi^2_{(3)} \geq (0.971)(1.87)] = .61$; there is no effect on the conclusions.

4.1.3.2 Confidence Intervals for θ_1/θ_2

Confidence intervals for θ_1/θ_2 provide a natural comparison of two exponential distributions and supplement a test of $\theta_1 = \theta_2$. A very simple procedure is to note from (4.1.4) that $\hat{\theta}_i$ $(i = 1, 2)$ can be treated as approximately $N(\theta_i, \theta_i^2 r_i^{-1})$ in large samples, so that $\log \hat{\theta}_i$ is approximately $N(\log \theta_i, r_i^{-1})$. Thus

$$Z = \frac{\log(\hat{\theta}_1/\hat{\theta}_2) - \log(\theta_1/\theta_2)}{(r_1^{-1} + r_2^{-1})^{1/2}} \tag{4.1.14}$$

is approximately $N(0, 1)$ and can be used as a pivotal quantity to get confidence intervals for $\log(\theta_1/\theta_2)$ and thus θ_1/θ_2.

Confidence intervals can also be found by inverting the likelihood ratio test for a hypothesis of the form $H_0 : \theta_1 = a\theta_2$, where $a > 0$ is a constant. The m.l.e.'s of θ_1 and θ_2 under H_0 are found by maximizing $L(a\theta_2, \theta_2)$ given by (4.1.9), and are

$$\tilde{\theta}_1 = a\tilde{\theta}_2 = \frac{T_1 + aT_2}{r_1 + r_2}.$$

The likelihood ratio statistic for testing H_0 is then

$$\begin{aligned} \Lambda &= 2\ell(\hat{\theta}_1, \hat{\theta}_2) - 2\ell(\tilde{\theta}_1, \tilde{\theta}_2) \\ &= 2r_1 \log(\tilde{\theta}_1/\hat{\theta}_1) + 2r_2 \log(\tilde{\theta}_2/\hat{\theta}_2). \end{aligned} \tag{4.1.15}$$

An α confidence interval for θ_1/θ_2 consists of the set of values a for which H_0 is not contradicted (or rejected) at the $1 - \alpha$ level of significance. If the approximation $\Lambda \sim \chi^2_{(1)}$ is used, this entails finding all values of a for which $\Lambda \leq \chi^2_{(1),\alpha}$.

If the samples from the two distributions are Type 2 censored, then an exact procedure is available, since $2r_i\hat{\theta}_i/\theta_i$ $(i = 1, 2)$ are independent $\chi^2_{(2r_i)}$ variables. Thus

$\hat{\theta}_1\theta_2/(\hat{\theta}_2\theta_1)$ has an $F_{(2r_1,2r_2)}$ distribution and can be used as a pivotal quantity for θ_1/θ_2.

The χ^2 approximation for (4.1.15) is slightly more accurate than the normal approximation for (4.1.14) with quite small samples, but (4.1.14) is a little easier to use. Unless the r_i are very small, the two methods tend to agree well.

Example 4.1.5. Suppose that in a small clinical trial to compare the duration of remission achieved by two drugs used in the treatment of leukemia, two groups of 20 patients produced $r_1 = 10$, $T_1 = 700$ ($\hat{\theta}_1 = 70$ weeks) and $r_2 = 10$, $T_2 = 540$ ($\hat{\theta}_2 = 54$ weeks) under a Type 1 censoring scheme and assumed exponential duration distributions. Let us obtain (approximate) .95 confidence intervals for θ_1/θ_2.

The two-sided .95 confidence interval for $\log(\theta_1/\theta_2)$ based on (4.1.14) is given by $\log(\hat{\theta}_1/\hat{\theta}_2) \pm 1.96(r_1^{-1} + r_2^{-1})^{1/2}$. This yields $-0.6170 \leq \log(\theta_1/\theta_2) \leq 1.1360$, which converts to the confidence interval $0.54 \leq \theta_1/\theta_2 \leq 3.11$. The approach based on the likelihood ratio statistic (4.1.15) requires that we find values of a $(= \theta_1/\theta_2)$ such that

$$\Lambda = 20\log(0.5 + 0.386a) + 20\log(0.5 + 0.648/a)$$

is less than $\chi^2_{(1),.95} = 3.841$. It is readily found that $0.51 \leq a \leq 3.15$; this confidence interval agrees closely with the previous one.

4.1.4 Planning Experiments or Life Tests

Section 2.5 discussed some general issues concerning the planning of studies on lifetime distributions, with emphasis on the estimation of specific parameters. Example 2.5.1 considered estimation of the mean θ of an exponential distribution, based on large-sample methods. Analogous results can be obtained for other procedures described in Sections 4.1.1–4.1.3; we provide two brief examples.

Example 4.1.6. In the case of a Type 2 censored life test (Section 4.1.2) the distribution of $2r\hat{\theta}/\theta$ is exactly $\chi^2_{(2r)}$, and gives exact confidence intervals for θ. The ratio of upper to lower confidence limits (UCL/LCL) for θ is a function of r, and r can be chosen to make the ratio acceptably small. For example, for a two-sided .90 confidence interval the ratio is

$$\frac{UCL}{LCL} = \frac{\chi^2_{(2r),.95}}{\chi^2_{(2r),.05}},$$

and $r = 10$ and $r = 20$ give ratios 2.89 and 2.10, respectively. Note that in choosing r, we leave open the choice of sample size n. This may be selected to control the duration of the life test; a larger value of n will lead to shorter durations. This point is discussed in subsection 4.1.4.1.

Example 4.1.7. Suppose that we wish to estimate the ratio θ_1/θ_2 of mean lifetimes for two exponential distributions, based on studies that involve Type 1 censoring. The approximate pivotal quantity (4.1.14) gives two-sided α confidence intervals for $\log(\theta_1/\theta_2)$ of the form

$$\log(\hat{\theta}_1/\hat{\theta}_2) \pm z_{1-\alpha/2}(r_1^{-1} + r_2^{-1})^{1/2}, \tag{4.1.16}$$

where z_p is the pth quantile for a $N(0, 1)$ variable. The precision can be adjusted by controlling $E(r_1^{-1})$ and $E(r_2^{-1})$ or, equivalently to the first-order level of approximation represented by (4.1.16), by controlling $E(r_1)$ and $E(r_2)$. For example, for a .95 confidence interval the width (UCL–LCL) of (4.1.16) is $3.92(r_1^{-1} + r_2^{-1})^{1/2}$, and experiments with $r_1 = r_2 = 10$ and $r_1 = r_2 = 20$, respectively, give UCL–LCL $= 1.753$ and 1.24. The ratio of upper to lower confidence limits for θ_1/θ_2 is therefore exp(UCL–LCL), or 5.77 and 3.45 for $r_1 = r_2 = 10$ and $r_1 = r_2 = 20$, respectively.

In some applications the objective is to carry out a formal test in which a specified null hypothesis H_0 is to be accepted or rejected. For example, in industrial or military applications tests are used to decide whether a batch of items is acceptable or not. In clinical trials or comparative life tests the objective is to make a decision concerning the distributions of two or more lifetime variables. The Neyman–Pearson theory of hypothesis testing provides a framework for decision making; the following two subsections outline the main ideas.

4.1.4.1 Tests for a Single Distribution

Any hypothesis concerning an exponential distribution can be expressed in terms of the mean θ. The most common problem involves testing a specific value θ_0 against values less than θ_0, that is,

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta < \theta_0, \tag{4.1.17}$$

where H_0 and H_1 are referred to as the null and alternative hypotheses. A formal hypothesis test is a decision rule for either accepting H_0 or rejecting it in favor of H_1, on the basis of observed data. The size of the test is

$$\alpha = Pr(\text{reject } H_0; \theta = \theta_0),$$

and the power function is defined by

$$P(\theta_1) = Pr(\text{reject } H_0; \theta = \theta_1).$$

Tests are generally designed so that the size (i.e., $P(\theta_0)$) and the power $P(\theta_1)$ at some value $\theta_1 < \theta_0$ are specified values.

Let us examine the construction of a test of (4.1.17) for the case of Type 2 censored data. It is plausible and easily seen from general results on formal testing (e.g., Cox and Hinkley 1974; Epstein and Sobel 1953) that for a given r and n the most powerful tests are to accept H_0 iff $\hat{\theta} > C$, where C is a specified value. Since

$2r\hat{\theta}/\theta \sim \chi^2_{(2r)}$ under Type 2 censoring (Section 4.1.2), the power function is then

$$\begin{aligned} P(\theta_1) &= Pr(\hat{\theta} \le C; \theta = \theta_1) \\ &= Pr\left(\chi^2_{(2r)} \le \frac{2rC}{\theta_1}\right). \end{aligned} \tag{4.1.18}$$

For any r we get a size α test by choosing $C = C_\alpha = \theta_0 \chi^2_{(2r),\alpha}/2r$. The power $P(\theta_1)$ of the test can be increased for $\theta_1 < \theta_0$ by increasing r: note that

$$\begin{aligned} P(\theta_1) &= Pr\left(\frac{2r\hat{\theta}_1}{\theta_1} \le \frac{2rC_\alpha}{\theta_1}\right) \\ &= Pr\left(\chi^2_{(2r)} \le \frac{2rC_\alpha}{\theta_1}\right), \end{aligned}$$

so to make $P(\theta_1) = 1 - \beta$, we need

$$\frac{2rC_\alpha}{\theta_1} = \chi^2_{(2r),1-\beta}$$

or

$$\frac{\chi^2_{(2r),\alpha}}{\chi^2_{(2r),1-\beta}} = \frac{\theta_1}{\theta_0}. \tag{4.1.19}$$

Hence, to make $P(\theta_1)$ equal to $1 - \beta$, we must choose r such that (4.1.19) is satisfied. There will not generally be an integral r value that exactly satisfies (4.1.19). However, it can be seen that for $\alpha < .5$ and $\beta < .5$ the quotient on the left-hand side of (4.1.19) is an increasing function of r and approaches one from below as $r \to \infty$. Since $\theta_1/\theta_0 < 1$, there is a smallest-value r_0 of r such that the left-hand side is $\ge \theta_1/\theta_0$, and then, for any $r \ge r_0$, $P(\theta_1) \ge 1 - \beta$. The choice $r = r_0$ therefore gives a test with the desired size and (approximately) the desired power at $\theta = \theta_1$. The larger $1 - \beta$ is, the larger r_0 will be. The entire power function for the test can be calculated from (4.1.18).

It will be observed that no particular value of n is indicated by the preceding arguments. Two tests with the same value of r but different values of n have identical power functions. However, although n does not enter into the power calculations, it is an important factor, since the larger n is, the less the time generally required to complete the test. One aspect of this is given in the following result.

LEMMA 4.1.1. Let $t_{(r)}$ be the rth smallest observation in a random sample of size n from the exponential distribution with mean θ. Then

$$E(t_{(r)}) = \theta \sum_{i=1}^{r} \frac{1}{n - i + 1}. \tag{4.1.20}$$

Proof. By Theorem 4.1.1, $W_1 = nt_{(1)}$ and $W_i = (n - i + 1)(t_{(i)} - t_{(i-1)})$, $i = 2, \ldots, r$, are independent random variables all having the same exponential distribution as the original observations. Thus $E(W_i) = \theta$, $i = 1, \ldots, r$. But $t_{(r)}$ can be written as

$$t_{(r)} = \frac{W_1}{n} + \frac{W_2}{n-1} + \cdots + \frac{W_r}{n-r+1},$$

and hence the stated result follows

Example 4.1.8. A particular electronic device has a lifetime distribution adequately modeled by an exponential distribution. In setting up a screening procedure for consignments of these devices, it is decided to institute a Type 2 censored life test plan, with $\theta_0 = 1000$ hours, $\theta_1 = 400$ hours, $\alpha = .05$, and $\beta = .10$. In other words, the test is to have only a 5% chance of rejecting a distribution with mean 1000 hours, but a 90% chance of rejecting one with mean 400 hours.

The smallest integer r such that the left side of (4.1.19) exceeds $\theta_1/\theta_0 = 0.4$ is $r = 11$, which gives $\chi^2_{(22),.05}/\chi^2_{(22),.90} = 12.338/30.813 = .4004$. Then $C_{.05} = 1000(12.338)/22 = 561$. The plan therefore stipulates that we use a Type 2 censored life test with $r = 11$ and reject H_0 if $\hat{\theta} \leq 561$.

To show the effect of n, the total number of items on test, we can use (4.1.20) to calculate expected durations of the test. One finds, for example, that for $n = 11, 13, 15$, and 20, $E(t_{(11)}) = 3.02\theta, 1.68\theta, 1.23\theta$, and $.77\theta$, respectively. A decision as to how large n should be can be based on considerations involving the costs of testing, the amount of time available for the test, and the possibility of departures from the exponential model.

Formal tests can be developed for other types of life test. It is readily shown (see Problem 4.4) that for tests with replacement of failed items as in Section 4.1.2, the results (4.1.18) and (4.1.19) still hold; the expected durations of the tests are, however, smaller than for tests without replacement. For Type 1 censored life tests, and others for which large-sample methods are used, approximations or simulation must be used to assess power and decide on test parameters. The following example illustrates how this can be done.

Example 4.1.9. Consider a test of (4.1.17) based on large-sample methods. The normal approximation (4.1.5), based on the parameter $\phi = \theta^{-1/3}$, is considerably more accurate in small samples than the analogous approximation (4.1.4) based on θ itself. We can replace $I_1(\phi)$ in (4.1.5) with $\mathcal{I}_1(\phi) = 9E(r)/\phi^2$ to the same level of approximation, so we consider the approximation

$$Z = 3[E(r)]^{1/2}\left(\frac{\hat{\phi}}{\phi} - 1\right) \sim N(0, 1). \tag{4.1.21}$$

Tests of (4.1.17) are of the form: reject H_0 iff $\hat{\theta} \leq C$, or equivalently, reject H_0 iff $\hat{\phi} \geq C^*$. By (4.1.21) we find that the power function in terms of ϕ is approximately

$$\begin{aligned} P(\phi_1) &= Pr\{\hat{\phi} \geq C^*; \phi = \phi_1\} \\ &= Pr\left\{Z \geq 3[E(r)]^{1/2}\left(\frac{C^*}{\phi_1} - 1\right)\right\}. \end{aligned} \tag{4.1.22}$$

For a test with size $P(\phi_0) = \alpha$ and power $P(\phi_1) = 1 - \beta$ at a specified value ϕ_1, we therefore require

$$3[E(r)]^{1/2}\left(\frac{C^*}{\phi_0} - 1\right) = N_{1-\alpha}$$

$$3[E(r)]^{1/2}\left(\frac{C^*}{\phi_1} - 1\right) = N_{\beta},$$

where N_p is the pth quantile for the distribution $N(0, 1)$. For given α, β, ϕ_0, and ϕ_1 we can choose C^* and $E(r)$ to satisfy (approximately) these equalities.

If the primary purpose of an experiment is to provide a decision in favor of H_0 or H_1, then sequential procedures can often be valuable. Discussion of sequential methods is beyond the scope of this book, but the basic idea is that the life test is monitored over time, so that the decision to accept or reject H_0 can be made as soon as there is sufficient evidence to reach such a decision.

Epstein and Sobel (1955) presented a test in which the decision made at time t essentially depends on the inequality

$$B < \left(\frac{\theta_0}{\theta_1}\right)^{r(t)} \exp\left[\left(\theta_1^{-1} - \theta_0^{-1}\right) T(t)\right] < A, \tag{4.1.23}$$

where $r(t)$ is the number of failures observed by time t and $T(t)$ is the total time on test up to time t, that is, the total lifetime lived by all items, failed and unfailed, up to time t. At time t experimentation continues as long as (4.1.23) is satisfied; on the other hand, if the function in the middle of (4.1.23) is $\leq B$, H_0 is rejected, and if it is $\geq A$, H_0 is accepted. A slight modification consists of truncating the tests to avoid very long test times. The constants B and A are selected to give the test size α and desired power $1 - \beta$ at $\theta = \theta_1$; it turns out that to a close approximation $A = (1 - \beta)/\alpha$ and $B = \beta/(1 - \alpha)$. Epstein and Sobel (1955) give approximate formulas for calculating the power function and other characteristics of this test when testing is with or without replacement.

The main advantage of a sequential plan is that the time needed to reach a decision about H_0 versus H_1 can be substantially reduced from that required by a similar nonsequential plan. If, however, one is not just interested in a decision rule, but also in estimation, sequential procedures create complications, though it is possible to obtain conservative confidence limits from them (e.g., Bryant and Schmee, 1979). Another qualification of the sequential tests is that their properties depend rather heavily on the exponentiality of the underlying lifetime distribution, and so the possibility of departures from the model needs to be considered.

References to related work are given in the Bibliographic Notes at chapter's end.

4.1.4.2 *Tests for Comparing Two Distributions*

Many clinical trials or life tests are designed to compare two lifetime distributions. In the very special case where the two distributions are exponential, the test is usually either

$$H_0 : \theta_1 = \theta_2 \quad \text{vs.} \quad H_1 : \theta_1 \neq \theta_2,$$

or a test with a one-sided alternative $H_1 : \theta_1 > \theta_2$ or $H_1 : \theta_1 < \theta_2$.

As in Example 4.1.9, the crucial factor in the power of a test of H_0 is the expected number of observed failures during the study. An expression analogous to (4.1.22) can be obtained, leading to a determination of approximate study requirements, as follows.

Suppose independent censored random samples (t_{ji}, δ_{ji}), $i = 1, \ldots, n_j$ $(j = 1, 2)$ are obtained in a study on lifetime distributions 1 and 2. Let $r_j = \sum \delta_{ji}$ denote the number of failures in the two samples $j = 1, 2$. From Section 4.1.1 and Example 2.5.1, the expected information about θ_j is $\mathcal{I}_j(\theta_j) = E(r_j)/\theta_j^2$ and this leads to the asymptotic approximation

$$\log \hat{\theta}_j \sim N(\log \theta_j, E(r_j)^{-1}), \qquad j = 1, 2. \tag{4.1.24}$$

Let $\delta = \log(\theta_1/\theta_2)$, so that a test of $H_0 : \theta_1 = \theta_2$ is equivalent to a test of $\delta = 0$. Consider the approximate pivotal quantity

$$Z = \frac{\log(\hat{\theta}_1/\hat{\theta}_2) - \delta}{[E(r_1)^{-1} + E(r_2)^{-1}]}, \tag{4.1.25}$$

noting that if δ is the true value and $E(r_1)$ and $E(r_2)$ are computed using δ, then Z is asymptotically $N(0, 1)$ as sample size increases. To test $H_0 : \delta = 0$ we use the statistic given by Z with $\delta = 0$.

Let $P(\delta_1)$ be the power function of a specific test of $H_0 : \delta = 0$,

$$P(\delta_1) = Pr(\text{reject } H_0; \delta = \delta_1).$$

For a two-sided test with specified size $P(0) = \alpha$, the approximate normal rejection region is given by $|Z| > -N_{\alpha/2}$, based on Z in (4.1.25) with $\delta = 0$. Suppose now that the power $P(\delta_1)$ for some specified β_1 is to be equal to $1 - \beta$. Now

$$\begin{aligned} P(\delta_1) &= Pr\{|Z| > -N_{\alpha/2}; \delta = \delta_1\} \\ &= Pr\{|\log(\hat{\theta}_1/\hat{\theta}_2)| > -V^{1/2}N_{\alpha/2}; \delta = \delta_1\}, \end{aligned}$$

where $V = E(r_1)^{-1} + E(r_2)^{-1}$. Assume that V under $\delta = 0$ and $\delta = \delta_1$ are essentially the same, so that Z is approximately $N(\delta_1, V)$ when $\delta = \delta_1$. This yields the approximation

$$P(\delta_1) = Pr(Z_1 < N_{\alpha/2} - \delta_1 V^{-1/2}) + Pr(Z_1 > -N_{\alpha/2} - \delta_1 V^{-1/2}),$$

where $Z_1 \sim N(0, 1)$. Suppose without loss of generality that $\delta_1 > 0$; then $Pr(Z_1 < N_{\alpha/2} - \delta_1 V^{-1/2}) \doteq 0$ and to make $P(\delta_1) = 1 - \beta$, we need approximately $-N_{\alpha/2} - \delta_1 V^{-1/2} = N_\beta$, or

$$V = \delta_1^2/(N_{\alpha/2} + N_\beta)^2. \tag{4.1.26}$$

If the study is to be designed so that $E(r_1) = E(r_2) = E(r/2)$, then $V = 4/r$ and (4.1.26) gives

$$E(r) = 4(N_{\alpha/2} + N_\beta)^2/\delta_1^2 \tag{4.1.27}$$

as the required expected total number of failures. More generally, if $E(r_1) = \pi_1 E(r)$ and $E(r_2) = \pi_2 E(r)$ with $\pi_1 + \pi_2 = 1$, then (4.1.26) gives

$$E(r) = (N_{\alpha/2} + N_\beta)^2/\pi_1\pi_2\delta_1^2. \tag{4.1.28}$$

The expected number of failures depends on the study design which, as discussed previously, involves the choices of sample size and duration of follow-up. In addition, it depends upon the unknown parameters θ_1 and θ_2, and so in order to select a design to meet stated power objectives it is necessary to use provisional values of θ_1 and θ_2. If individuals have fixed censoring times C_{ji} $(j = 1, 2)$ for individuals $i = 1, \ldots, n_j$ from distributions 1 and 2, then

$$E(r) = \sum_{j=1}^{2} \sum_{i=1}^{n_j} (1 - e^{-C_{ji}/\theta_j}).$$

In special settings such as clinical trials, guidelines for study design have been given, allowing for factors such as staggered entry of individuals to the study, losses to follow-up during the study, and the lengths of time available for the accrual of subjects and their follow-up (e.g., Rubinstein et al. 1981; Lachin and Foulkes 1986).

Example 4.1.10. Suppose that a two-sided test of $H : \theta_1 = \theta_2$ is wanted with size $\alpha = .05$ and power $1 - \beta = .90$ when $\theta_1/\theta_2 = 2$, that is, when $\delta_1 = \log 2 = .693$. By (4.1.28) we then require

$$E(r) = E(r_1) + E(r_2) = 27.06/\pi_1\pi_2, \tag{4.1.29}$$

where $\pi_j = E(r_j)/E(r)$. Suppose further that in the study $n/2$ individuals from each of distributions 1 and 2 are to be followed for the same length of time C; then

$$E(r_j) = \frac{n}{2}(1 - e^{-C/\theta_j}), \qquad j = 1, 2. \tag{4.1.30}$$

If assumed values for θ_1 and $\theta_2 = .5\theta_1$ are considered, then values for n and C, which satisfy (4.1.30), can be found. For example, suppose that $\theta_1 = 1$ and that it is

possible to run the study with a maximum follow-up time of $C = 1$. Then (4.1.29) and (4.1.30) imply that $\pi_1 = .422 = 1 - \pi_2$, and

$$\frac{n}{2}(1 - e^{-1} - e^{-2}) = \frac{27.06}{\pi_1 \pi_2},$$

giving $n = 148.2$. Thus an estimated minimum of 149 individuals is required to achieve the desired power of .90 at $\delta_1 = \log 2$. It would be sensible in practice to assume a conservatively high value for θ_1, because if θ_1 is larger than assumed then the values of $E(r_1)$ and $E(r_2)$ will be smaller, giving less power than desired.

4.2 INFERENCE PROCEDURES FOR SOME OTHER MODELS

In this section we provide brief discussions of two other distributions that are sometimes used as lifetime models: the gamma and inverse Gaussian distributions. In addition we illustrate how maximum likelihood inference and model checking can be implemented for general univariate models, given right-censored data.

4.2.1 The Gamma Distribution

The two-parameter gamma distribution, discussed in Section 1.3.5, has p.d.f. of the form

$$f(t; \alpha, k) = \frac{1}{\alpha \Gamma(k)} \left(\frac{t}{\alpha}\right)^{k-1} \exp(-t/\alpha) \qquad t > 0$$

where $\alpha > 0$ and $k > 0$ are scale and shape parameters, respectively. The survivor function is

$$S(t; \alpha, k) = 1 - I(k, t/\alpha)$$

where $I(k, x)$ is the scaled incomplete gamma integral (1.3.16).

With uncensored data, some inference procedures have fairly simple exact forms, as discussed in a number of books on mathematical statistics (e.g., Cox and Hinkley 1974). We outline a few results, then consider censored data.

4.2.1.1 Uncensored Data

The log-likelihood function for k and α from a complete random sample $t_1, \ldots, t_n$ is

$$\begin{aligned} \ell(k, \alpha) &= \sum_{i=1}^{n} \log f(t_i; \alpha, k) \\ &= -nk \log \alpha - n \log \Gamma(k) + n(k-1) \log \bar{t} - n\bar{t}/\alpha, \end{aligned} \qquad (4.2.1)$$

where

$$\bar{t} = \sum_{i=1}^{n} t_i/n \qquad \text{and} \qquad \tilde{t} = \left(\prod_{i=1}^{n} t_i\right)^{1/n}$$

are the arithmetic and geometric means. Setting $\partial \ell/\partial \alpha$ and $\partial \ell/\partial k$ equal to 0 and rearranging slightly, we get the likelihood equations

$$k\alpha = \bar{t}, \qquad \log k - \psi(k) = \log(\bar{t}/\tilde{t}), \qquad (4.2.2)$$

where $\psi(k) = \Gamma'(k)/\Gamma(k)$ is the digamma function (see Appendix B). The m.l.e.'s $\hat{\alpha}$ and $\hat{k}$ are easily found by solving (4.2.2); note that the second equation can be solved to obtain $\hat{k}$, and then $\hat{\alpha} = \bar{t}/\hat{k}$. Alternatively, we can maximize (4.2.1) using some other approach, as discussed in Section 4.2.1.2 and Appendix D.

The statistics $\bar{t}$ and $\tilde{t}$ are jointly sufficient for α and k, and provide exact tests and confidence intervals that have certain optimality properties. In particular, scale-invariant tests of

$$H_0 : k = k_0 \quad \text{vs.} \quad H_1 : k > k_0$$

can be based on $W = \tilde{t}/\bar{t}$, whose distribution does not depend on α (e.g., Cox and Hinkley 1974, Sec. 5.3). It can be shown that large values of W provide evidence against H_0, so that the p-value (significance level) associated with an observed value w_{obs} of W is $Pr(W \geq w_{obs}; k = k_0)$. A $1 - p$ UCL for k is correspondingly the largest value k_0 that gives a significance level of p or greater. Engelhardt and Bain (1978a) discuss approximations to the distribution of W that are helpful. A simple alternative for computing significance levels is to use simulation to estimate $Pr(W \geq w_{obs}; k = k_0)$. To do this we merely need to generate samples from the gamma distribution with $\alpha = 1$ and $k = k_0$, and compute $w = \tilde{t}/\bar{t}$. By repeating this sufficiently many times we can estimate the probability that $W \geq w_{obs}$ as precisely as desired.

Uniformly most powerful unbiased tests for α can also be obtained; they are based on the conditional distribution of W given $\tilde{t}$. Engelhardt and Bain (1977, 1978a) discuss approximations for obtaining p-values or confidence intervals for α.

The parameters k and α are usually of less direct interest than distribution quantiles or survival probabilities. An alternative for making inferences about k, α or other characteristics of the gamma distribution is to use maximum likelihood methods. Although these involve large-sample approximations, they are easily implemented, are adaptable to arbitrary functions of k and α, and can deal with censored data. Moreover, their accuracy can be improved, if necessary in small samples, through parametric bootstrap simulations or second-order corrections (see Appendix C). We now consider likelihood methods for either censored or uncensored data. Example 4.2.1 compares likelihood inferences and the exact procedures of this section for some uncensored data.

4.2.1.2 *Likelihood Methods for Censored or Uncensored Data*

The likelihood function for a possibly censored random sample (t_i, δ_i), $i = 1, \ldots, n$, as described in Section 2.2, is given by (2.2.3) and the expressions for the gamma p.d.f. and survivor function (s.f.) as

$$L(k, \alpha) = \prod_{i=1}^{n} \left[\frac{1}{\alpha \Gamma(k)} \left(\frac{t_i}{\alpha} \right)^{k-1} e^{-t_i/\alpha} \right]^{\delta_i} [1 - I(k, t_i/\alpha)]^{1-\delta_i},$$

with $I(k, \alpha)$ given by (1.3.16). The corresponding log-likelihood function is

$$\ell(k, \alpha) = -rk \log \alpha - r \log \Gamma(k) + (k-1) \sum_{i=1}^{n} \delta_i \log t_i - \sum_{i=1}^{n} \delta_i t_i/\alpha$$

$$+ \sum_{i=1}^{n} (1 - \delta_i) \log[1 - I(k, t_i/\alpha)], \tag{4.2.3}$$

where $r = \sum \delta_i$ is the number of uncensored lifetimes.

Standard large-sample procedures are based on the asymptotic normality of the m.l.e.'s $(\hat{k}, \hat{\alpha})$. Asymptotic covariance matrices for $(\hat{k}, \hat{\alpha})$ obtained by inverting observed or expected information matrices involve second derivatives of incomplete gamma integrals (1.3.16) and are somewhat complicated. Since most of the common software packages do not handle the gamma distribution, the simplest computational approach is to maximize (4.2.3) using optimization software that does not require expressions for derivatives (see Appendix D), and gives an estimate of the asymptotic covariance matrix at $(\hat{k}, \hat{\alpha})$ obtained by numerical differentiation. An alternative approach for tests or confidence intervals about parameters, which is especially preferable in small samples, is to use likelihood ratio procedures, described in Appendix C. We will describe these methods in some detail.

A contour plot of the joint relative log-likelihood function $r(k, \alpha) = \ell(k, \alpha) - \ell(\hat{k}, \hat{\alpha})$ provides an informative picture of the information about k, α or functions of them. In addition, the extent to which contours are approximately ellipsoidal (quadratic) indicates whether confidence intervals based on large-sample normal approximations for $(\hat{k}, \hat{\alpha})$ will be accurate and in agreement with results based on likelihood ratio procedures. Instead of $r(k, \alpha)$, we can choose to plot the likelihood ratio statistic $-2r(k, \alpha)$, that is,

$$\Lambda(k, \alpha) = 2\ell(\hat{k}, \hat{\alpha}) - 2\ell(k, \alpha). \tag{4.2.4}$$

Approximate joint confidence regions for (k, α) with confidence coefficient p are given as the set of points (k, α) satisfying $\Lambda(k, \alpha) \leq \chi^2_{(2),p}$.

Inferences concerning k or α are obtained from their maximized or profile log-likelihood functions, or equivalent likelihood ratio statistics. For example, to test the hypothesis $H_0 : k = k_0$, we can use the likelihood ratio statistic

$$\Lambda_1(k_0) = 2\ell(\hat{k}, \hat{\alpha}) - 2\ell(k_0, \tilde{\alpha}(k_0)), \tag{4.2.5}$$

where $\tilde{\alpha}(k_0)$ is the m.l.e. for α when $k = k_0$, obtained by maximizing $\ell(k_0, \alpha)$ with respect to α. In large samples the distribution of $\Lambda_1(k_0)$ is approximately $\chi^2_{(1)}$ when $k = k_0$, and it appears this approximation is reasonably accurate even for small-sample sizes. An approximate two-sided p confidence interval for k is obtained as the set of values k_0 satisfying $\Lambda_1(k_0) \leq \chi^2_{(1),p}$.

Tests and confidence intervals for α are obtained in a similar way by using

$$\Lambda_2(\alpha_0) = 2\ell(\hat{k}, \hat{\alpha}) - 2\ell(\tilde{k}(\alpha_0), \alpha_0),$$

where $\tilde{k}(\alpha_0)$ maximizes $\ell(k, \alpha_0)$.

Getting tests or confidence intervals for quantiles or the gamma distribution's survivor function is a little more complicated, since the survivor function has no simple closed form. Suppose, for example, that a confidence interval for $S(t_0)$ is wanted, for a specified time t_0. Since $S(t_0) = 1 - I(k, t_0/\alpha)$, we consider hypotheses of the form

$$H_0 : I(k, t_0/\alpha) = 1 - s_0. \tag{4.2.6}$$

If $\tilde{k}$ and $\tilde{\alpha}$ are the m.l.e.'s of k and α subject to the constraint (4.2.6), then under H_0 the likelihood ratio statistic

$$\Lambda(s_0) = 2\ell(\hat{k}, \hat{\alpha}) - 2\ell(\tilde{k}, \tilde{\alpha}) \tag{4.2.7}$$

is approximately $\chi^2_{(1)}$. Large values of Λ provide evidence against H_0, and an approximate p confidence interval for $S(t_0)$ consists of the set of values s_0 satisfying $\Lambda(s_0) \leq \chi^2_{(1),p}$.

Tests and confidence intervals for quantiles can be obtained in a similar way. The γth quantile t_γ satisfies $I(k, t_\gamma/\alpha) = \gamma$, so for a specified γ we consider the hypotheses

$$H_0 : I(k, t_0/\alpha) = \gamma,$$

which are exactly the same form as (4.2.6). Tests of H_0 are therefore carried out as for (4.2.6). However, to obtain a p confidence interval for t_γ for a specified γ, we fix the value $s_0 = 1 - \gamma$ in (4.2.6) and find the set of values t_0 such that $\Lambda(s_0)$ in (4.2.7) is $\leq \chi^2_{(1),p}$.

To implement the likelihood ratio method for quantiles or survival probabilities we must maximize $\ell(k, \alpha)$ subject to the constraint (4.2.6), for specified values t_0 and s_0. This is easily done as follows:

1. Define $M(k) = \ell(k, \alpha(k))$, where $\alpha(k)$ is defined implicitly by (4.2.6) for given k. Note that to find $\alpha(k)$ we can merely solve the equation in (4.2.6) for $t_0^* = t_0/\alpha$, and then $\alpha = t_0/t_0^*$. The value t_0^* satisfies $I(k, t_0^*) = 1 - s_0$ and is simply the $1 - s_0$ quantile $Q(1 - s_0; k)$ for the one-parameter gamma distribution $Ga(k)$ of (1.3.17); standard software gives these values.
2. Use an optimization procedure for $M(k)$ that does not require analytical derivatives (see Appendix D).

Example 4.2.1. The data that follow are survival times in weeks for 20 male rats that were exposed to a high level of radiation. The data are due to Furth et al. (1959) and have been discussed by Engelhardt and Bain (1977) and others. The times are

152, 152, 115, 109, 137, 88, 94, 77, 160, 165,
125, 40, 128, 123, 136, 101, 62, 153, 83, 69.

The arithmetic and geometric means of the 20 lifetimes are $\bar{t} = 113.45$ and $\tilde{t} = 107.07$. The m.l.e.'s of k and α are easily found from (4.2.2) or by direct maximization of (4.2.1) to be $\hat{k} = 8.80$, $\hat{\alpha} = 12.89$. The asymptotic covariance matrix for $(\hat{k}, \hat{\alpha})$ obtained by inverting the observed information matrix gives the standard errors (estimated standard deviations) $se(\hat{k}) = 2.73$, $se(\hat{\alpha}) = 4.12$ and estimated asymptotic correlation $\widehat{corr}(\hat{k}, \hat{\alpha}) = -.97$. We note that $\hat{k}$ and $\hat{\alpha}$ are highly correlated; recall from (4.2.2) that $\hat{k}\hat{\alpha} = \bar{t}$ in the case of uncensored data.

Approximate .95 confidence intervals for k and α, obtained as $\hat{k} \pm 1.96 se(\hat{k})$ and $\hat{\alpha} \pm 1.96 se(\hat{\alpha})$, are $3.45 \le k \le 14.15$ and $4.82 \le \alpha \le 20.96$. By way of comparison, the .95 confidence interval for k obtained by the invariant procedure based on $W = \bar{t}/\tilde{t}$ (Engelhardt and Bain 1978a) is $4.03 \le k \le 14.40$, and the .95 interval for α based on the uniformly most powerful unbiased test (Engelhardt and Bain 1977) is $6.5 \le \alpha \le 23.7$. The likelihood ratio confidence intervals are $4.5 \le k \le 15.3$ and $7.3 \le \alpha \le 26.0$. The agreement between the three methods is reasonably good, with the exact confidence limits lying between those for the Wald and likelihood ratio methods.

Let us now consider a confidence interval for the median lifetime $t_{.50}$, which would generally be of more interest than intervals for α or k. Confidence intervals for the mean $\mu = k\alpha$ are also of interest, and easily obtained. By (4.2.6), $t_{.5}$ satisfies $I(k, t_{.5}/\alpha) = .50$; that is, $t_{.5} = \alpha Q(.5, k)$, where $Q(p, k)$ is the pth quantile for the one-parameter gamma distribution (1.3.17). The m.l.e. for $t_{.5}$ is easily obtained as $\hat{t}_{.5} = \hat{\alpha} Q(.5, \hat{k}) = 109.2$, but computation of a standard error is difficult. Therefore we use the likelihood ratio procedure described just before this example in order to get a confidence interval. This involves the calculation of likelihood ratio statistic values $\Lambda(t_{.5})$ for specified values of $t_{.5}$; note that $\Lambda(t_{.5}) = 2\ell(\hat{k}, \hat{\alpha}) - 2\ell(\tilde{k}, \tilde{\alpha})$, where $\tilde{k}, \tilde{\alpha}$ maximize $\ell(k, \alpha)$ subject to the restriction $\alpha Q(.5, k) = t_{.5}$. The approximate .95 confidence interval for $t_{.5}$ consists of values for which $\Lambda(t_{.5}) \le 3.84$, and gives $92.9 \le t_{.5} \le 127.2$.

To illustrate the censored data case, we suppose that the data had been censored at $t = 150$ weeks; this would result in five censored survival times, all equal to 150. In this case $\hat{k} = 5.79$, $\hat{\alpha} = 21.3$, $se(\hat{k}) = 2.12$, $se(\hat{\alpha}) = 8.54$, and Wald approximate .95 confidence intervals for k and α are $4.63 \le k \le 9.95$ and $4.56 \le \alpha \le 38.0$. The likelihood ratio .95 confidence intervals are $2.6 \le k \le 11.1$ and $10.7 \le \alpha \le 52.6$, there being a fairly big discrepancy between the two confidence intervals for α. The reason for this is the markedly nonquadratic shape of the likelihood ratio statistic $\Lambda_2(\alpha)$, which is shown in Figure 4.2. The agreement between the Wald and likelihood ratio intervals is improved if we use the parameterization $\psi = \log \alpha$. In this

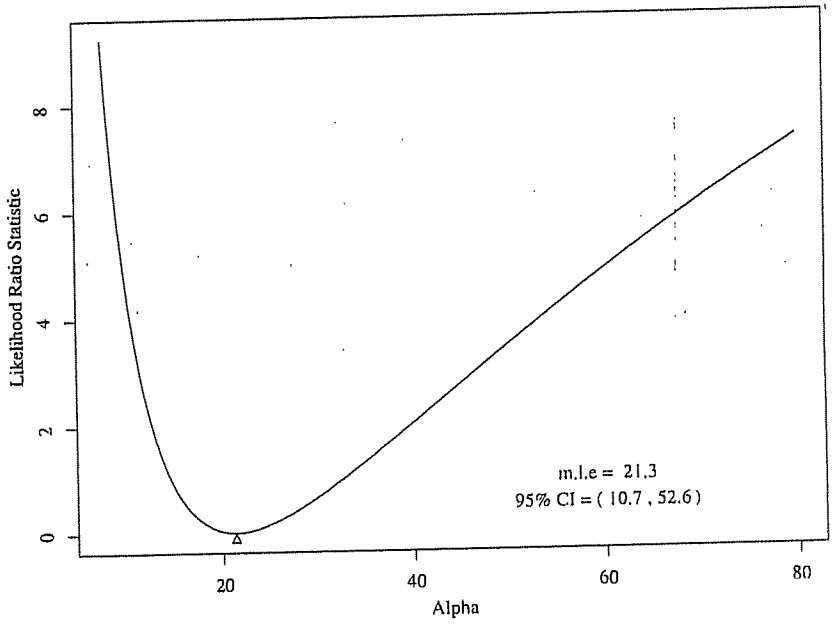

Figure 4.2. Likelihood ratio statistic for gamma scale parameter α.

case we have $\hat{\psi} = \log \hat{\alpha} = 3.059$, $se(\hat{\psi}) = se(\hat{\alpha})/\hat{\alpha} = .401$, and an approximate .95 confidence interval $2.273 \le \psi \le 3.845$ from $\hat{\psi} \pm 1.96 se(\hat{\psi})$. This transforms to $9.71 \le \alpha \le 46.8$, which is fairly close to the likelihood ratio interval. A check of the likelihood ratio statistic or the profile likelihood for ψ shows it to be approximately quadratic.

In order to get a confidence interval for $t_{.5}$ with censored data, it is simplest to use the likelihood ratio statistic $\Lambda(t_{.5})$. Figure 4.3 shows a plot of the statistic; the .95 confidence interval, consisting of values for which $\Lambda(t_{.5}) \le 3.84$, is $95.1 \le t_{.5} \le 144.6$. The m.l.e. is $\hat{t}_{.5} = 116.3$ weeks.

Graphical model checks for the gamma distribution can be based on probability or quantile plots, as described in Section 3.3.1. For the former, for example, we plot the values (3.3.1) against $S(t_j; \hat{k}, \hat{\alpha}) = 1 - I(t_j/\hat{\alpha}, \hat{k})$, where the t_j are the observed failure times. For the censored data case here, this is the same as a plot of the points $((j - 0.5)/20, I(t_j^*/\hat{\alpha}, \hat{k}))$ for $j = 1, \ldots, 15$, where $t_1^* < \cdots < t_{15}^*$ are the ordered, uncensored survival times. This plot is roughly linear, and provides no indication that the gamma model is unsatisfactory.

A second approach is simply to plot the estimated survivor function $S(t; \hat{k}, \hat{\alpha})$, and the Kaplan–Meier estimate $\hat{S}(t)$ on the same graph; equivalently, we could plot the corresponding distribution function estimates. The fit of the model is readily apparent, and we have the advantage of an untransformed plot of survival probabilities. Figure 4.4 in the next section presents such a plot for an inverse Gaussian model.

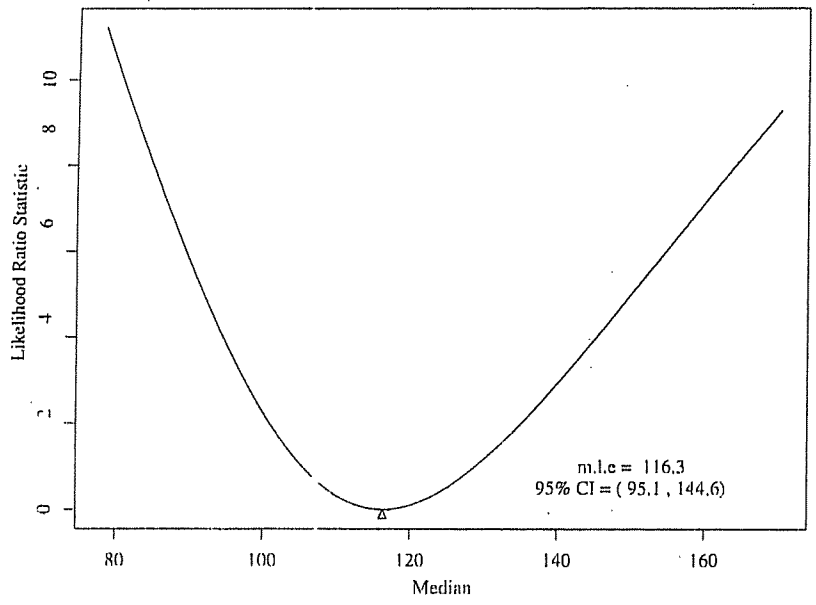

Figure 4.3. Likelihood ratio statistic for the gamma median $t_{.5}$.

4.2.2 The Inverse Gaussian Distribution

The inverse Gaussian distribution $IG(\mu, \lambda)$ discussed in Section 1.3.7 has p.d.f. given by (1.3.23) and distribution function by (1.3.24). For the case of an uncensored random sample $t_1, \ldots, t_n$, the m.l.e.'s of μ and λ are easily seen to be

$$\hat{\mu} = \bar{t}, \qquad \hat{\lambda} = n \Big/ \sum_{i=1}^{n} (t_i^{-1} - \bar{t}^{-1}).$$

In addition, it can be shown that $n\lambda/\hat{\lambda} \sim \chi^2_{(n-1)}$, that $\bar{t} \sim IG(\mu, n\lambda)$, and that $\bar{t}$ and $\hat{\lambda}$ are independent. Confidence limits or tests for λ are consequently easily obtained, and uniformly most powerful unbiased tests and associated confidence limits for μ are also available. Chhikara and Folks (1977, 1989), Jorgensen (1981), and Johnson et al. (1994, Ch. 15) discuss these and other results for inverse Gaussian models.

We consider here the general case involving possibly censored data (t_i, δ_i), $i = 1, \ldots, n$. The log-likelihood function for μ and λ is

$$\ell(\mu, \lambda) = \sum_{i=1}^{n} \delta_i \log f(t_i; \mu, \lambda) + (1 - \delta_i) \log[1 - F(t_i; \mu, \lambda)], \qquad (4.2.8)$$

where $f(t; \mu, \lambda)$ and $F(t; \mu, \lambda)$ are given by (1.3.23) and (1.3.24), respectively. Derivatives of $\ell(\mu, k)$ are messy but straightforward to evaluate, and the log-likelihood is readily maximized both by optimization software that requires expressions for first and second derivatives and by software that does not. Confidence intervals or tests for μ or λ are easy to obtain via likelihood ratio procedures or the asymptotic normality of $(\hat{\mu}, \hat{\lambda})$. Confidence intervals or tests for quantiles or survival probabilities can be obtained by likelihood ratio methods with a little effort, given the form (1.3.24) for the cumulative distribution function (c.d.f.). It is slightly simpler to use $(\hat{\mu}, \hat{\lambda})$ and the estimated asymptotic covariance matrix $I(\hat{\mu}, \hat{\lambda})^{-1}$, evaluating the latter either from algebraic expressions or by using optimization software that evaluates the second derivative matrix at $(\hat{\mu}, \hat{\lambda})$ by numerical methods.

Example 4.2.2. Whitmore (1983) considered data on the times to failure of 20 aluminum reduction cells. Failure times, in units of 1000 days, are given below with asterisks denoting a censored observation:

.468, .725, .838, .853, .965, 1.139, 1.142, 1.304, 1.317, 1.427,
1.554, 1.658, 1.764, 1.776, 1.990, 2.010, 2.224, 2.279*, 2.244*, 2.286*.

General optimization software readily finds m.l.e.'s for μ and λ from (4.2.8) as $\hat{\mu} = 1.61$, $\hat{\lambda} = 5.96$, with standard errors $se(\hat{\mu}) = 0.20$, $se(\hat{\lambda}) = 2.06$, and estimated

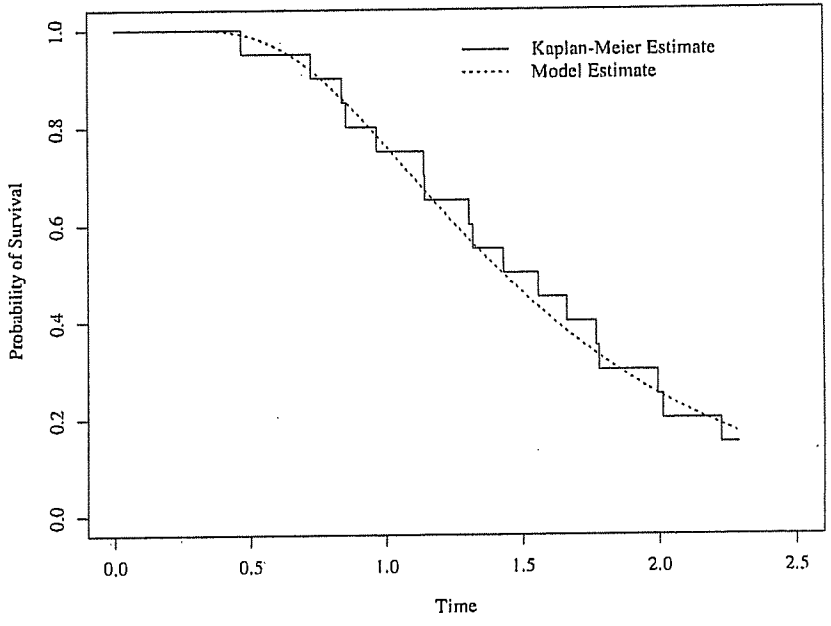

Figure 4.4. Inverse Gaussian and Kaplan–Meier estimates of $S(t)$ for aluminum reduction cells.

Table 4.2. Approximate 90% Confidence Intervals for $F(.5)$, $F(1.0)$, and $F(1.5)$

Method	$F(.5)$	$F(1.0)$	$F(1.5)$
Wald (1)	(0, .038)	(.11, .37)	(.40, .69)
Wald (2)	(.002, .082)	(.14, .39)	(.40, .68)
Bootstrap (3)	(0, .053)	(.09, .36)	(.38, .69)
Bootstrap (4)	(.001, .073)	(.11, .39)	(.40, .70)

asymptotic correlation −0.15. To assess the fit of the model we plot the estimated inverse Gaussian survivor function $S(t; \hat{\mu}, \hat{\lambda})$ and the Kaplan–Meier estimate $\hat{S}(t)$ in Figure 4.4. There is no indication that the model is inadequate.

Let us obtain approximate .90 confidence intervals for failure probabilities $F(t)$, given by expression (1.3.24). Table 4.2 shows intervals based on four methods:

1. The Wald interval $\hat{\psi} \pm 1.645 se(\hat{\psi})$, where $\psi = F(t)$ is given by (1.3.24) and $se(\hat{\psi})$ is obtained by a straightforward but tedious application of the asymptotic variance formula (B2).
2. The Wald interval based on transforming $\hat{\delta} \pm 1.645 se(\hat{\delta})$, where $\hat{\delta} = \log(\hat{\psi}/(1-\hat{\psi}))$ and $se(\hat{\delta}) = se(\hat{\psi})/(\hat{\psi}(1-\hat{\psi}))$.
3. The nonparametric percentile bootstrap method (Efron and Tibshirani 1993, Ch. 13).
4. The bias-corrected (BC) bootstrap method (Efron and Tibshirani 1993, Ch. 14).

Results are shown in Table 4.2 for $F(.5)$, $F(1.0)$, and $F(1.5)$. Except for $F(.5)$ there is reasonable agreement across the four methods, given the widths of the intervals. In the case of $F(.5)$, the intervals based on methods (2) and (4) are to be preferred.

4.2.3 Models with Polynomial-Based Hazard Functions

As discussed in Section 1.3.9, a variety of lifetime models beyond the standard log-location-scale, gamma, and inverse Gaussian models are sometimes used. Examples of settings that are not well described by any of the standard models are when the hazard is bathtub-shaped or bimodal. Models in which $h(t)$ or some transform of it are low-order polynomials sometimes provide a reasonable fit to data in such cases. Taking $\log h(t)$ to be of polynomial form is attractive, since no restrictions on the parameter values are required. The Gompertz distribution, for which $\log h(t) = \alpha_0 + \alpha_1 t$, has closed-form expressions for density and s.f.'s and has been used a good deal. Models with polynomials of degree two or higher, which can represent nonmonotonic hazards, do not give a closed form for $S(t)$. This is not in principle a major impediment to their use, given the availability of numerical integration software for evaluation of $H(t)$ and $S(t)$.

Models for which $h(t)$ is polynomial are more easily handled, since $H(t)$ is also polynomial in that case. However, the requirement $h(t) \geq 0$ means that parameters must satisfy certain constraints. Models with fractional powers of t, for example, $h(t) = \alpha_0 + \alpha_1 t^{1/2} + \alpha_2 t$, are also easy to handle. The use of a quadratic model for $h(t)$ is illustrated in Example 4.3.1 of Section 4.3. More flexible families of models can be provided by taking $h(t)$ or $\log h(t)$ to be piecewise polynomial. This leads to (regression) spline models introduced in Section 1.3.8; we will consider cubic splines briefly.

A cubic regression spline model for $h(t)$ is one for which $h(t)$ is piecewise cubic. Let $a_1 < \cdots < a_k$ be a sequence of specified cut points or "knots," and define $a_0 = 0$ and $a_{k+1} = \infty$; the cubic spline then consists of cubic polynomials over the intervals (a_{j-1}, a_j), which are forced to join smoothly at the knots. In particular, $h(t)$ is defined so as to be continuous and have continuous first and second derivatives at $a_1, \ldots, a_k$. The model can be represented parametrically in various ways; one simple form is

$$h(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \sum_{j=1}^{k} \theta_j (t - a_j)_+^3, \tag{4.2.9}$$

where a_+ denotes $\max(a, 0)$. With $k = 0$, (4.2.9) defines $h(t)$ as a single cubic polynomial, and with $k \geq 1$ the hazard function consists of $k+1$ cubic pieces. It is easily seen that if there are $k+1$ pieces, the number of parameters in the model is $k+4$, reflecting the fact that $h(t)$ has the aforementioned continuity restrictions.

Splines are also associated with smoothing procedures discussed in Section 3.4 and were used in Example 3.4.1. Our interest in them here is as parametric models, and we rarely want to consider k bigger than two or three, so that the total number of parameters is seven or fewer. The parametric form (4.2.9) is usually poor for computation or estimation, it being preferable to have parameters that are at least roughly orthogonal. In practice, representations

$$h(t; \boldsymbol{\alpha}) = \sum_{j=1}^{p} \alpha_j B_j(t)$$

in terms of known functions $B_j(t)$ are normally used.

Cubic splines with one or two well-chosen knots can provide flexible enough models for $h(t)$ to fit a wide range of lifetime data. Because a fitted cubic polynomial may extrapolate poorly and is sensitive to changes in the data, the cubic piece over (a_k, ∞) is sometimes replaced with a linear function. Various presentations of spline models have been given in the literature in conjunction with censored data. See Rosenberg (1995) for spline models for $h(t)$; Kooperberg et al. (1995) for spline models for $\log h(t)$; Abrahamowicz et al. (1992) for spline models for the p.d.f. $f(t)$, and Kooperberg and Stone (1992) for spline models for $\log f(t)$.

The log-likelihood function based on a censored random sample (t_i, δ_i), $i = 1, \ldots, n$, can be written in the form (2.2.17), giving

$$\ell(\boldsymbol{\alpha}) = \sum_{i=1}^{n} [\delta_i \log h(t_i; \boldsymbol{\alpha}) - H(t_i; \boldsymbol{\alpha})], \tag{4.2.10}$$

where $h(t;\boldsymbol{\alpha})$ and $H(t;\boldsymbol{\alpha})$ are the hazard and cumulative hazard functions, and $\boldsymbol{\alpha}$ denotes unknown parameters. If $h(t;\boldsymbol{\alpha})$ is piecewise polynomial, then so is $H(t;\boldsymbol{\alpha})$, and (4.2.10) has a closed form. More generally, numerical integration is needed to evaluate $\ell(\boldsymbol{\alpha})$. It is possible to treat the cut points or knot positions a_j as parameters, or to prespecify them. The latter is easier, although when the a_j are based on an inspection of the data, the precision of estimates is overstated.

It should be said that spline models with even one or two knots have fairly many parameters, and should not be considered a substitute for parsimonious parametric models. Their main use is in difficult settings where simpler parametric models appear inadequate.

4.3 GROUPED, INTERVAL CENSORED, OR TRUNCATED DATA

4.3.1 Grouped Lifetimes

Grouped lifetime data are interval-censored data where each individual has the same potential observation intervals. In particular, suppose that lifetimes are observed to fall into $k+1$ intervals $I_j = [a_{j-1}, a_j)$, $j = 1, \ldots, k+1$, where $0 = a_0 < a_1 < \cdots < a_k < a_{k+1} = \infty$. Let d_j be the number of lifetimes in I_j, from a random sample of size n. In settings where it is possible to see individuals or units only at the time points $a_1, \ldots, a_k$, we often know only the d_j and not the exact lifetimes for each individual.

When lifetimes from a continuous distribution are grouped, estimation can be based on the exact multinomial likelihood function for the observed data $(d_1, \ldots, d_k)$. If the underlying distribution of T has c.d.f. $F(t;\boldsymbol{\theta})$, then $(d_1, \ldots, d_k)$ has a multinomial probability function

$$\frac{n!}{d_1! \cdots d_k! d_{k+1}!} \pi_1^{d_1} \cdots \pi_k^{d_k} \pi_{k+1}^{d_{k+1}}, \tag{4.3.1}$$

where $\pi_j = Pr(a_{j-1} \leq T < a_j) = F(a_j;\boldsymbol{\theta}) - F(a_{j-1};\boldsymbol{\theta})$. The likelihood function for $\boldsymbol{\theta}$ can therefore be taken as

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{k+1} [F(a_j;\boldsymbol{\theta}) - F(a_{j-1};\boldsymbol{\theta})]^{d_j}. \tag{4.3.2}$$

Maximization of $\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$ can usually be easily achieved with general-purpose optimization software. The score function and information matrix for $\boldsymbol{\theta}$ are, respectively,

$$\frac{\partial \ell}{\partial \boldsymbol{\theta}} = \sum_{j=1}^{k+1} \frac{d_j}{\pi_j} \frac{\partial \pi_j}{\partial \boldsymbol{\theta}}$$

$$I(\boldsymbol{\theta}) = \frac{-\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \sum_{j=1}^{k+1} \frac{d_j}{\pi_j^2} \left(\frac{\partial \pi_j}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \pi_j}{\partial \boldsymbol{\theta}'}\right) - \frac{d_j}{\pi_j} \frac{\partial^2 \pi_j}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}, \tag{4.3.3}$$

and inferences about $\boldsymbol{\theta}$ can be based on the general procedures described in Appendix C.

If censoring can occur in intervals other than the last (i.e., there can be withdrawals in some intervals), the exact likelihood function cannot be written down without further assumptions. One possibility is that all withdrawals occur at the ends of intervals. In this case, if w_j represents the number of withdrawals in I_j, the likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{k+1} [F(a_j;\boldsymbol{\theta}) - F(a_{j-1};\boldsymbol{\theta})]^{d_j} S(a_j;\boldsymbol{\theta})^{w_j}. \tag{4.3.4}$$

If, however, withdrawals occur at unknown times within I_j, then some assumption about the withdrawal mechanism is needed, just as in the case of life table estimation (see Section 3.6 and Problems 3.17, 3.18). The likelihood (3.7.4) in Problem 3.18 is often a reasonable adjustment to (4.3.4).

Example 4.3.1. The data given in Table 4.3 are from results concerning the time to second failure for 104 bus motors (Davis 1952), with time being the number of thousand miles driven. The data suggest a model with a nonmonotonic hazard function, in particular, with the bathtub shape discussed in Section 1.2.4. One family of models that might be considered is that with quadratic hazard functions $h(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2$. In this case, the cumulative hazard function is $H(t;\boldsymbol{\theta}) = \theta_1 t + \theta_2 t^2 + \theta_3 t^3$, and the distribution function is $F(t;\boldsymbol{\theta}) = 1 - \exp[-H(t;\boldsymbol{\theta})]$, where $\theta_j = \alpha_{j-1}/j$ $(j = 1, 2, 3)$.

Using the likelihood (4.3.2) with $k = 6$, $(a_1, \ldots, a_6) = (20, 40, 60, 80, 100, 120)$, and $(d_1, \ldots, d_7) = (19, 13, 13, 15, 15, 18, 11)$, we rewrite the model as $H(t;\boldsymbol{\theta}') = \theta_1'(t/100) + \theta_2'(t/100)^2 + \theta_3'(t/100)^3$ for numerical stability. We then find using general optimization software that the m.l.e. for $\boldsymbol{\theta}'$ is $\hat{\boldsymbol{\theta}}' = (1.315, -1.695, 1.747)$, giving

$$H(t;\hat{\boldsymbol{\theta}}) = 1.315(t/100) - 1.695(t/100)^2 + 1.747(t/100)^3.$$

To assess the agreement between the fitted model and the data, we can calculate expected frequencies

$$e_j = 104[F(a_j;\hat{\boldsymbol{\theta}}) - F(a_{j-1};\hat{\boldsymbol{\theta}})] \qquad j = 1, \ldots, 7.$$

Table 4.3. Frequency Distribution for Bus Motor Failure Data

Thousands of Miles	0–20	20–40	40–60	60–80	80–100	100–120	≥ 120
Observed frequency	19	13	13	15	15	18	11
Expected frequency	19.64	12.29	12.44	15.70	17.43	14.47	12.04

This gives the expected frequencies shown in Table 4.3. The agreement between expected and observed frequencies is quite good. The Pearson chi-squared goodness-of-fit statistic $\sum(d_j - e_j)^2/e_j$ (see Section 10.2.3) gives a value of 1.41 and an associated p-value of $Pr(\chi^2_{(3)} \geq 1.41) = .703$, indicating no evidence against the family of models. Further discussion of goodness-of-fit tests is provided in Chapter 10.

4.3.2 Interval-Censored Data

Interval censored data, as described in Sections 2.3.1 and 3.5.3, generate likelihood functions of the form (2.3.1):

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} [F(R_i; \boldsymbol{\theta}) - F(L_i; \boldsymbol{\theta})], \tag{4.3.5}$$

where $F(t; \boldsymbol{\theta})$ is the c.d.f. for lifetime and the ith lifetime has been observed to lie in the interval $(L_i, R_i]$. The score function and information matrix are, respectively,

$$\frac{\partial \ell}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{\partial \Delta F_i / \partial \boldsymbol{\theta}}{\Delta F_i} \tag{4.3.6}$$

$$I(\boldsymbol{\theta}) = \frac{-\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \sum_{i=1}^{n} \left\{ \left(\frac{\partial \Delta F_i / \partial \boldsymbol{\theta}}{\Delta F_i} \right) \left(\frac{\partial \Delta F_i / \partial \boldsymbol{\theta}'}{\Delta F_i} \right) - \frac{\partial^2 \Delta F_i / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}{\Delta F_i} \right\}, \tag{4.3.7}$$

where $\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$, and we use the notation $\Delta F_i = F(R_i; \boldsymbol{\theta}) - F(L_i; \boldsymbol{\theta})$.

The m.l.e. $\hat{\boldsymbol{\theta}}$ can be found by solving $\partial \ell / \partial \boldsymbol{\theta} = \mathbf{0}$ or by direct maximization of $\ell(\boldsymbol{\theta})$, and inferences about $\boldsymbol{\theta}$ can be based on the methods of Appendix C. Some survival analysis software can handle interval-censored data for log-location-scale models; see the Computational Notes at the end of the chapter. More generally, general optimization software described in Appendix D can be used.

Example 4.3.2. In Example 3.5.5 we considered some data on the times to the appearance of cracks in metal turbine wheels. Each of 432 wheels was examined on a single occasion; the failure times are thus interval censored, with $(L_i, R_i]$ equal to either $(0, C_i]$ or $(C_i, \infty]$, where C_i is the time of inspection for wheel i. This is often referred to as current-status data.

A nonparametric estimate of the c.d.f. $F(t)$ was obtained in Example 3.5.5. Here we obtain a parametric estimate based on the assumption that

$$F(t; \alpha, \beta) = 1 - \exp[-(t/\alpha)^\beta] \qquad t \geq 0$$

is of Weibull form. Since either $L_i = 0$ or $R_i = \infty$ for each item, (4.3.5) can be written as

$$L(\alpha, \beta) = \prod_{i=1}^{n} F(C_i; \alpha, \beta)^{\delta_i} [1 - F(C_i; \alpha, \beta)]^{1-\delta_i},$$

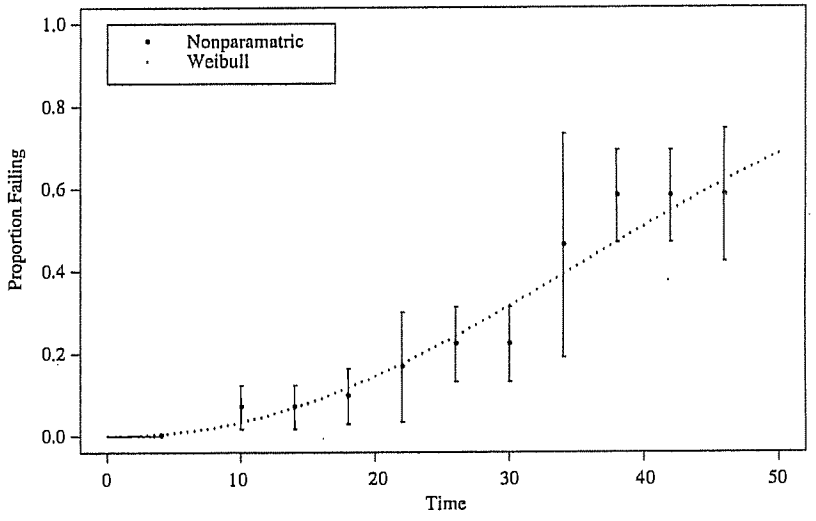

Figure 4.5. Weibull and nonparametric estimates of $F(t)$ for time to crack initiation.

where $\delta_i = I(T_i \leq C_i)$. Maximization of $\ell(\alpha, \beta) = \log L(\alpha, \beta)$ or the equivalent log-likelihood for the extreme value parameters $u = \log \alpha$, $b = \beta^{-1}$ gives the m.l.e.'s $\hat{\alpha} = 46.78$, $\hat{\beta} = 2.176$. Standard errors obtained from the inverse of $I(\hat{\alpha}, \hat{\beta})$, as given by (4.3.7), are 2.99 and .271, respectively. Figure 4.5 shows the Weibull m.l.e. $F(t; \hat{\alpha}, \hat{\beta})$ and the nonparametric estimate $\hat{F}(t)$ obtained in Example 3.5.5, along with the pointwise approximate .95 confidence limits associated with the nonparametric estimate obtained earlier.

The Weibull model provides a reasonable fit to the data, as evidenced by a comparison of the parametric and nonparametric estimates. In many applications there is a desire to use a parametric model to extrapolate beyond the observed failure times, but this is naturally risky. As an illustration, we note that a log-logistic model (1.3.12) also provides a satisfactory fit to the data here. The m.l.e.'s of u and b in (1.3.12) are $\hat{u} = 3.680$ and $\hat{b} = 0.394$. The Weibull and log-logistic estimates of $F(t)$ agree well up to $t = 50$, but very different m.l.e.'s of .960 and .856 are obtained for $F(80)$ under the Weibull and log-logistic models, respectively. The maximum values of the log-likelihood under the two models are -189.29 and -189.73; the Weibull model is slightly favored, but there is no significant difference between them.

4.3.3 Truncated Data

Left- and right-truncated data have been discussed in Sections 2.4 and 3.5. Lifetimes may also be subject to more general forms of truncation or selection. A rather general setting is where a lifetime T_i is forced by the observation or selection process to lie in the interval $(u_i, v_i]$, so that u_i is a left-truncation and v_i a right-truncation time.

The cases $u_i = 0$ and $v_i = \infty$ give ordinary right truncation and left truncation, respectively. If an independent interval censoring mechanism applies, so that T_i is observed to lie in the interval $(L_i, R_i] \subseteq (u_i, v_i]$, then the likelihood function from n independent individuals is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{F(R_i; \boldsymbol{\theta}) - F(L_i; \boldsymbol{\theta})}{F(v_i; \boldsymbol{\theta}) - F(u_i; \boldsymbol{\theta})}, \tag{4.3.8}$$

where $F(t; \boldsymbol{\theta})$ is the parametric c.d.f. for an untruncated lifetime.

Maximum likelihood estimation and associated inference procedures can be implemented using methods described in Appendix D. Some survival analysis software handles truncation and interval censoring for log-location-scale models; see the Computational Notes at the end of the chapter. It should be noted that severe truncation limits the amount of information about model parameters, producing likelihood functions that are flat in certain regions, and possibly nonelliptical. This can render maximization of $\ell(\boldsymbol{\theta})$ more difficult, and large-sample inference procedures inaccurate. We consider two examples of truncated data to illustrate the differences between mild and severe truncation effects.

Example 4.3.3. (Example 3.5.3 revisited). Examples 2.4.4 and 3.5.3 discussed right-truncated data on the time T from HIV infection to AIDS for a group of 124 persons aged 5–59, whose HIV infections resulted from blood transfusions. The right truncation arose from the fact that for an individual to be included in the data set they had to be diagnosed with AIDS by June 30, 1986. The AIDS latency times, t_i, and truncation times, v_i, in months, are given in Appendix G.

Figure 3.16 showed the nonparametric m.l.e. of $F_R(t; v_{max}) = F(t)/F(v_{max})$, where $F(t)$ is the c.d.f. for latency time and $v_{max} = 99.5$ months is the largest truncation time in the data set. If a parametric model $F(t; \boldsymbol{\theta})$ is specified, then it is possible to estimate the unconditional (untruncated) distribution of T. As we will see, however, such estimates are not precise, and it is impossible to differentiate among models that give very different estimates.

Consider a Weibull model with c.d.f.

$$F(t; \alpha, \beta) = 1 - \exp[-(t/\alpha)^{\beta}] \qquad t \geq 0.$$

The special case of (4.3.8), which corresponds to right-truncation only ($u_i = 0$), and exact observation of t_i ($F_{R_i}; \boldsymbol{\theta}) - F(L_i; \boldsymbol{\theta}) \propto f(t_i; \boldsymbol{\theta})$), gives the likelihood function

$$L(\alpha, \beta) = \prod_{i=1}^{n} \frac{f(t_i; \alpha, \beta)}{F(v_i; \alpha, \beta)}, \tag{4.3.9}$$

where $f(t; \alpha, \beta)$ is the Weibull p.d.f. (1.3.5) with $\alpha = \lambda^{-1}$. To make the likelihood more elliptical we employ the extreme value parameters $u = \log \alpha$, $b = \beta^{-1}$

when maximizing (4.3.9), and find estimates and asymptotic standard errors (given in brackets) from the observed information matrix as $\hat{u} = 5.40(2.96)$, $\hat{b} = .48(.06)$.

Although the extreme value scale parameter b and Weibull shape parameter β are precisely estimated, the parameters u and $\alpha = \exp(u)$ are not. This implies as well that quantiles of T, given by $t_p = \exp\{u + b\log(-\log(1-p))\}$, are also imprecisely estimated. For example, a naive approximate .95 confidence interval for $\alpha = t_{.632}$ is given by $\exp\{\hat{u} \pm 1.96se(\hat{u})\}$, which yields the interval (.7 months, 73,130 months); this is both uninformative and nonsensical.

Figure 4.6 shows contours of the log-likelihood function $\ell(u, b) = \log L(u, b)$, from (4.3.9); the maximum value is $\ell(\hat{u}, \hat{b}) = -35.05$. The lack of information about the parameter u is clearly indicated.

As a check on the Weibull model, we compare the nonparametric estimate of $F(t)/F(99.5)$ from Example 3.5.3 with the Weibull estimate $F(t; \hat{\alpha}, \hat{\beta})/F(99.5; \hat{\alpha}, \hat{\beta})$ in Figure 4.7. The estimates agree well, bearing in mind the large standard errors for the nonparametric estimate at larger values of t.

The truncation in this problem is severe; it is known from other HIV-AIDS studies where truncation of latency times was not an issue that median latency times for the types of individuals represented here are of the order of 10 years. Thus, the data here represent only the lower end of the distribution. Nonparametrically we cannot estimate unconditional probabilities, $F(t)$, or quantiles, t_p, at all, as discussed in

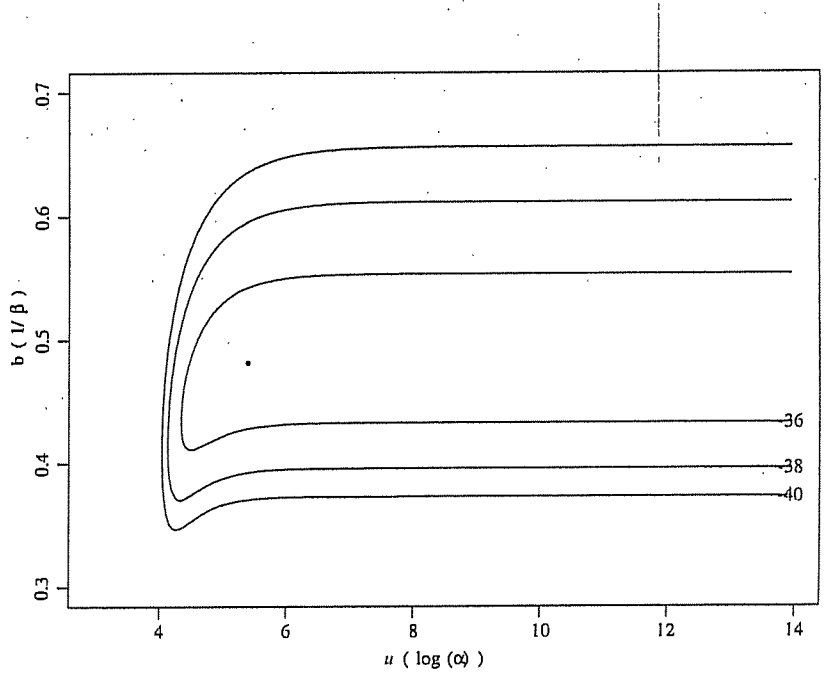

Figure 4.6. Contours of log-likelihood $\ell(u, b)$ for right-truncated AIDS latency-time data.

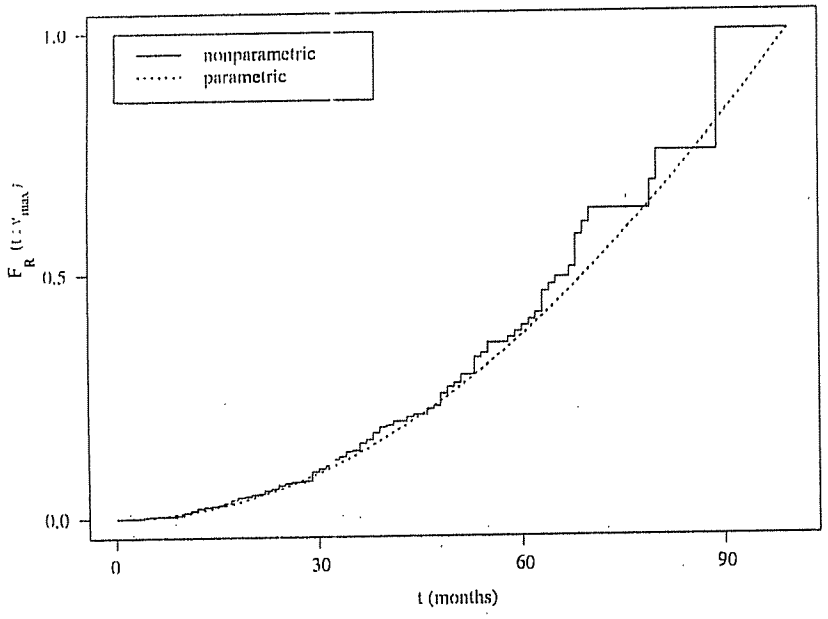

Figure 4.7. Weibull and nonparametric estimates of $F(t)/F(99.5)$ for AIDS latency times.

Example 3.5.3, but even with the assumption of a parametric model there is little information about such quantities.

Example 4.3.4. (Example 3.5.1 revisited). Left-truncated data on the lifetimes T (in thousands of km driven) of automobile brake pads were discussed in Examples 2.4.2 and 3.5.1. A nonparametric estimate of the conditional s.f., $S_L(t; u_{\min}) = S(t)/S(u_{\min})$, was obtained in Example 3.5.1, where $S(t)$ represents the unconditional s.f. for T and $u_{\min} = 7.0$ km is the minimum truncation time across the 98 vehicles represented in the data set. Analysis using several parametric models indicated that a log-normal distribution fits the data well; a plot of the nonparametric and log-normal estimates of $S(t)/S(u_{\min})$ was given in Figure 3.14. We comment briefly here on the log-normal model, which was fitted by using the likelihood function (4.3.8) corresponding to left truncation at u_i, with exact observation of t_i. This gives

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{f(t_i; \mu, \sigma)}{S(u_i; \mu, \sigma)},$$

where $f(t; \mu, \sigma)$ and $S(t; \mu, \sigma)$ are the log-normal p.d.f. and survivor functions given by (1.3.10) and (1.3.11), respectively. Maximization of $L(\mu, \sigma)$ gives m.l.e.'s

and standard errors (in brackets), as $\hat{\mu} = 4.109(.045)$, $\hat{\sigma} = .421(.033)$; both μ and σ are precisely estimated. It is known that almost no brake pads have lifetimes less than 7.0 thousand km. Under the log-normal model the unconditional probability $S(u_{\min}; \hat{\mu}, \hat{\sigma})$ is over .99999, so the estimated truncation effect is negligible. Consequently, the log-likelihood function and information about parameters μ and σ is essentially the same as for a complete, untruncated sample of size $n = 98$.

4.4 MIXTURE MODELS

Mixture models were introduced in Section 1.3.10. Continuous mixtures, in which the survivor function for T is of the form (1.3.29), are in principle easy to deal with, provided the model parameters are well identified. In this section we focus on discrete mixtures, where $S(t)$ is of the form (1.3.26). Only models with two components are considered; mixtures with more components are encountered rather infrequently in lifetime distribution applications.

Maximum likelihood estimation with parametric mixture models can be implemented with general optimization software, as described in Appendix D. Depending on the extent to which the component distributions in a mixture overlap, the likelihood function may be flat in certain regions and preclude precise estimation of individual parameters. Determining the m.l.e. may also be difficult in some cases. Discrete mixtures are most conveniently used in settings where the data suggest there are two or more well-separated components to $f(t)$ or $h(t)$. We consider a pair of examples in which this is the case.

Example 4.4.1. (Example 3.3.3 revisited). Colon cancer recurrence times were discussed in Example 3.3.3 for two groups of patients: a Drug Therapy and a Control group. Times to recurrence, T, are measured in days from treatment. Nonparametric Kaplan–Meier estimates of the survivor functions $S(t)$ for each group suggest that the hazard function for recurrence drops to a low value by some point, perhaps because some fraction of patients are cured and will never experience disease recurrence. It was noted in Example 3.3.3 that standard distributions for which $S(t) \to 0$ for t large, such as the Weibull and log-logistic, do not fit these data.

Plausible models are ones for which a fraction $1 - p$ of patients is assumed to have no chance of disease recurrence. These are sometimes referred to as cure-rate models, and have s.f.'s of the form (1.3.29):

$$S(t) = pS_0(t) + 1 - p, \tag{4.4.1}$$

where $0 < p < 1$ and $S_0(t)$ is a s.f. with $S(0) = 1$ and $S(\infty) = 0$. We consider first a model where $S_0(t)$ is of log-logistic form (1.3.13),

$$S_0(t) = \{1 + (t/\alpha)^{\beta}\}^{-1} \qquad t \geq 0. \tag{4.4.2}$$

The p.d.f. corresponding to (4.4.1) is then $-pS_0'(t)$, or

$$f(t;\alpha,\beta) = \frac{p(\beta/\alpha)(t/\alpha)^{\beta-1}}{\{1+(t/\alpha)^{\beta}\}^2} \qquad t \geq 0.$$

The likelihood function from a censored sample of recurrence times is

$$L(\alpha,\beta,p) = \prod_{i=1}^{n} f(t_i;\alpha,\beta,p)^{\delta_i} S(t_i;\alpha,\beta,p)^{1-\delta_i},$$

where $S(t;\alpha,\beta,p)$ is given by (4.4.1) and (4.4.2). Parameter estimates and standard errors (in brackets) for the Control and Therapy groups are obtained with no difficulty by standard optimization procedures, and are as follows:

Control :	$\hat{\alpha} = 419.5(39.8)$,	$\hat{\beta} = 1.58(0.13)$,	$\hat{p} = 0.608(0.032)$
Therapy :	$\hat{\alpha} = 479.0(51.1)$,	$\hat{\beta} = 1.68(0.17)$,	$\hat{p} = 0.426(0.032)$.

A plot of the s.f. estimates $S(\cdot;\hat{\alpha},\hat{\beta},\hat{p})$ is shown in Figure 4.8 as Model (1) for each of the treatment groups, along with the Kaplan–Meier estimates for each group;

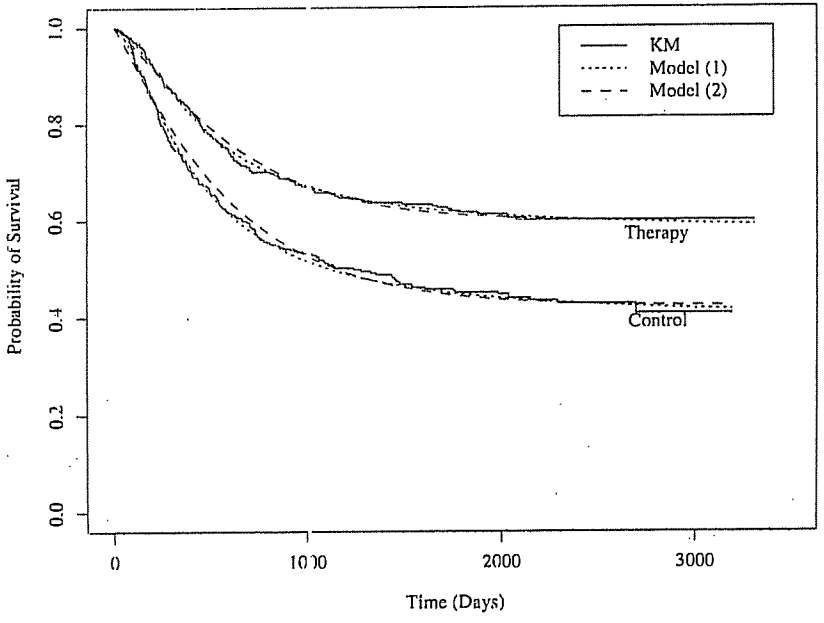

Figure 4.8. Parametric mixture and Kaplan–Meier estimates of $S(t)$ for colon cancer recurrence.

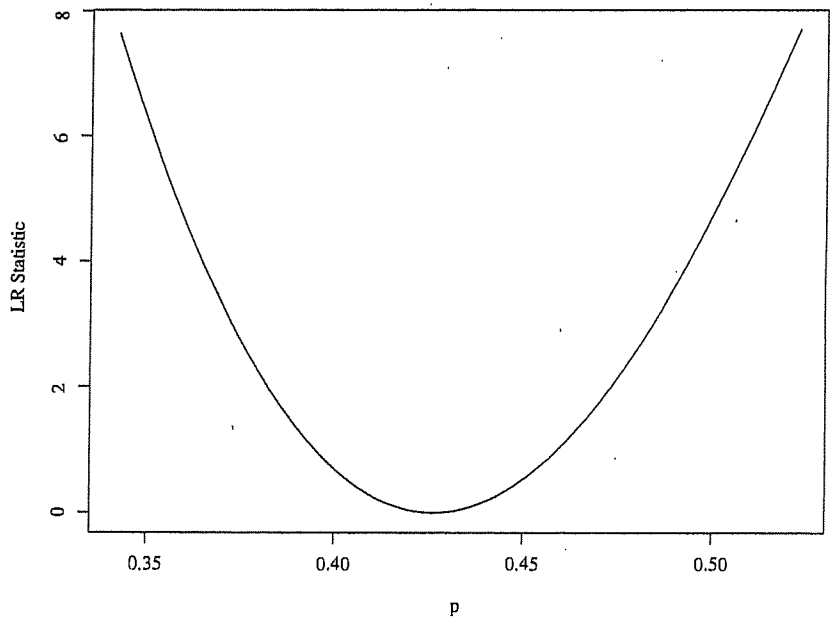

Figure 4.9. Likelihood ratio statistic for the proportion of patients, p, who experience cancer recurrence.

the parametric model fits well. Parametric estimates of $S(t)$ based on (4.4.1) with a Weibull distribution $S_0(t) = \exp\{-(t/\alpha)^{\beta}\}$ in place of (4.4.2) are also plotted as Model (2). The Weibull mixture model fits well, too, though slightly less so than the log-logistic mixture.

The mixture models provide estimates $1 - \hat{p}$ of the fraction of the population who are long-term survivors (i.e., have no cancer recurrence). In settings where censoring is heavy there may be high correlations among the m.l.e.'s, and an imprecise estimate of p. That is not the case here, because follow-up of individuals was long and very few failures with large values of t were seen. Confidence intervals for p or $1-p$ by the naive Wald method and by the likelihood ratio statistic $\Lambda(p) = 2\ell(\hat{\alpha},\hat{\beta},\hat{p}) - 2\ell(\tilde{\alpha}(p),\tilde{\beta}(p),p)$ agree quite closely. For example, $\Lambda(p)$ is plotted for the Therapy group and the log-logistic model in Figure 4.9, and is seen to be approximately quadratic. Approximate .95 confidence intervals $\hat{p} \pm 1.96se(\hat{p})$ and $\{p : \Lambda(p) \leq 3.84\}$ for the Therapy group both give approximately $.36 \leq p \leq .49$. The corresponding confidence interval for the Control group is $.55 \leq p \leq .67$, indicating a clear difference in the proportion of long-term survivors under the two treatments.

Example 4.4.2. (Example 3.4.1 revisited). Example 3.4.1 discussed data on the times to failure for 60 electrical appliances subjected to a life test. Exploration

of the data there suggested that the hazard function might have two components, one consisting of a fairly small portion of the distribution and giving small failure times, and one giving a wide range of larger failure times. A discrete mixture may be plausible physically, with the left-most component representing items with defects that make them liable to fail early, and so we explore this possibility.

Examination of the data and of the Nelson–Aalen and Kaplan–Meier estimates can suggest plausible values for p, and parametric models for the components $S_1(t)$, $S_2(t)$ of a mixture with survivor function

$$S(t) = pS_1(t) + (1 - p)S_2(t) \tag{4.4.3}$$

and corresponding p.d.f.

$$f(t) = pf_1(t) + (1 - p)f_2(t). \tag{4.4.4}$$

For example, the Kaplan–Meier estimate $\hat{S}_{KM}(t)$ shown in Figure 4.10 suggests a value of p in the .1–.2 or .8–.9 range; a plot of $-\log \hat{S}_{KM}(t)$ or of the Nelson–Aalen estimate $\hat{H}_{NA}(t)$ suggests that models for $S_1(t)$ and $S_2(t)$ with monotone hazard functions may be suitable. Consequently, we will consider a mixture of two Weibull

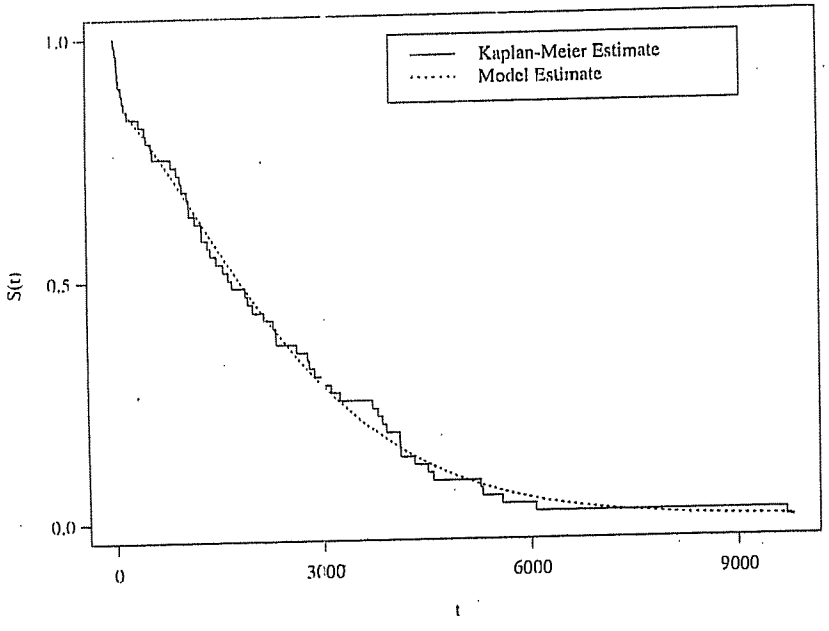

Figure 4.10. Parametric mixture and Kaplan–Meier estimates of $S(t)$ for electrical appliance failures.

components, with $S_1(t)$ and $S_2(t)$ in (4.4.3) given by

$$S_j(t) = \exp[-(t/\alpha_j)^{\beta_j}] \qquad j = 1, 2. \tag{4.4.5}$$

The log-likelihood function based on the 60 failure times $t_1, \ldots, t_{60}$ given in Example 3.4.1 is of the form

$$\ell(\alpha_1, \beta_1, \alpha_2, \beta_2, p) = \sum_{i=1}^{60} \log f(t_i; \alpha_1, \beta_1, \alpha_2, \beta_2, p), \tag{4.4.6}$$

with $f(t)$ given by (4.4.4) with $f_j(t) = -S_j'(t)$, $j = 1, 2$. The model (4.4.4) and (4.4.5) with the two parameter vectors $(\alpha, \beta, \alpha', \beta', p)$ and $(\alpha', \beta', \alpha, \beta, 1 - p)$ are the same, so without loss of generality we make the restriction $0 < p < .5$. The m.l.e.'s are then readily found by using general optimization software to maximize (4.4.6); estimates and standard errors (in brackets) are $\hat{\alpha}_1 = 95.4(25.8)$, $\hat{\beta}_1 = 1.66(.49)$, $\hat{\alpha}_2 = 2774.5(314.2)$, $\hat{\beta}_2 = 1.40(.18)$, $\hat{p} = .137(.051)$. Asymptotic correlations of the m.l.e.'s are all under .46. Initial estimates for the optimization procedure are suggested by examination of $\hat{S}_{KM}(t)$ and the data: $p = .15$ and values for α_1 and α_2 (which are the .632 quantiles for $S_1(t)$ and $S_2(t)$) of approximately 100 and 2800 seem reasonable. Initial values for β_1 and β_2 of 1 are often effective, and used here, though plots developed from the Kaplan–Meier or Nelson–Aalen estimates can also be used to suggest estimates.

Figure 4.10 shows plots of the estimated survivor function

$$\hat{S}(t) = \hat{p}\, \exp[-(t/\hat{\alpha}_1)^{\hat{\beta}_1}] + (1 - \hat{p}) \exp[-(t/\hat{\alpha}_2)^{\hat{\beta}_2}]$$

and of the Kaplan–Meier estimate. The Weibull mixture clearly agrees with the observed data.

4.5 THRESHOLD PARAMETERS

As mentioned in Section 1.3, threshold parameters are occasionally introduced into models; these are values $\gamma > 0$ such that lifetimes must satisfy the restriction $T \geq \gamma$. For example, the three-parameter Weibull distribution includes a threshold parameter γ and has p.d.f.

$$f(t; \gamma, \alpha, \beta) = \frac{\beta}{\alpha}\left(\frac{t-\gamma}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{t-\gamma}{\alpha}\right)^{\beta}\right] \qquad t \geq \gamma. \tag{4.5.1}$$

The existence of a time γ before which failure is impossible is sometimes plausible, but data are often quite uninformative concerning its value. In addition, nonstandard behavior can occur when maximum likelihood methods are applied to such models. We will consider these issues briefly, then examine Weibull and exponential distributions with threshold parameters.

4.5.1 General Remarks

If a model has p.d.f. of the form

$$f(t; \boldsymbol{\theta}, \gamma) = f_0(t - \gamma; \boldsymbol{\theta}) \qquad t \geq \gamma \tag{4.5.2}$$

and if (t_i, δ_i), $i = 1, \ldots, n$ is a censored random sample of lifetimes, then the parameter γ must satisfy $\gamma \leq t_{(1)} = \min(t_i)$, since we require $f(t_i; \boldsymbol{\theta}, \gamma) \geq 0$ and $0 \leq S(t_i; \boldsymbol{\theta}, \gamma) \leq 1$ for each t_i. Peculiarities in the likelihood function

$$L(\boldsymbol{\theta}, \gamma) = \prod_{i=1}^{n} f_0(t_i - \gamma; \boldsymbol{\theta})^{\delta_i} S_0(t_i - \gamma; \boldsymbol{\theta})^{1-\delta_i} \tag{4.5.3}$$

can arise. For example, for the three-parameter log-normal distribution where $\log(T - \gamma)$ is $N(\mu, \sigma^2)$ it can be seen (see Problem 4.20) that $L(\mu, \sigma, \gamma)$ is unbounded as γ approaches $t_{(1)}$ from below. The same behavior occurs for the three-parameter Weibull model (4.5.1) if $\beta < 1$, since in that case $f(t) \to \infty$ as $t \to \gamma+$. This behavior can be avoided if we recognize that there is a finite degree of precision within which a continuous variate can be measured, and replace values for the p.d.f. $f(t_i)$ in the likelihood function with $Pr(t_i - \Delta \leq T_i \leq t_i + \Delta) = F(t_i + \Delta) - F(t_i - \Delta)$, where Δ represents the precision of measurement. If Δ is small, then $F(t_i + \Delta) - F(t_i - \Delta)$ is typically very close to $2\Delta f(t_i)$ and there is no need for this adjustment, but in the case of $t_{(1)}$ it keeps the likelihood finite and restricts γ to be $\leq t_{(1)} - \Delta$. An alternative approach, due to Cheng and Iles (1987), is to leave the restriction as $\gamma \leq t_{(1)}$, but replace $f(t_{(1)})$ in the likelihood function with $F(t_{(1)} + \Delta) - F(t_{(1)})$.

Estimation of γ is also generally nonregular for certain values of $\boldsymbol{\theta}$ in a model (4.5.2); this is discussed for the Weibull model in Section 4.5.2. In practice a convenient and satisfactory approach is to estimate $\boldsymbol{\theta}$ and to examine the fit of models (4.5.2) with γ assumed known. This is easy to do, since when γ is known we simply consider observations $s_i = t_i - \gamma$ for the lifetime distribution $f_0(s_i; \boldsymbol{\theta})$, for which estimation is typically regular. Plausible values for γ can be determined from the profile likelihood function

$$L_p(\gamma) = L(\hat{\boldsymbol{\theta}}(\gamma), \gamma)$$

where $L(\boldsymbol{\theta}, \gamma)$ is given by (4.5.3) and $\hat{\boldsymbol{\theta}}(\gamma)$ is the m.l.e. for $\boldsymbol{\theta}$ when γ is known. A plot of $L_p(\gamma)$ shows plausible values for γ, and in some cases it is possible to calibrate $L_p(\gamma)$ by reference to a $\chi^2_{(1)}$ distribution for an associated likelihood ratio statistic. The fit of models for fixed values of γ can be assessed informally through plots or more formally through the methods in Chapter 10. Unboundedness of the likelihood function (4.5.3) manifests itself in the profile $L_p(\gamma)$. This is usually not a problem, but if necessary we can replace $f(t_{(1)}; \boldsymbol{\theta}, \gamma)$ with either $F(t_{(1)} + \Delta; \boldsymbol{\theta}, \gamma) - F(t_{(1)} - \Delta; \boldsymbol{\theta}, \gamma)$ or $F(t_{(1)} + \Delta; \boldsymbol{\theta}, \gamma) - F(t_{(1)}; \boldsymbol{\theta}, \gamma)$, as described earlier. This makes the likelihood function (4.5.3) and $L_p(\gamma)$ bounded. The details associated

with different models vary slightly, but the discussion in the next subsection for the Weibull distribution deals with the crucial points.

Some authors allow the parameter γ to take on any real value, whereas we insist that $\gamma \geq 0$. Although there are certain mathematical advantages to leaving γ arbitrary, it is naturally nonnegative in lifetime models. Sometimes an estimation procedure may produce a negative confidence limit for γ; in that case we replace the limit with 0. Cox and Hinkley (1974, pp. 224–226) provide discussion on this point.

4.5.2 The Three-Parameter Weibull Distribution

The log-likelihood function from a censored random sample (t_i, δ_i), $i = 1, \ldots, n$ arising from (4.5.1) is, from (4.5.3),

$$\ell(\alpha, \beta, \gamma) = r \log \beta - r\beta \log \alpha + (\beta - 1) \sum_{i=1}^{n} \delta_i \log(t_i - \gamma) - \sum_{i=1}^{n} \left(\frac{t_i - \gamma}{\alpha} \right)^{\beta}, \tag{4.5.4}$$

where $r = \sum \delta_i$ is the number of uncensored lifetimes. This function is unbounded, since for any $\beta < 1$, $\ell(\alpha, \beta, \gamma) \to \infty$ as $\gamma \to t_{(1)}-$. Consequently, a solution to the likelihood equations $\partial\ell/\partial\alpha = 0$, $\partial\ell/\partial\beta = 0$, $\partial\ell/\partial\gamma = 0$ does not produce a global maximum for the likelihood. It appears from empirical investigation (e.g., Pike 1966; Rockette et al. 1974; Lockhart and Stephens 1994) that the likelihood equations must have two or fewer solutions. It has been proven that when $\ell(\alpha, \beta, \gamma)$ has a local maximum, there is also a second solution to the likelihood equations that gives a saddle point.

Situations in which the Weibull distribution is used with a threshold parameter typically have $\beta \geq 1$, and we restrict attention to this case. With the restriction $\beta \geq 1$, the likelihood function is bounded and it may have a local maximum at a point $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ with $\hat{\alpha} > 0$, $\hat{\beta} > 1$, $\hat{\gamma} < t_{(1)}$. When there is no local maximum, $\ell(\alpha, \beta, \gamma)$ is maximized over the region with $\alpha > 0$, $\beta \geq 1$, $\gamma \leq t_{(1)}$ by

$$\hat{\gamma} = t_{(1)}, \qquad \hat{\beta} = 1, \qquad \hat{\alpha} = \sum_{i=1}^{n} \frac{t_i - \hat{\gamma}}{r}. \tag{4.5.5}$$

This can also give the global maximum of $\ell(\alpha, \beta, \gamma)$ when a local maximum exists, so it is necessary to compare the likelihood function values at (4.5.5) and at the local maximum in order to determine the global maximum. In this way, maximization of $\ell(\alpha, \beta, \gamma)$ gives essentially the same estimate as does maximization of the likelihood obtained by adjusting $f(t_{(1)}; \alpha, \beta, \gamma)$ as described in the preceding section, assuming the restriction $\beta \geq 1$ is retained.

A good way to obtain the m.l.e. and determine plausible values for γ is to compute the profile log-likelihood function $\ell_p(\gamma) = \ell(\hat{\alpha}(\gamma), \hat{\beta}(\gamma), \gamma)$, where $\hat{\alpha}(\gamma)$ and $\hat{\beta}(\gamma)$ are the m.l.e.'s of α and β, with γ held fixed. These estimates are easily found by treating the values $s_i = t_i - \gamma$ as a censored sample from the two-parameter Weibull distribution. As described in Section 5.2, estimates are provided by many software packages. A graph of $\ell_p(\gamma)$ shows plausible values of γ and in most cases allows $\hat{\gamma}$

to be accurately determined; note that only values $\gamma \leq t_{(1)}$ are considered and that we restrict attention to values $\beta \geq 1$. When the m.l.e. $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ has been obtained, the log relative profile likelihood or the corresponding likelihood ratio statistic $\Lambda(\gamma) = 2\ell(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) - 2\ell_p(\gamma)$ can also be calculated. Asymptotic theory shows that if $\beta > 2$, then $\Lambda(\gamma_0)$ has a limiting $\chi^2_{(1)}$ distribution when $\gamma_0 > 0$ is the true value of γ, though the adequacy of the $\chi^2_{(1)}$ approximation in moderate-size samples has not been thoroughly investigated. If $\beta < 2$, then the limiting distribution of $\Lambda(\gamma_0)$ is not $\chi^2_{(1)}$. In the case where it is known that $\beta = 1$, we have a two-parameter exponential model, and it is noted in the next section that $\Lambda(\gamma_0)$ is then asymptotically $\chi^2_{(2)}$. For practical purposes a conservative guideline is to think of plausible values for γ as ones with $\Lambda(\gamma) \leq 5.99$, which is the .95 quantile for $\chi^2_{(2)}$. If plausible values for β appear to be greater than 2, then we can use instead the $\chi^2_{(1)}$ quantile 3.84. In most cases the log-likelihood function $\ell(\alpha, \beta, \gamma)$ is very flat near the m.l.e. and there is a wide range of plausible values for γ.

The threshold parameter has a different function than α or β in (4.5.1). Interval estimation and tests about distribution characteristics are best carried out with γ treated as known; the sensitivity of inferences to variation in γ can be examined. Point estimates and confidence intervals for α and β often vary widely as γ is varied; this reflects the fact that the likelihood function $\ell(\alpha, \beta, \gamma)$ is flat near the m.l.e., and that the m.l.e.'s are highly correlated. The data are informative only about certain functions of α, β, and γ and not the individual parameters. As illustrated in the example below, estimates of quantiles $t_p = \gamma + \alpha[-\log(1-p)]^{1/\beta}$ are often quite stable as γ varies, and it would then be reasonable to quote confidence limits for t_p with γ assumed equal to $\hat{\gamma}$.

Example 4.5.1. Pike (1966) gave some data from a laboratory investigation in which the vaginas of rats were painted with the carcinogen DMBA, and the number of days T until a carcinoma appeared was recorded. The data below are for a group of 19 rats (Group 1 in Pike's paper); the two observations with asterisks are censoring times.

143, 164, 188, 188, 190, 192, 206, 209, 213, 216, 220, 227, 230, 234, 246, 265, 304, 216*, 244*

These data were given in Problem 3.5, where it was suggested that probability plots for Weibull distributions with and without a threshold parameter be considered. We will fit the three-parameter model to the data here.

Table 4.4 shows estimates $\hat{\alpha}(\gamma)$, $\hat{\beta}(\gamma)$, and associated values of the profile log-likelihood $\ell_p(\gamma)$. The estimates $\hat{\alpha}(\gamma)$, $\hat{\beta}(\gamma)$, and log-likelihood values are obtained by maximizing (4.5.4) with γ fixed, which is the two-parameter Weibull log-likelihood function (see Section 5.2), with t_i replaced by $t_i - \gamma$. From this we see that a local maximum of $\ell(\alpha, \beta, \gamma)$ occurs at about $\hat{\gamma} = 122$, $\hat{\alpha} = 108.4$, $\hat{\beta} = 2.712$, though the log-likelihood is very flat in the region of this point. Values of the profile relative likelihood function $R_p(\gamma) = L_p(\gamma)/L_p(\hat{\gamma})$ and of the likelihood ratio

Table 4.4. m.l.e.'s and Profile Relative Likelihood for γ

γ	$\hat{\alpha}(\gamma)$	$\hat{\beta}(\gamma)$	$\ell_p(\gamma)$	$R_p(\gamma)$	$\Lambda(\gamma)$
0	234.3	6.08	−88.233	.403	1.818
60	173.2	4.49	−87.831	.602	1.015
100	131.8	3.38	−87.467	.867	.285
110	121.2	3.08	−87.381	.945	.113
120	110.6	2.78	−87.327	.998	.004
122	108.4	2.71	−87.324	1.000	.000
125	105.2	2.61	−87.330	.994	.012
130	99.7	2.44	−87.382	.944	.115
135	94.0	2.24	−87.542	.804	.436
140	88.0	1.99	−88.064	.477	1.480
142	85.2	1.80	−88.773	.235	2.896
143	81.1	1.00	−91.718	.012	8.846

statistic $\Lambda(\gamma) = -2\log R_p(\gamma) = 2\ell(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) - 2\ell_p(\gamma)$ are also given in Table 4.4. Figure 4.11 shows plots of these two functions; although the relative likelihood function is equivalent to the likelihood ratio statistic, which we normally plot, we show $R_p(\gamma)$ here as well to emphasize the shape of the profile likelihood function for γ. It is seen that no values of γ in (0,143) are particularly implausible except for those very close to 143. Note that we have restricted β to be ≥ 1; if we allow values $\beta < 1$, there is a local minimum of $\ell_p(\gamma)$ very close to $t_{(1)} = 143$, and both $\ell_p(\gamma)$ and $\ell(\alpha, \beta, \gamma)$ become arbitrarily large as $\gamma \to 143$.

Note that $\hat{\alpha}(\gamma)$ and $\hat{\beta}(\gamma)$ vary considerably with γ. Unless we are willing to restrict γ to a narrow range, precise estimation of α or β is not possible. However, estimates of quantiles or survival probabilities are quite stable as γ varies. For example, the m.l.e. for the pth quantile with a given value of γ is

$$\hat{t}_p(\gamma) = \gamma + \hat{\alpha}(\gamma)[-\log(1-p)]^{1/\hat{\beta}(\gamma)}.$$

Estimates for $p = .10, .50, .90$ at $\gamma = 60, 100, 140$ are as follows:

$$\begin{array}{lll} t_{.10}(60) = 164.9 & t_{.50}(60) = 219.6 & t_{.90}(60) = 268.6 \\ t_{.10}(100) = 167.7 & t_{.50}(100) = 218.2 & t_{.90}(100) = 268.7 \\ t_{.10}(140) = 168.3 & t_{.50}(140) = 213.2 & t_{.90}(140) = 274.0. \end{array}$$

One would expect these estimates to be stable, because Weibull plots of the data with the different γ values indicate an adequate fit to the data in each instance.

The value $\gamma = 0$ is of special interest because it corresponds to the absence of a threshold for failure. With these data $\gamma = 0$ is clearly plausible, and there is no need to consider a formal test. If a test were of interest, it would be satisfactory to use $\Lambda(\gamma)$ as a test statistic, with a p-value computed with the approximation $Pr(\Lambda(0) \geq \Lambda_{obs}) = .5Pr(\chi^2_{(1)} \geq \Lambda_{obs})$. This is because, whereas $\Lambda(\gamma_0)$ is asymptotically $\chi^2_{(1)}$ when $\beta > 2$ and $\gamma_0 > 0$, the likelihood ratio statistic $\Lambda(0)$ has a limiting distribution

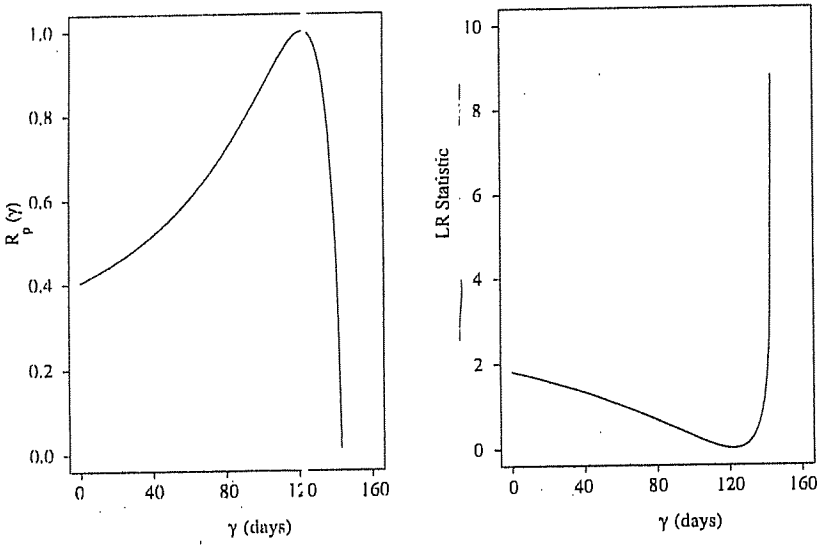

Figure 4.11. Relative likelihood function and likelihood ratio statistic for Weibull threshold parameter (time to vaginal cancer data).

when $\gamma = 0$ with $Pr(\Lambda(0) = 0) = .5$ and $Pr(\Lambda(0) > x) = .5Pr(\chi^2_{(1)} > x)$ for $x > 0$. Alternatively, if we allow γ and thus $\hat{\gamma}$ to be negative, then $\Lambda(0)$ has a $\chi^2_{(1)}$ limiting distribution when $\gamma = 0$; a one-sided test of $\gamma = 0$ against alternatives with $\gamma > 0$ then gives the same p-value as for the case with restriction $\gamma \geq 0$.

4.5.3 The Two-Parameter Exponential Distribution

The two-parameter exponential distribution has p.d.f.

$$f(t; \theta, \gamma) = \frac{1}{\theta} e^{-(t-\gamma)/\theta} \qquad t \geq \gamma. \tag{4.5.6}$$

Although there seem to be relatively few documented applications of this model to real data, it is rather easy to deal with and has received considerable theoretical attention. It is also of interest as a special case of the Weibull model in the preceding section. A few results are summarized here.

The log-likelihood function from a censored random sample (t_i, δ_i), $i = 1, \ldots, n$ is

$$\ell(\theta, \gamma) = -r \log \theta - \sum_{i=1}^{n} \frac{(t_i - \gamma)}{\theta}, \tag{4.5.7}$$

where $r = \sum \delta_i$ is the number of uncensored lifetimes. Bearing in mind that $\gamma \leq t_{(1)}$, the smallest observed time, it is easy to see that for any $\theta > 0$, $\ell(\theta, \gamma)$ is maximized at $\gamma = t_{(1)}$. Since $\ell(\theta, t_{(1)})$ is maximized at $\sum(t_i - t_{(1)})/r$, the m.l.e. based on (4.5.7) is

$$\hat{\theta} = \sum_{i=1}^{n} (t_i - t_{(1)})/r, \qquad \hat{\gamma} = t_{(1)}. \tag{4.5.8}$$

As elsewhere, it is tacitly assumed that $r > 0$; when $r = 0$, the likelihood does not possess a finite maximum.

Tests and interval estimates for θ are readily obtained, and are essentially the same as procedures for θ in the one-parameter exponential model, given in Section 4.4.1, with $t_i - t_{(1)}$ used in place of t_i. For example, the likelihood ratio statistic for θ is

$$\begin{aligned} \Lambda(\theta) &= 2\ell(\hat{\theta}, \hat{\gamma}) - 2\ell(\theta, \hat{\gamma}(\theta)) \\ &= 2r \left[\frac{\hat{\theta}}{\theta} - 1 - \log\left(\frac{\hat{\theta}}{\theta}\right) \right], \end{aligned} \tag{4.5.9}$$

and we can use the limiting $\chi^2_{(1)}$ distribution of $\Lambda(\theta)$ in order to obtain tests or confidence intervals.

Inferences about γ via its likelihood ratio statistic are also straightforward, though the limiting distribution for $\Lambda(\gamma)$ turns out to be $\chi^2_{(2)}$ rather than the regular $\chi^2_{(1)}$ (Hogg 1956). The likelihood ratio statistic is $\Lambda(\gamma) = 2\ell(\hat{\theta}, \hat{\gamma}) - 2\ell(\hat{\theta}(\gamma), \gamma)$, and since $\hat{\theta}(\gamma) = \sum(t_i - \gamma)/r$, we find from (4.5.7) that

$$\Lambda(\gamma) = 2r \log \left[1 + \frac{n(\hat{\gamma} - \gamma)}{r\hat{\theta}} \right]. \tag{4.5.10}$$

In the case of Type 2 censored data, discussed in Section 2.2.1, exact distributional results are available. In this case the t_i values with $\delta_i = 1$ are $t_{(1)} < \cdots < t_{(r)}$, the first r order statistics, and those with $\delta_i = 0$ are equal to $t_{(r)}$; the m.l.e.'s are as given in (4.5.8). It is readily seen that $\hat{\theta}$ and $\hat{\gamma}$ are jointly sufficient for θ and γ; their distributions are given in the following theorem.

THEOREM 4.5.1. Let $\hat{\theta}$ and $\hat{\gamma}$ be the m.l.e.'s based on a Type 2 censored sample consisting of the r smallest observations in a random sample of n from the distribution (4.5.6). Then $\hat{\theta}$ and $\hat{\gamma}$ are independent, and $2n(\hat{\gamma} - \gamma)/\theta$ and $2r\hat{\theta}/\theta$ are distributed as $\chi^2_{(2)}$ and $\chi^2_{(2r-2)}$, respectively.

Proof. The random variables $T_{(1)} - \gamma, \ldots, T_{(r)} - \gamma$ are the first r order statistics in a random sample of size n from the one-parameter exponential distribution (1.3.3). By Theorem 4.1.1, the quantities $W_1 = n(T_{(1)} - \gamma)$ and

$$W_i = (n - i + 1)(T_{(i)} - T_{(i-1)}) \qquad i = 2, \ldots, r$$

are independent and have one-parameter exponential distributions. Hence $2n(\hat{\gamma} - \gamma)/\theta = 2W_1/\theta \sim \chi^2_{(2)}$ and

$$\frac{2r\hat{\theta}}{\theta} = 2\sum_{i=2}^{r}\frac{W_i}{\theta} \sim \chi^2_{(2r-2)},$$

by the same arguments as in Theorem 4.1.1.

Confidence intervals or tests for θ or γ are easily obtained from the results of the theorem. Inferences for θ can be based on the pivotal quantity $2r\hat{\theta}/\theta \sim \chi^2_{(2r-2)}$, and inferences for γ can be based on the pivotal

$$\frac{n(r-1)(\hat{\gamma}-\gamma)}{r\hat{\theta}} \sim F_{(2,2r-2)}. \tag{4.5.11}$$

That this has an F distribution follows directly from the results of Theorem 4.5.1. Quantiles for $F_{(2,2r-2)}$ have a closed form: the pth quantile is

$$F_{(2,2r-2),p} = (r-1)[(1-p)^{-1/(r-1)} - 1]. \tag{4.5.12}$$

It is rather easy to show (see Problem 4.18) that when r becomes large, confidence limits for θ and γ based on these pivotal quantities become the same as those based on the likelihood ratio statistics (4.5.9) and (4.5.10), used in conjunction with a $\chi^2_{(1)}$ and $\chi^2_{(2)}$ distribution, respectively.

Confidence intervals for quantiles t_p or for $S(t)$ are awkward to obtain via standard large-sample methods in the general case of censored data. For Type 2 censored data, however, confidence intervals for $t_p = \gamma + \theta[-\log(1-p)]$ can be based on the pivotal quantity

$$Z_p = \frac{\hat{\gamma} - t_p}{\hat{\theta}}. \tag{4.5.13}$$

This is readily seen to be pivotal by the results of Theorem 4.5.1; see (4.5.14). If $z_{p,q}$ is the qth quantile of Z_p, then $\hat{\gamma} - z_{p,q}\hat{\theta}$ is a lower q confidence limit for t_p. Confidence intervals for $S(t)$ can also be obtained from (4.5.13). To get a lower q confidence limit for $S(t_0)$ for a specified t_0, we determine p such that $\hat{\gamma} - z_{p,q}\hat{\theta} = t_0$; see the discussion in Section 3.2.3 concerning the relationship between confidence intervals for quantiles and survival probabilities.

The distribution of (4.5.13) has been studied by various authors; an exact formula and an approximation that can be used to get confidence limits are given below. However, Z_p has a simple representation in terms of independent χ^2 random variables: by simple manipulation of (4.5.13) and Theorem 4.5.1, we have that

$$Z_p = \frac{rV_1}{nV_2} + \frac{2r[-\log(1-p)]}{V_2}, \tag{4.5.14}$$

where $V_1 \sim \chi^2_{(2)}$ and $V_2 \sim \chi^2_{(2r-2)}$ are independent. Quantiles or survival probabilities for Z_p are easily obtained to any desired degree of accuracy by simulation.

Engelhardt and Bain (1978b) gave expressions that can be used instead of simulation. If p and q are such that $(1-p)^n \geq 1-q$, then the lower q confidence limit for t_p has closed form

$$\hat{\gamma} + \frac{r}{n}\left[1 - \left(\frac{(1-p)^n}{1-q}\right)^{1/(r-1)}\right]\hat{\theta}. \tag{4.5.15}$$

The q lower confidence limit on $S(t_0)$ in the same case is

$$(1-q)^{1/n}\left[1 - \frac{n(t_0-\hat{\gamma})}{r\hat{\theta}}\right]^{(r-1)/n}. \tag{4.5.16}$$

For the case where $(1-p)^n < 1-q$, there is no exact expression for the confidence limits. Engelhardt and Bain give the following approximations and show they are sufficiently accurate for virtually all practical purposes. The lower q confidence limit for t_p is

$$\hat{\gamma} + \frac{r}{n}\left[-m(p) - N_q\left(\frac{m^2(p)}{r} + \frac{1}{r^2}\right)^{1/2}\right]\hat{\theta}, \tag{4.5.17}$$

where $m(p) = [1 + n\log(1-p)]/(r-2.5)$ and N_q is the q quantile for the standard normal distribution. The corresponding lower q confidence limit for $S(t_0)$ is

$$\exp\left[-\frac{1}{n} + \frac{r(r-2.5)}{an}\left(Y - \frac{N_q}{r}\left(rY^2 + a\right)^{1/2}\right)\right], \tag{4.5.18}$$

where $Y = n(\hat{\gamma} - t_0)/\hat{\theta}$ and $a = r^2(1 - N_q/r)$.

Although (4.5.13) is not an exact pivotal quantity in the case of arbitrarily censored data, it is approximately pivotal in large samples. A reasonable approach to confidence interval estimation or tests in that case is to use the nonparametric bootstrap (see Appendix D.2) to estimate the distribution of Z_p. A satisfactory approach is to select n observations (t_i, δ_i) from the observed data, with replacement, then to compute estimates $\hat{\gamma}^*$ and $\hat{\theta}^*$ and the value $z_p^* = (\hat{\gamma}^* - \hat{t}_p)/\hat{\theta}^*$, where $\hat{t}_p = \hat{\gamma} + [-\log(1-p)]\hat{\theta}$ is the m.l.e. from the observed data. Repeating this B times (say $B = 1000$), we consider the z_p^* as a random sample from the distribution of Z_p, and use it to estimate quantiles or survival probabilities.

Example 4.5.2. Engelhardt and Bain (1978b) and others considered data on the mileages at which 19 military personnel carriers failed in service. There is no cen-

soring, and the mileages are

162, 200, 271, 320, 393, 508, 539, 629, 706, 777,
884, 1008, 1101, 1182, 1463, 1603, 1984, 2355, 2880.

Since there is no censoring, Theorem 4.1.1 tells us that if model (4.5.6) is correct, then conditional on $T_{(1)} = t_{(1)}$, the variables $T_{(i)} - t_{(1)}$ $(i = 2, \ldots, 19)$ have the distribution of the order statistics of a sample of size 18 from a one-parameter exponential distribution with mean θ. A probability plot of the values $t_{(i)} - t_{(1)}$, as described in Section 3.3.1, indicates that an exponential model is consistent with the data.

Maximum likelihood estimates of θ and γ from (4.5.8) are $\hat{\gamma} = 162$, $\hat{\theta} = 835.2$. Confidence intervals for γ can be obtained from the fact that $18(\hat{\gamma} - \gamma)/\hat{\theta} \sim F_{(2,36)}$, by (4.5.11) with $r = n = 19$. For example, since $Pr(F_{(2,36)} \leq 3.254) = .95$, we get the .95 confidence interval $\gamma \geq \hat{\gamma} - 3.254\hat{\theta}/18$, or $11.0 \leq \gamma (\leq 162)$, which is very wide. For a two-sided .95 confidence interval for θ, we use $38\hat{\theta}/\theta \sim \chi^2_{(36)}$ and $Pr(21.38 \leq \chi^2_{(36)} \leq 54.40) = .95$ to get $583.4 \leq \theta \leq 1484$.

Let us also obtain a lower .90 confidence limit for the quantile $t_{.10}$ of the distribution. We can use (4.5.15) in this case, or determine the distribution of $Z_{.10}$ in (4.5.13) by simulation. We use the former since it is exact; this gives the interval $t_{.10} \geq 114.6$.

4.6 PREDICTION INTERVALS

Some applications involve the prediction of future observations in a population or process, based on existing data. For example, one may wish to predict the number of parts that will need to be replaced in a system over the next three months, or the time to platelet recovery for a leukemia patient who has received a bone marrow transplant. Prediction is different than estimation of a distributional characteristic because we are interested in a finite number (perhaps only one) of individuals rather than the entire conceptual population that the distribution represents.

Suppose that a future observation is represented by the random variable Y, with c.d.f. $f(y; \boldsymbol{\theta})$. If $\boldsymbol{\theta}$ is known, then the quantities $y_\alpha(\boldsymbol{\theta})$, satisfying $Pr(Y \leq y_\alpha(\boldsymbol{\theta}); \boldsymbol{\theta}) = \alpha$, provide prediction limits for Y. The so-called plug-in method of setting prediction limits with $\boldsymbol{\theta}$ unknown is to replace $\boldsymbol{\theta}$ with an estimate, $\hat{\boldsymbol{\theta}}$, based on existing data. The nominal α upper prediction limit $y_\alpha(\hat{\boldsymbol{\theta}})$ does not in this case satisfy either $Pr(Y \leq y_\alpha(\hat{\boldsymbol{\theta}})|\hat{\theta}; \boldsymbol{\theta}) = \alpha$ or $Pr(Y \leq y_\alpha(\hat{\boldsymbol{\theta}}); \boldsymbol{\theta}) = \alpha$, where we consider both Y and $\hat{\boldsymbol{\theta}}$ as random variables. If $\hat{\boldsymbol{\theta}}$ is based on a large sample then, assuming it is a consistent estimator of $\boldsymbol{\theta}$, the preceding probabilities will typically be close to α. With small data sets it is sensible to recognize the uncertainty inherent in $\hat{\boldsymbol{\theta}}$. This leads to the concept of prediction intervals, which we now discuss briefly.

To start, let $Y_1, \ldots, Y_n$ be a random sample from a distribution $F(y; \boldsymbol{\theta})$ in some parametric family, and let Y' represent an independent "future" observation from the same distribution. It is assumed that $\boldsymbol{\theta}$ is unknown, but can be estimated from $Y_1, \ldots, Y_n$. An α prediction interval for Y' is a random interval

$[A(Y_1, \ldots, Y_n), B(Y_1, \ldots, Y_n)]$ such that

$$Pr[A(Y_1, \ldots, Y_n) \leq Y' \leq B(Y_1, \ldots, Y_n)] = \alpha. \tag{4.6.1}$$

Such "exact" prediction intervals can be obtained only in certain situations, but if there exists a pivotal random variable

$$U = g(Y_1, \ldots, Y_n, Y') \tag{4.6.2}$$

whose distribution is free of $\boldsymbol{\theta}$, and if probability statements

$$Pr(a \leq U \leq b) = \alpha \tag{4.6.3}$$

can be inverted into the form (4.6.1), then exact prediction intervals exist. The interval $[A(y_1, \ldots, y_n), B(y_1, \ldots, y_n)]$ based on observed data $(y_1, \ldots, y_n)$ is a realized prediction interval and has a similar interpretation to a confidence interval.

Example 4.6.1. Let $Y_1, \ldots, Y_n$ be a random sample from the exponential distribution (1.3.3) with c.d.f. $F(y; \theta) = 1 - \exp(-y/\theta)$, and let Y' be an independent future observation from the same distribution. By Corollary 4.1.1 of Theorem 4.1.1, it follows that $2Y'/\theta \sim \chi^2_{(2)}$ and $2\sum_{i=1}^n Y_i/\theta \sim \chi^2_{(2n)}$, and therefore that

$$U = nY'/\sum_{i=1}^{n} Y_i \sim F_{(2,2n)} \tag{4.6.4}$$

is a pivotal quantity. Thus, letting $F_{(2,2n),\alpha}$ be the α quantile for the $F_{(2,2n)}$ distribution and noting that $\hat{\theta} = \overline{Y}$ is the m.l.e. of θ, we have $Pr(U \leq F_{(2,2n),\alpha}) = \alpha$, and so

$$Pr(Y' \leq \hat{\theta} F_{(2,2n),\alpha}) = \alpha. \tag{4.6.5}$$

Thus $\hat{\theta} F_{(2,2n),\alpha}$ is an α upper prediction limit for the future observation Y'. For example, if $n = 10$, then $F_{(2,20),.95} = 3.49$ and the .95 upper prediction limit is $3.49\hat{\theta}$.

By comparison, $y_{.95}(\theta) = 3.00\theta$ for the exponential distribution, so the plug-in .95 prediction limit would be $3.00\hat{\theta}$. The unconditional probability $Pr(Y' \leq 3.00\hat{\theta})$ is substantially less than .95 in this case. As n increases, the proper .95 prediction limit $\hat{\theta} F_{(2,2n),.95}$ approaches $3.00\hat{\theta}$, reflecting the fact that $\hat{\theta}$ converges to the true value of θ as $n \to \infty$. For $n = 30$ and $n = 60$, for example, the .95 limits obtained from (4.6.5) are $3.15\hat{\theta}$ and $3.07\hat{\theta}$. For $n = 60$ there would be little harm in simply using the plug-in limit.

The preceding example is special in yielding exact prediction intervals. If there had been a Type 1 censored random sample (t_i, δ_i), $i = 1, \ldots, n$, for example, then exact intervals would no longer exist. This is analogous to the situation concerning

confidence intervals for θ, where exact intervals are unavailable with Type 1 censored data. If the random variable $\mathbf{Y}$ represents potential observed data, then, as with confidence intervals, we can look for prediction intervals $[A(\mathbf{Y}), B(\mathbf{Y})]$ for which

$$Pr[A(\mathbf{Y}) \le Y' \le B(\mathbf{Y})] \to \alpha \tag{4.6.6}$$

as the sample size $n \to \infty$. If the probability on the left approaches α sufficiently fast as n increases, then for sufficiently large n the interval may reasonably be termed an approximate α prediction interval for the future observation Y'.

Plug-in limits based on consistent estimators of $\boldsymbol{\theta}$ provide intervals satisfying (4.6.6). In particular, if $y'_p(\boldsymbol{\theta})$ is the pth quantile for Y' and $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{Y})$ is consistent, then for α_1, α_2 such that $\alpha_1 + \alpha_2 = 1 - \alpha$ the interval $[y'_{\alpha_1}(\hat{\boldsymbol{\theta}}), y'_{1-\alpha_2}(\hat{\boldsymbol{\theta}})]$ is an approximate α prediction interval. However, with small or moderate sample sizes the actual coverage probability $Pr[A(\mathbf{Y}) \le Y' \le B(\mathbf{Y})]$ may not be as close to α as desired. Two approaches are often used to improve coverage probability accuracy. The first is to look for approximate pivotal quantities $U = g(\mathbf{Y}, Y')$ whose distribution depends very little on $\boldsymbol{\theta}$, even for small sample sizes, and to obtain prediction intervals by inverting probability statements (4.6.3). The second approach is termed calibration, and consists of determining (usually by simulation) the actual coverage probability $\alpha'(\boldsymbol{\theta})$ associated with a plug-in prediction interval with nominal coverage α. If desired, the value of α can then be adjusted so as to make $\alpha'(\boldsymbol{\theta})$ greater than or equal to some nominal value. In practice, what is usually done is to adjust α to make $\alpha'(\hat{\boldsymbol{\theta}})$ equal to some desired coverage probability, where $\hat{\boldsymbol{\theta}}$ is the m.l.e. for $\boldsymbol{\theta}$ based on the observed data $\mathbf{y}$.

If $F(y; \boldsymbol{\theta})$ is the c.d.f. for continuous Y', then $F(Y'; \boldsymbol{\theta})$ is a Uniform(0, 1) random variable and appealing approximate pivotal quantities for prediction are

$$U' = F(Y'; \hat{\boldsymbol{\theta}}(\mathbf{Y})) \tag{4.6.7}$$

or monotonic functions of U. The distribution of U can be estimated by simulation for any value of $\boldsymbol{\theta}$ by generating independent data Y' and $\mathbf{Y}$. This is usually done for the value $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ only, where $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{y})$ is the m.l.e. for $\boldsymbol{\theta}$ based on the observed data. This procedure is sometimes referred to as a parametric bootstrap (see Appendix D.2).

It will not be possible to simulate censored data $\mathbf{Y}$ under the estimated model with $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ unless the censoring process is known. One can if necessary generate $\mathbf{Y}$ and $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ using nonparametric bootstrap sampling (see Appendix D.2). Two general notes of caution are that large numbers of simulations may be needed to estimate the distribution of pivotals or to calibrate plug-in prediction limits well, and that the accuracy of approximate methods has been studied only in a few special settings.

Example 4.6.2. Consider the exponential model of Example 4.6.1 and a Type 1 censored random sample $y_i = (t_i, \delta_i)$, $i = 1, \ldots, n$, which arises as follows. Each individual has a lifetime T_i from the exponential distribution and a known potential censoring time, C_i, and we observe $t_i = \min(T_i, C_i)$ and $\delta_i = I(t_i = T_i)$. The m.l.e.

of θ is $\hat{\theta} = \sum t_i / \sum \delta_i$ and the approximate pivotal (4.6.7) is $U = 1 - \exp(-Y'/\hat{\theta})$. If we can closely approximate the distribution of U or a monotonic function of U, such as $W = Y'/\hat{\theta}$, then prediction intervals can be given. In particular, if a and b are such that $Pr(a \le W \le b) = \alpha$, then $[A(\hat{\theta}), B(\hat{\theta})] = (a\hat{\theta}, b\hat{\theta})$ is an (approximate) α prediction interval for Y'.

The calibration approach is essentially the same. For simplicity suppose we want an upper prediction limit $B(\mathbf{Y})$ for Y'. Since $y_\alpha(\theta) = -\theta \log(1 - \alpha)$ is the αth quantile for Y', the plug-in α upper prediction limit is $y_\alpha(\hat{\theta}) = -\hat{\theta} \log(1 - \alpha)$. However, the true coverage probability associated with the prediction limit is

$$h(\alpha) = Pr[Y' \le y_\alpha(\hat{\theta}); \theta], \tag{4.6.8}$$

where we suppress notationally that $h(\alpha)$ also depends on θ. The process of calibration consists of determining the function $h(\alpha)$. Once this is done we can obtain a prediction interval with the desired coverage. For example, for a .95 interval we use the plug-in limit $y_\alpha(\hat{\theta})$ with α chosen so that $h(\alpha) = .95$. Since $y_\alpha(\hat{\theta}) = -\hat{\theta} \log(1 - \alpha)$ in the present setting, (4.6.8) implies that

$$h(\alpha) = Pr[W \le -\log(1 - \alpha); \theta],$$

where $W = Y'/\hat{\theta}$. Thus calibration is here equivalent to determining the distribution of W.

To illustrate the use of simulation for calibration or determination of the distribution of W, let us consider the following artificial example involving an uncensored sample of size $n = 10$, as in Example 4.6.1. In this case $W = Y'/\hat{\theta}$ is distributed exactly as $F_{(2,20)}$ so the accuracy of approximations obtained via simulation can be examined. Suppose the 10 observed lifetimes are 0.695, 0.148, 0.911, 0.344, 1.034, 0.718, 0.296, 1.178, 0.802, 0.825, giving $\hat{\theta} = 0.695$. Table 4.5 shows the exact .05 and .95 quantiles of W along with parametric and nonparametric bootstrap estimates, obtained, respectively, as follows.

1. Independent pseudorandom observations y' and $y^*_1, \ldots, y^*_{10}$ are generated from Exp(.695), giving a value $w^* = y'/\bar{y}^*$. This is repeated B times, giving values $w^*_1, \ldots, w^*_B$.

Table 4.5. Exact and Simulated Quantiles of W

Quantile	Exact Value		Parametric Bootstrap	Nonparametric Bootstrap
$w_{.05}$	.051	$B = 2{,}000$	.060	.054
		$B = 10{,}000$	.045	.051
		$B = 50{,}000$	.051	.049
$w_{.95}$	3.493	$B = 2{,}000$	3.517	3.084
		$B = 10{,}000$	3.418	3.051
		$B = 50{,}000$	3.507	3.079

2. A value y' is generated from Exp(.695) and a nonparametric bootstrap sample $(y_1^*, \ldots, y_{10}^*)$ is generated by sampling 10 values, with replacement, from the observed lifetimes. This gives $w^* = y'/\bar{y}^*$, and the process is repeated B times to give $w_1^*, \ldots, w_B^*$.

In each case the .05 and .95 quantiles of W can be estimated as $w_{(.05B)}^*$ and $w_{(.95B)}^*$.

The parametric bootstrap estimates are very accurate if the value of B is sufficiently large. The nonparametric estimates do not change much as B increases beyond 2000 for the small sample size ($n = 10$) here; the .05 quantile of W is estimated well, but not the .95 quantile. These results should be viewed in the context that the exact .05 and .95 quantiles of $W = Y'/\hat{\theta}$ are .051 and 3.493. The nonparametric bootstrap here improves only slightly on the plug-in method, which gives the .95 quantile as 2.996.

The discussion so far has dealt with the prediction of a single independent observation Y' from the distribution under consideration. Predictions for functions of two or more random variables may also be of interest, say

$$V = g(Y_1', \ldots, Y_m').$$

For example, V might be the sum $\sum Y_i'$ or the rth order statistic $Y_{(r)}'$, which are both of interest in reliability contexts. If the c.d.f. $F_V(v; \boldsymbol{\theta})$ is available in closed form, then it may be possible to use $U = F_V(V; \hat{\boldsymbol{\theta}})$ as an approximate (or in some special cases, exact) pivotal quantity from which prediction intervals for V can be obtained. In most such cases, it will be necessary to estimate the distribution of U by simulation, as illustrated in Example 4.6.2. The calibration approach can also be used and is equivalent to the pivotal method. This involves calibrating plug-in prediction limits $v_\alpha(\hat{\boldsymbol{\theta}})$ for V, where $F_V[v_\alpha(\boldsymbol{\theta}); \boldsymbol{\theta}] = \alpha$, and can be done using simulation. In some applications V may not be independent of $\mathbf{Y}$ and $\hat{\boldsymbol{\theta}}$. This does not complicate matters substantially provided that the joint distribution of V and $\hat{\boldsymbol{\theta}}$ can be approximated by simulation.

Prediction of discrete random variables can also be of interest, for example, $V = \sum I(Y_i' > y_0)$, the number of lifetimes among future $Y_1', \ldots, Y_m'$, that exceed some stated value y_0. In this case, intervals with exact nominal coverage probabilities such as .95 or .99 usually don't exist even if $\boldsymbol{\theta}$ is known, but intervals with approximately such coverage can be sought. The best approach is calibration of plug-in prediction limits. The following example involves a discrete variable V that is also not independent of $\hat{\boldsymbol{\theta}}$.

Example 4.6.3. The lifetimes T of certain electromechanical units can be assumed to follow a distribution $F(t; \boldsymbol{\theta})$. Suppose that n units enter service at the same time and that after a time τ has elapsed, $r < n$ of the units have failed. It is wished to obtain prediction limits for the number of remaining units V that will fail in the time interval (τ, τ').

Given r, V has a Binomial$(m, p(\boldsymbol{\theta}))$ distribution, where $m = n - r$ and

$$p(\boldsymbol{\theta}) = \frac{F(\tau'; \boldsymbol{\theta}) - F(\tau; \boldsymbol{\theta})}{1 - F(\tau; \boldsymbol{\theta})}.$$

If $\boldsymbol{\theta}$ were known, we could determine, say, an α upper prediction limit $B_\alpha(\boldsymbol{\theta}, r)$ as the smallest integer such that

$$Pr[V \le B_\alpha(\boldsymbol{\theta}, r) | r; \boldsymbol{\theta}] \ge \alpha. \tag{4.6.9}$$

The plug-in prediction limit is $B_\alpha(\hat{\boldsymbol{\theta}}, r)$, where $\hat{\boldsymbol{\theta}}$ is the m.l.e. of $\boldsymbol{\theta}$ based on the data $\mathbf{y}$ observed up to time τ; this consists of r and information about the times of the r failures over $(0, \tau)$. The calibration approach is then to estimate the unconditional coverage probability $h(\alpha) = Pr[V^* \le B_\alpha(\hat{\boldsymbol{\theta}}^*, r^*); \boldsymbol{\theta}]$, where $(V^*, \hat{\boldsymbol{\theta}}^*, r^*)$ are random variables representing the data over $(0, \tau)$ and the number of failures over (τ, τ'). This is done by assuming that $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(\mathbf{y})$, the m.l.e. based on the observed data. The probability $h(\alpha)$ can be estimated by simulation. Noticing that

$$h(\alpha) = E\{Pr[V^* \le B_\alpha(\hat{\boldsymbol{\theta}}^*, r^*) | \hat{\boldsymbol{\theta}}^*, r^*]\},$$

and that the probability inside the expectation is given by the distribution Binomial$(n - r^*, p(\hat{\boldsymbol{\theta}}))$, allows us to avoid simulating values V^*.

In many prediction problems the distribution of V may be analytically intractable, necessitating the calculation of even plug-in limits by simulation. Calibration is then more computation-intensive. If the data set on which the plug-in limits are based is moderately large, it is usually safe to forgo calibration. Since it is generally advisable to assess the effects of model variation on prediction, one can replace calibration with a sensitivity analysis on the effects of changes in $\boldsymbol{\theta}$ and the model on plug-in prediction limits.

We conclude by mentioning that Bayesian methods of prediction are attractive. Let $\pi(\boldsymbol{\theta})$ be a prior distribution for $\boldsymbol{\theta}$ and let

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{L(\boldsymbol{\theta}; \mathbf{y})\pi(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}; \mathbf{y})\pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

be the posterior distribution for $\boldsymbol{\theta}$, given observed data $\mathbf{y}$ that provide the likelihood function $L(\boldsymbol{\theta}; \mathbf{y})$. The Bayesian predictive distribution for the independent future observation, V, given $\mathbf{y}$, then has probability density or mass function

$$f(v|\mathbf{y}) = \int f_V(v; \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta}, \tag{4.6.10}$$

where $f_V(v; \boldsymbol{\theta})$ is the p.d.f. or probability mass function for V when $\boldsymbol{\theta}$ is known. If V and $\mathbf{Y}$ are not independent, then $f_V(v; \boldsymbol{\theta})$ is replaced by $f_V(v|\mathbf{y}; \boldsymbol{\theta})$ in (4.6.10).

Bayesian prediction is readily implemented in many problems using numerical methods or simulation. As the sample size giving $\mathbf{y}$ increases, Bayesian α-probability

prediction intervals converge to plug-in α prediction intervals. Aitchison and Dunsmore (1975) and Geisser (1993) provide overviews of the Bayesian approach.

BIBLIOGRAPHIC NOTES

The exponential distribution was featured in many early papers on lifetime distributions, particularly with reference to industrial life testing; see, for example Sukhatme (1937), Epstein and Sobel (1953, 1954, 1955), Epstein (1954), Bartholomew (1957), Mendenhall (1958) and Govindarajulu (1964). Johnson, Kotz, and Balakrishnan (1994, Ch. 19) provide a more extensive list of references. Anscombe (1964) and Sprott (1973) emphasized the use of parameter transformations to improve the accuracy of large-sample procedures as in Section 4.1.

Industrial life test plans under exponential distributions have been thoroughly studied in the early references just cited, and later by Aroian (1976), Bryant and Schmee (1979), Kao et al. (1979), and others. Life test acceptance plans published in various reliability standards (see Blischke and Murthy 2000, pp. 697–701) have been based on this work; a note of caution is that these plans are sensitive to departures from the assumed exponential model (e.g., Zelen and Dannemiller 1961; Harter and Moore 1976; Fryer and Holt 1976). Plans for comparative experiments have been studied in the context of clinical trials and other areas. Early examples are found in Armitage (1975), Breslow and Haug (1972), and Louis (1977), where sequential plans are emphasized. Bernstein and Lagakos (1978), Rubinstein et al. (1981), Lachin and Foulkes (1986), and others provide detailed examinations. Books on clinical trials (e.g., Whitehead 1992; Piantadosi 1997) discuss the planning of comparative experiments under various types of assumptions.

Inference for the gamma model has been considered by Engelhardt and Bain (1978a), Chao and Glaser (1978) and others for the complete data case. The inverse Gaussian model has been considered by Chhikara and Folks (1977, 1989), Jorgensen (1981), and Whitmore (1975, 1983). Johnson et al. (1994, Chs. 17, 15) provide numerous references for the gamma and inverse Gaussian models.

Models with polynomial hazard functions were considered by Bain (1974), Canfield and Borgman (1975), and Gaver and Acar (1979). The Gompertz model with $\log h(t)$ linear has been widely studied, and Gehan and Siddiqui (1973) consider models for which some transform of $h(t)$ is linear in the parameters. The use of piecewise polynomial functions, especially splines that are everywhere smooth, has received a great deal of recent attention. In addition to the preceding references, see, for example, Etazadi-Amoli and Ciampi (1987) and, for a discrete-time application, Efron (1988). Additional references are given in the Bibliographic Notes for Chapter 3.

Silvapulle and Burridge (1986) discuss unimodality properties of likelihood functions based on grouped data; Heitjan (1989) provides a review of inference procedures. Lindsey (1998) discusses inference based on interval-censored data for parametric models, and Lindsey and Ryan (1998) discuss models with piecewise polynomial hazard functions. The parametric treatment of truncated data is similarly

straightforward in principle, but if truncation is severe, one may encounter likelihood functions that are uninformative about certain parameters. Kalbfleisch and Lawless (1988b, 1989) consider applications where this is the case.

In some applications censoring times may be missing for censored individuals; for example, see Suzuki (1985a,b, 1995), Kalbfleisch and Lawless (1988a, b), Hu and Lawless (1996), and Hu et al. (1998). This book does not address this topic except in Problem 3.11, and likewise does not deal with missing data on covariates. Lawless et al. (1999) provide some general discussion and references on these topics.

There is a substantial literature on discrete mixtures of parametric models (e.g., Titterington et al. 1985; Böhning 2000), with mixtures of normal (Folkes 1979; Aitkin and Wilson 1980; Johnson et al. 1994, Sec. 13.10.2), exponential (Johnson et al. 1994, Sec. 19.9), and Weibull (Kao 1959; Falls 1970) distributions among well-studied models. Maller and Zhou (1996) consider cure-rate models as in Example 4.4.1. Meeker and Escobar (1998, Sec.11.5) consider similar models in reliability applications. For examples of continuous mixture models, see Whitmore (1986).

Inferences about threshold parameters have been studied a good deal, in part because of the possibility of nonregular asymptotic behavior for certain parameter values. Cheng and Traylor (1995) provide a survey of the area; see also Smith (1985, 1995). Estimation for the three-parameter Weibull model has been studied extensively (e.g., Pike 1966; Rockette et al. 1974; Lemon 1975; Lockhart and Stephens 1994; Smith 1995). The two-parameter exponential distribution has also been widely studied, particularly for the case of complete or Type 2 censored data (e.g., see Engelhardt and Bain 1978b; Pierce 1973 and references therein). Johnson et al. (1994) contains many additional references on Weibull and exponential models with threshold parameters.

Prediction problems are considered by Aitchison and Dunsmore (1975), Hahn and Meeker (1991), Geisser (1993), and Meeker and Escobar (1998, Ch. 12; 1999), who all give numerous references and examples of applications. Beran (1990) and Hall et al. (1999) consider calibration using bootstrap simulations. Barndorff-Nielsen and Cox (1994, Sec. 9.4; 1996) discuss different approaches to prediction, and asymptotic coverage properties for prediction intervals. Aitchison and Dunsmore (1975), Geisser (1993), and Meeker and Escobar (1998, Ch. 14) discuss Bayesian prediction intervals. For Bayesian point prediction, see Skouras and Dawid (1998).

COMPUTATIONAL NOTES

The gamma, inverse Gaussian and polynomial hazard function models of Section 4.2 are not included in major statistical software packages, but it is relatively easy to implement maximum likelihood methods using standard optimization software, as discussed in Appendix D. Interval censoring and truncation for data from the common log-location-scale models (Weibull, log-normal, log-logistic) are handled by several packages, including S-Plus. Software for fitting discrete parametric mixtures to censored data is discussed by Böhning (2000). With censored data we can use optimization software, with care taken to explore the shape of the likelihood function for models with several parameters.

PROBLEMS AND SUPPLEMENTS

4.1 The following data are times $t_1, \ldots, t_n$ between successive failures of air conditioning equipment in a Boeing 720 airplane (Proschan, 1963): 74, 57, 48, 29, 502, 12, 70, 21, 29, 386, 59, 27, 153, 26, and 326. Assuming that the data come from an exponential distribution with mean θ, compare confidence intervals for θ obtained by using the approximate pivotals below with exact intervals obtained by using the fact that $2\sum t_i/\theta \sim \chi^2_{(2n)}$: (1) $\sqrt{n}(\hat{\theta} - \theta)/\hat{\theta} \sim N(0, 1)$; (2) $3\sqrt{n}\hat{\phi}^{-1}(\hat{\phi} - \phi) \sim N(0, 1)$, where $\phi = \theta^{-1/3}$; (3) the likelihood ratio statistic.

(Section 4.1)

4.2 Consider the likelihood function $L(\theta)$ obtained in the case of Type 1 censored sampling from the exponential distribution and let $\phi = g(\theta)$ be an arbitrary one-to-one transformation. Show that $(\partial^3 \log L/\partial\phi^3)_{\hat{\phi}} = 0$ if and only if $\phi \propto \theta^{-1/3}$. (The likelihood function for ϕ thus looks more "normal" than that for θ and suggests that treating $\hat{\phi}$ as normally distributed is preferable to treating $\hat{\theta}$ as normally distributed in obtaining confidence intervals.)

(Section 4.1; Anscombe 1964; Sprott 1973)

4.3 The following data are remission times, in weeks, for a group of 30 leukemia patients in a certain type of therapy; starred observations are censoring times: 1, 1, 2, 4, 4, 6, 6, 6, 7, 8, 9, 9, 10, 12, 13, 14, 18, 19, 24, 26, 29, 31*, 42, 45*, 50*, 57, 60, 71*, 85*, 91.

(a) Estimate the median remission time by three methods: (1) by using the nonparametric method of Section 3.2.4; (2) by assuming that the underlying distribution of remission times is exponential; and (3) by assuming that the distribution of remission times is gamma. Compare confidence intervals based on the three methods.

(b) Similarly compare estimates of $S(26)$, the probability a remission lasts more than 26 weeks, using the nonparametric Kaplan–Meier estimate and the two parametric models, respectively.

(c) Is there any evidence against either of the parametric models?

(Sections 4.1, 4.2)

4.4 Suppose that an acceptance plan is desired which, under the one-parameter exponential model, will reject $H_0 : \theta = 1000$ with probability .10 when $\theta = 1000$ hours, and with probability .95 when $\theta = 300$ hours. Obtain Type 2 censored plans, both with and without replacement of failed units, and graph the power functions for the plans.

(Section 4.1)

4.5 *Sensitivity of exponential distribution tests to model departures.*

(a) Let $t_1, \ldots, t_n$ be a complete random sample from an exponential distribution with mean θ. Consider life test plans that test $H_0 : \theta = 1000$ versus $H_1 : \theta < 1000$ and have size 0.10. Graph the power functions of the tests for sample sizes $n = 10$ and $n = 20$.

(b) Suppose that $t_1, \ldots, t_n$ actually come from a Weibull distribution with p.d.f. $(\beta/\alpha)(t/\alpha)^{\beta-1}\exp[-(t/\alpha)^\beta]$, $t > 0$, where $\beta = 1.5$ and $\alpha = \theta/\Gamma(1 + 1/1.5)$; this distribution also has mean θ. It can be shown that the distribution of $\sum t_i/\alpha$ is well approximated by a χ^2 distribution,

$$\sum_{i=1}^{n} \frac{t_i}{\alpha} \sim c\chi^2_{(b)},$$

where c and b are selected so that $c\chi^2_{(b)}$ has the same mean and variance as $\sum t_i/\alpha$. Show that this yields the values $c = [\Gamma(1 + 2/1.5) - \Gamma(1 + 1/1.5)^2]/2\Gamma(1 + 1/1.5)$ and $b = nc^{-1}\Gamma(1 + 1/1.5)$.

(c) Use the χ^2 approximation of part (b) to examine the power function of the tests in part (a) when the underlying distribution is a Weibull distribution with $\beta = 1.5$ rather than an exponential distribution.

(*Remark:* A one-sided size .05 test of $\beta = 1$ vs. $\beta > 1$ in a Weibull model has power at $\beta = 1.5$ approximately equal to .4 and .7 for $n = 10$ and $n = 20$, so this degree of nonexponentiality is not certain to be detected.)

(Section 4.1)

4.6 *Predicting the duration of a life test.* Sometimes it is desired to predict the total duration of a life test on the basis of early results in the test. Suppose, for example, that a test is to terminate at the time $t_{(r)}$ of the rth failure. If the sth failure has just occurred ($1 \le s < r$), we can predict $t_{(r)}$.

(a) If the data came from a one-parameter exponential distribution with mean θ, prove that $t_{(r)} - t_{(s)}$ and

$$T_s = \sum_{i=1}^{s} t_{(i)} + (n - s)t_{(s)}$$

are independent, and that $U = (t_{(r)} - t_{(s)})/T_s$ is pivotal, with distribution function

$$Pr(U \le t) = 1 - \frac{(n-s)!}{(r-s-1)!(n-r)!} \cdot \sum_{i=0}^{r-s-1} \binom{r-s-1}{i} (-1)^i/(n-r+i+1) \cdot [1 + (n-r+i+1)t]^s.$$

Show how U can be used to obtain prediction intervals for $t_{(r)}$, based on $t_{(1)}, \ldots, t_{(s)}$.

(b) Describe how simulation can instead be used to obtain the distribution of U.

(Section 4.1; Lawless 1971)

4.7 The following observations are failure times (in minutes) for a sample of 15 electronic components in an accelerated life test:

1.4, 5.1, 6.3, 10.8, 12.1, 18.5, 19.7, 22.2,

23.0, 30.6, 37.3, 46.3, 53.9, 59.8, 66.2

(a) Assuming that the data came from a gamma distribution, obtain the m.l.e.'s $\hat{k}$ and $\hat{\alpha}$ of the shape and scale parameters.

(b) Let $Q_p(k, \alpha)$ represent the pth quantile of the two-parameter gamma distribution, given k and α. That is, $Q_p(k, \alpha)$ satisfies

$$I\left(k, \frac{Q_p}{\alpha}\right) = p,$$

where $I(k, x)$ is the incomplete gamma integral (1.3.16). Examine the adequacy of the gamma model by plotting the points

$$\left[Q_{(i-.5)/n}(\hat{k}, \hat{\alpha}), t_{(i)}\right] \qquad i = 1, \ldots, n,$$

where $t_{(i)}$ is the ith smallest observation in the sample of n.

(Sections 3.3, 4.2; Wilk et al. 1962)

4.8 Chhikara and Folks (1977) gave the data below on repair times (in hours) for 46 failures of an airborne communications receiver.

0.2	0.3	0.5	0.5	0.5	0.5	0.6	0.6	0.7	0.7
0.7	0.8	0.8	1.0	1.0	1.0	1.0	1.1	1.3	1.5
1.5	1.5	1.5	2.0	2.0	2.2	2.5	2.7	3.0	3.0
3.3	3.3	4.0	4.0	4.5	4.7	5.0	5.4	5.4	7.0
7.5	8.8	9.0	10.3	22.0	24.5				

(a) Treating the times as exact continuous observations, fit an inverse Gaussian distribution to the data.

(b) Informally assess the fit of the model. Use simulation or other means to consider whether the two largest repair times seem consistent with an inverse Gaussian model.

(Section 4.2.2)

4.9 *Parametric vs. nonparametric estimates.* Consider a study where all units still alive at time c are withdrawn, so that their lifetimes are right censored. Suppose it is wished to estimate $S(t_0)$, where $0 < t_0 < c$.

(a) Compare the precision of the nonparametric estimate $\hat{S}(t_0) = (n - r)/n$, where r is the number of units failing by time t_0, with that of the m.l.e. of $S(t)$ when the underlying distribution is exponential.

(b) Outline how the precision of the Kaplan–Meier estimate $\hat{S}(t)$ could be compared with the exponential m.l.e. of $S(t)$ under general Type 1 censoring (Section 2.2.1).

(Sections 3.2, 4.1; Miller 1983)

4.10 *Loss of information under grouping.* Consider settings where n individuals are inspected at specified times $a_1, \ldots, a_k$ so that only the numbers of deaths between successive inspection times are observed. That is, we observe d_j = number of lifetimes in $(a_{j-1}, a_j]$ for $j = 1, \ldots, k+1$, where $a_0 = 0 < a_1 < \cdots < a_k < a_{k+1} = \infty$.

(a) Assuming that lifetimes follow an exponential distribution with mean θ, obtain the information $I(\theta)$ and asymptotic variance of $\sqrt{n}(\hat{\theta} - \theta)$ under both the observation scheme just given and under observation of exact failure times, but with Type 1 censoring at time a_k.

(b) Assume further that $a_j = ja$ for $j = 1, \ldots, k$, and evaluate the expected information $\mathcal{I}(\theta)$ under both observation schemes. Make a numerical comparison of grouped and exact observation for the values $k = 1, \ldots, 5$ when a_k corresponds to the .50 quantile of the underlying exponential distribution. Repeat the comparison if a_k corresponds to the .90 quantile.

(Sections 4.1, 4.3)

4.11 *Loss of information under current status observation.* Consider $n = km$ individuals, where k and m are positive integers, and suppose that m individuals are observed at each of k times $C_1, \ldots, C_k$, it being determined in each case whether their lifetime exceeds the respective C_j or not. Suppose that C_j corresponds to the $(j - .5)/k$ quantile of the underlying lifetime distribution; we could of course only approximate this in practice, since the quantile values would be unknown.

(a) Assuming that the underlying lifetime distribution is exponential with mean θ, compare the information $I(\theta)$ and asymptotic variance of $\sqrt{n}(\hat{\theta} - \theta)$ for current-status data with each of $k = 1, 2, 4, 8$ and with exact observation of the lifetimes.

(b) Outline a numerical study to compare the precision with which the pth quantile, t_p, of an underlying log-logistic distribution (1.3.12) would be estimated, under exact observation of lifetimes and under current-status observation with each of $k = 1, 2, 4, 8$.

(Section 4.3)

4.12 The life table data in Table 4.6 are from a study involving 112 patients with plasma cell myeloma treated at the National Cancer Institute (Carbone et al. 1967).

(a) Use plots of empirical estimates of the survivor and hazard functions to suggest possible models.

(b) Fit Weibull and Gompertz models to these data, using maximum likelihood. Compare results using the likelihood function (4.3.4), which

Table 4.6. Survival Times for Patients with Plasma Cell Myeloma

Interval (Months)	Number at Risk at Start (n_l)	Number of Withdrawals (w_l)
[0, 5.5)	112	1
[5.5, 10.5)	93	1
[10.5, 15.5)	76	3
[15.5, 20.5)	55	0
[20.5, 25.5)	45	0
[25.5, 30.5)	34	1
[30.5, 40.5)	25	2
[40.5, 50.5)	10	3
[50.5, 60.5)	3	2
[60.5, ∞)	0	0

assumes withdrawals occurs at the ends of intervals, with those using the likelihood (3.7.4) in Problem 3.17. Informally assess the fit of each model.

(c) Compare the estimates in part (b) with the m.l.e.'s obtained by assuming that failure times or censoring times in an interval are all equal to the interval midpoint.

(d) Compare also the variance estimates based on the approaches in part (b) and the approximation of part (c). What do you conclude?

(Section 4.3)

4.13 Consider the interval-censored data on breast cosmesis given in Problem 3.13.

(a) Fit Weibull and log-logistic distributions to the data from each of the two treatment groups. Plot estimates of the s.f.'s $S(t)$ under the two models, along with the nonparametric estimates obtained from Problem 3.13.

(b) Fit models with (1) a cubic hazard function with $h(0) = 0$, and (2) a cubic spline hazard function with a single knot at 18 months for each treatment group. Compare the estimates of $S(t)$ with those of part (a).

(Sections 4.2, 4.3; Lindsey and Ryan 1998; Lindsey 1998)

4.14 *Mixtures with known components.* Suppose that Y has p.d.f.

$$f(y; p) = pf_1(y) + (1 - p)f_2(y),$$

where $0 < p < 1$ and f_1 and f_2 are completely specified p.d.f.'s.

(a) Show that the expected information from a complete sample of size n from the distribution is

$$I(p) = \frac{n}{p(1-p)}\left(1 - \int_{-\infty}^{\infty} \frac{f_1(y)f_2(y)}{f(y; p)}\,dy\right).$$

Note that this reduces to the binomial information $n[p(1-p)]^{-1}$ when f_1 and f_2 do not overlap.

(b) If

$$f_1(y) = \frac{1}{(2\pi)^{1/2}}e^{-(y-\mu)^2/2} \quad \text{and} \quad f_2(y) = \frac{1}{(2\pi)^{1/2}}e^{-y^2/2}$$

evaluate $I(.5)$ numerically for several values of μ. Comment on the precision with which one can estimate p as a function of μ.

(Section 4.4; Hill 1963)

4.15 Consider the mixture model (4.4.1) with long-term survivors in the case where the survivor function is

$$S(t) = 1 - p + pe^{-t/\theta} \qquad t \geq 0.$$

(a) Data from a study on 100 subjects who were followed for half a year gave 22 failure times t_i with $\sum t_i = 4.41$ years and 78 censoring times, each equal to 0.5 year. Plot contours of the log-likelihood function

$$\ell(\theta, p) = 22\log(p/\theta) - 4.41/\theta + 78\log(1 - p + pe^{-.5/\theta})$$

and consider interval estimation of p.

(b) After 3.0 years of follow-up there were 49 failures observed, with $\sum t_i = 31.10$, and 51 censoring times, each equal to 3.0. Plot contours of the log-likelihood funtion $\ell(\theta, p)$ in this case, and consider estimation of p.

(Section 4.4)

4.16 Using general optimization software, fit a mixture of two log-logistic distributions to the data in Example 4.4.2 using maximum likelihood. Compare the estimate of $S(t)$ under this model with the estimates shown in Figure 4.10.

(Section 4.4)

4.17 Consider the data in Example 4.5.1, to which a three-parameter Weibull model (4.5.1) was fitted. Obtain m.l.e.'s of α, β, and γ using the alternative likelihood function discussed in Section 4.5.1, in which $f(t_{(l)}; \alpha, \beta, \gamma)$ is replaced with $F(t_{(l)}+\Delta) - F(t_{(l)}-\Delta)$, where $F(t)$ is the c.d.f. corresponding to (4.5.1) and Δ is a small value; use $\Delta = .5$. Compare the estimates with those in the example, and also compare the profile log-likelihood function $\ell(\hat{\alpha}(\gamma), \hat{\beta}(\gamma), \gamma)$ with that in Table 4.4.

(Section 4.5.2)

4.18 Prove the assertion made immediately following (4.5.12). Note that as $r \to \infty$, an $F_{(2,r)}$ random variables converges in distribution to $.5\chi^2_{(2)}$.

(Section 4.5.3)

4.19 The data below represent failure times, in minutes, for two types of electrical insulation in an experiment in which the insulation was subjected to a continuously increasing voltage stress.

Type A	219.3	79.4	86.0	150.2	21.7	18.5
	121.9	40.5	147.1	35.1	42.3	48.7
Type B	21.8	70.7	24.4	138.6	151.9	75.3
	12.3	95.5	98.1	43.2	28.6	46.9

Examine graphically whether the two sets of data might be considered to be random samples from different two-parameter exponential distributions (see Example 4.5.2). If this appears reasonable, compare the two distributions and, in particular, test that they have the same threshold parameter value.

(Section 4.5.3)

4.20 *The three-parameter log-normal distribution.* For the three-parameter log-normal distribution, $\log(T - \gamma)$ is normally distributed with mean μ and variance σ^2, where $T \geq \gamma$ and $\gamma \geq 0$ is a threshold parameter.

(a) If γ is known, determine the m.l.e.'s $\hat{\mu}(\gamma)$ and $\hat{\sigma}(\gamma)$ of μ and σ from a complete sample of size n. Thus obtain the profile likelihood function $L_{\max}(\gamma)$. Show that

$$\lim_{\gamma \to t_{(1)}^-} L_{\max}(\gamma) = \infty,$$

Consider the ramifications of this for maximum likelihood estimation.

(b) Consider the rat-tumor data of Example 4.5.1 as having arisen from a three-parameter log-normal distribution. Compute and examine the profile likelihood function $L_{\max}(\gamma)$. Obtain the value $\hat{\gamma}$ that gives a local maximum of $L_{\max}(\gamma)$. Treating this as the m.l.e., estimate all three parameters. Determine a range of plausible values for γ and examine the effect of γ on estimates of distribution quantiles.

(c) Compare $L_{\max}(\gamma)$ and $\hat{\gamma}$ in part (b) with those obtained by using the modified likelihood function described in Problem 4.17.

(Section 4.5; Griffiths 1980)

4.21 Disease remission times T for patients undergoing a certain type of treatment are well described by an exponential distribution. A set of 50 patients gave $\sum t_i = 61.5$ years and the m.l.e. $\hat{\theta} = 1.23$ years for the mean duration of remission. In a future set of 100 patients let V denote the number whose remission time exceeds 1 year. Obtain a lower .95 prediction limit $A(\hat{\theta})$ for V using the plug-in method and then determine by simulation the unconditional coverage probability $Pr[V > A(\hat{\theta})]$ for this procedure.

(Section 4.6)

4.22 *Bayesian prediction*

(a) Consider an exponential lifetime distribution with mean θ, and censored lifetime data, giving the likelihood function

$$L(\theta) = \frac{1}{\theta^r} \exp(-T/\theta),$$

where $r = \sum \delta_i$ and $T = \sum t_i$. Assuming that $\lambda = \theta^{-1}$ has a two-parameter gamma prior with density function

$$\pi(\lambda; a, b) = \frac{1}{\Gamma(b)} a^b \lambda^{b-1} e^{-a\lambda} \quad (\lambda > 0),$$

obtain the predictive distribution for a future lifetime Y, using (4.6.10).

(b) Obtain a lower .95 Bayesian prediction interval for Y based on uncensored data with $r = n = 20$ and $T = 2000$ and a gamma prior with $b = a$ approaching 0. Compare this prediction limit with a frequentist limit based on the method of Example 4.6.1. The limiting improper prior $\pi(\lambda) = \lambda^{-1}$ is sometimes taken as a noninformative prior for λ. Note that for the preceding general gamma prior, $E(\lambda) = ba^{-1}$ and $\mathrm{Var}(\lambda) = ba^{-2}$.

(Section 4.6; Aitchison and Dunsmore 1975; Martz and Waller 1982)

CHAPTER 5

Inference Procedures for Log-Location-Scale Distributions

5.1 INFERENCE FOR LOCATION-SCALE DISTRIBUTIONS

Location-scale distributions, introduced in Section 1.3.6, have survivor functions of the form

$$S(y; u, b) = S_0\left(\frac{y-u}{b}\right) \qquad -\infty < y < \infty, \tag{5.1.1}$$

where $u(-\infty < u < \infty)$ is a location parameter, $b > 0$ is a scale parameter, and $S_0(\,)$ is a fully specified survivor function defined on $(-\infty, \infty)$. If T is a lifetime variable and $Y = \log T$ has distribution (5.1.1), then we say that T has a log-location-scale distribution. The survivor function for T may be written as

$$\begin{aligned} S^*(t; \alpha, \beta) &= S_0\left(\frac{\log t - u}{b}\right) \\ &= S_0^*[(t/\alpha)^\beta], \end{aligned} \tag{5.1.2}$$

where $\alpha = \exp(u)$, $\beta = b^{-1}$ and for $0 < w < \infty$, $S_0^*(w) = S_0(\log w)$. The Weibull, log-logistic, and log-normal distributions are all of this form; the corresponding location-scale parameter distributions of Y are the extreme value, logistic, and normal, respectively.

Log-location-scale distributions are the most widely used parametric lifetime models, and regression models in which u (and sometimes b) in (5.1.1) are functions of covariates are of fundamental importance in both parametric and semiparametric frameworks. This chapter presents inference procedures for location-scale and log-location-scale models without covariates. Following a general development of methodology Weibull, log-logistic, log-normal, and other models are discussed in more detail. We focus on settings where exact lifetimes or censoring times are

observed, with no truncation or selection effects present. For problems involving interval censoring or truncation, reference may be made to Sections 4.3 and 4.4. Similarly, we do not consider threshold parameters, which were treated in Section 4.5.

5.1.1 Likelihood-Based Methods

This section summarizes general results for location-scale models; illustrations are deferred to later sections that deal with specific distributions. We will discuss maximum likelihood and associated methods in terms of the parameters (u, b). The log-likelihood function for u and b is usually closer to quadratic than that for (α, β), and large-sample normal approximations for $\hat{u}$, $\hat{b}$ tend to be more accurate than those for $\hat{\alpha}$, $\hat{\beta}$. A still better choice of parameters for obtaining the maximum likelihood estimates (m.l.e.'s) and implementing normal approximations is $(u, \log b)$; survival software often uses this. We will retain the more natural (u, b) parameterization for general discussion, bearing in mind that $(u, \log b)$ may be used for computation.

For a censored random sample of lifetimes (t_i, δ_i), $i = 1, \ldots, n$ the likelihood function under the model (5.1.1) for $\log T$ is

$$L(u, b) = \prod_{i=1}^{n} \left[\frac{1}{b} f_0\left(\frac{y_i - u}{b}\right)\right]^{\delta_i} S_0\left(\frac{y_i - u}{b}\right)^{1-\delta_i}, \tag{5.1.3}$$

where $y_i = \log t_i$ and $f_0(z) = -S_0'(z)$ is the probability density function (p.d.f.) corresponding to $S_0(z)$. Letting $z_i = (y_i - u)/b$ and $r = \sum \delta_i$, we find the log-likelihood function for u and b as

$$\ell(u, b) = -r \log b + \sum_{i=1}^{n} [\delta_i \log f_0(z_i) + (1 - \delta_i) \log S_0(z_i)]. \tag{5.1.4}$$

Since $\partial z_i/\partial u = -b^{-1}$ and $\partial z_i/\partial b = -z_i b^{-1}$, the score functions are

$$\frac{\partial \ell}{\partial u} = -\frac{1}{b} \sum_{i=1}^{n} \left[\delta_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1 - \delta_i) \frac{\partial \log S_0(z_i)}{\partial z_i}\right] \tag{5.1.5}$$

$$\frac{\partial \ell}{\partial b} = -\frac{r}{b} - \frac{1}{b} \sum_{i=1}^{n} \left[\delta_i z_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1 - \delta_i) z_i \frac{\partial \log S_0(z_i)}{\partial z_i}\right], \tag{5.1.6}$$

and the second derivatives of $\ell(u, b)$ are

$$\frac{\partial^2 \ell}{\partial u^2} = \frac{1}{b^2} \sum_{i=1}^{n} \left[\delta_i \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1 - \delta_i) \frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right] \tag{5.1.7}$$

$$\frac{\partial^2 \ell}{\partial b^2} = \frac{r}{b^2} + \frac{2}{b^2} \sum_{i=1}^{n} \left[\delta_i z_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1 - \delta_i) z_i \frac{\partial \log S_0(z_i)}{\partial z_i}\right] + \frac{1}{b^2} \sum_{i=1}^{n} \left[\delta_i z_i^2 \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1 - \delta_i) z_i^2 \frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right] \tag{5.1.8}$$

$$\frac{\partial^2 \ell}{\partial u \partial b} = \frac{1}{b^2} \sum_{i=1}^{n} \left[\delta_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1 - \delta_i) \frac{\partial \log S_0(z_i)}{\partial z_i}\right] + \frac{1}{b^2} \sum_{i=1}^{n} \left[\delta_i z_i \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1 - \delta_i) z_i \frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right]. \tag{5.1.9}$$

The observed information matrix is

$$I(u, b) = \begin{pmatrix} -\partial^2 \ell/\partial u^2 & -\partial^2 \ell/\partial u \partial b \\ -\partial^2 \ell/\partial b \partial u & -\partial^2 \ell/\partial b^2 \end{pmatrix},$$

and the usual large-sample normal approximation to the joint distribution of the estimators $\hat{u}$ and $\hat{b}$ is, as described in Appendix C, to treat $(\hat{u}, \hat{b})$ as approximately bivariate normal with mean (u, b) and covariance matrix $I(\hat{u}, \hat{b})^{-1}$. It may be noted that since $\hat{u}, \hat{b}$ satisfy the equations $\partial \ell/\partial u = 0$, $\partial \ell/\partial b = 0$, the first term in (5.1.9) equals 0 at $(\hat{u}, \hat{b})$ and the second term in (5.1.8) equals $-2r/\hat{b}^2$.

The Fisher information matrix $\mathcal{I}(u, b) = E\{I(u, b)\}$ is not available unless the censoring process is fully specified. It can be obtained in the case of Type 1 or Type 2 censoring, described in Section 2.2.1. For Type 1 censoring, let C_i be the prespecified censoring time for individual i, and let $R_i = (\log C_i - u)/b$. Then $Pr(\delta_i = 1) = F_0(R_i) = 1 - S_0(R_i)$, and given that $\delta_i = 1$, the standardized variable $Z_i = (Y_i - u)/b$ has the truncated p.d.f.

$$g(z) = \frac{f_0(z)}{F_0(R_i)} \qquad -\infty < z \le R_i. \tag{5.1.10}$$

Once again, there is some simplification from the fact that $E(\partial \ell/\partial u) = 0$ and $E(\partial \ell/\partial b) = 0$, which implies that the first term in (5.1.9) has expectation zero and the second term in (5.1.8) has expectation $-2r/b^2$. The remaining terms can be evaluated using (5.1.10) and $Pr(\delta_i = 1) = F_0(R_i)$. Problem 5.4 considers the case of an extreme value distribution.

The primary use of $\mathcal{I}(u, b)$ is for design purposes; the matrix $\mathcal{I}(u, b)^{-1}$ at specified values of u and b can be used to estimate the precision of estimators based on a given sample size and censoring pattern. In place of direct calculation, we can alternatively estimate $\mathcal{I}(u, b)$ by simulation. Design-related issues are discussed in Section 5.6.

5.1.1.1 Wald-Type Confidence Procedures

Approximate confidence intervals or tests for u, b, and other parameters can be obtained by treating $(\hat{u}, \hat{b})$ as bivariate normal with mean (u, b) and covariate matrix

$\hat{V} = I(\hat{u}, \hat{b})^{-1}$. In particular, let $se(\hat{\theta}) = \widehat{\text{Asvar}}(\hat{\theta})^{1/2}$ denote the standard error of an estimate $\hat{\theta}$, so that

$$se(\hat{u}) = \hat{V}_{11}^{1/2}, \qquad se(\hat{b}) = \hat{V}_{22}^{1/2}.$$

Confidence intervals for u and b can be obtained from the approximate pivotal quantities

$$Z_1 = \frac{\hat{u} - u}{se(\hat{u})}, \qquad Z_2 = \frac{\hat{b} - b}{se(\hat{b})}, \tag{5.1.11}$$

both of which are approximately $N(0, 1)$ for large samples. For example, the fact that $Pr(-1.96 \le Z_1 \le 1.96) \doteq .95$ leads to the approximate .95 confidence interval $\hat{u} - 1.96se(\hat{u}) \le u \le \hat{u} + 1.96se(\hat{u})$. Confidence intervals for b can alternatively be based on the approximate $N(0, 1)$ pivotal quantity

$$Z_2' = \frac{\log \hat{b} - \log b}{se(\log \hat{b})},$$

where $se(\log \hat{b}) = se(\hat{b})/\hat{b}$ by (B4) of Appendix B. Intervals based on Z_2' are slightly more accurate in small samples than ones based on Z_2.

The pth quantile for Y is $y_p = u + w_p b$, where $w_p = F_0^{-1}(p)$ is the pth quantile of $F_0(z) = 1 - S_0(z)$ in (5.1.1). The standard error of the m.l.e. $\hat{y}_p = \hat{u} + w_p \hat{b}$ is

$$se(\hat{y}_p) = (\hat{V}_{11} + 2w_p \hat{V}_{12} + w_p^2 \hat{V}_{22})^{1/2}, \tag{5.1.12}$$

and

$$Z_p = \frac{\hat{y}_p - y_p}{se(\hat{y}_p)} \tag{5.1.13}$$

is an approximate standard normal pivotal quantity that can be used to get confidence intervals for y_p.

Confidence intervals for survival probabilities $S(y_0)$ or for $F(y_0) = 1 - S(y_0)$ can be obtained for a specified y_0 by using the approximate $N(0, 1)$ pivotal quantity

$$Z_0 = \frac{(\hat{\psi} - \psi)}{se(\hat{\psi})}, \tag{5.1.14}$$

where $\psi = S_0^{-1}(S(y_0)) = (y_0 - u)/b$. The asymptotic variance formula (B2) of Appendix B gives $se(\hat{\psi}) = (\mathbf{a}'\hat{V}\mathbf{a})^{1/2}$, where $\mathbf{a}' = (-\hat{b}^{-1}, -(y_0 - \hat{u})\hat{b}^{-2})$. An alternative procedure is to use the relationship between quantiles and the survivor function along with the pivotal quantity (5.1.13). Since $S(y_0) \ge 1 - p$ if and only if $y_p \ge y_0$, we can obtain a lower q confidence limit for $S(y_0)$ by finding the value

of p such that y_0 is a lower q confidence limit for y_{1-p}. This requires that we find $\psi = S_0^{-1}(S(y_0))$ such that

$$\frac{\hat{b}(\hat{\psi} - \psi)}{[\hat{V}_{11} + 2\psi^2\hat{V}_{12} + \psi^2\hat{V}_{22}]^{1/2}} = -N_q, \tag{5.1.15}$$

where N_q is the qth quantile for $N(0, 1)$. This usually gives a result close to that based on (5.1.14), which requires that we solve

$$\frac{\hat{b}(\hat{\psi} - \psi)}{[\hat{V}_{11} + 2\hat{\psi}^2\hat{V}_{12} + \hat{\psi}^2\hat{V}_{22}]^{1/2}} = -N_q \tag{5.1.16}$$

to get a lower q confidence limit for ψ, and thus $S(y_0) = S_0(\psi)$. Note that (5.1.16) is slightly easier to use, since (5.1.15) involves solving a quadratic equation to get ψ.

The normal approximations upon which the preceding procedures are based can be inaccurate for small samples; Jeng and Meeker (2000) and Doganaksoy and Schmee (2000) provide discussions and references for extreme value and normal distributions. An alternative to the normal approximations is to use bootstrap simulation procedures to estimate the distributions of pivotals Z_1, Z_2, and Z_p given earlier (Appendix D.2; Efron and Tibshirani 1993, Ch. 12; Davison and Hinkley 1997, Ch. 5). Analytical adjustments designed to improve accuracy can also be made. A cautionary note is that all of these procedures can perform poorly in small samples with heavy censoring. Of course, small samples typically contain only very limited information about the lifetime distribution and this limits the practical impact of such data.

Likelihood ratio procedures are slightly more trouble to implement than Wald procedures, but often perform better in small or medium-size samples; we consider them next.

5.1.1.2 *Likelihood Ratio Procedures*

To test hypotheses $H: u = u_0$ or $H: b = b_0$, respectively, one can use the likelihood ratio statistics

$$\Lambda_1(u_0) = 2\ell(\hat{u}, \hat{b}) - 2\ell(u_0, \hat{b}(u_0))$$

and

$$\Lambda_2(b_0) = 2\ell(\hat{u}, \hat{b}) - 2\ell(\hat{u}(b_0), b_0),$$

where $\hat{u}(b_0)$ maximizes $\ell(u, b_0)$ and $\hat{b}(u_0)$ maximizes $\ell(u_0, b)$. Large values of $\Lambda_1(u_0)$ and $\Lambda_2(b_0)$ provide evidence against the hypotheses, and approximate p-values may be calculated by using the fact that under the respective hypotheses, the statistics $\Lambda_1(u_0)$ and $\Lambda_2(b_0)$ are asymptotically distributed as $\chi^2_{(1)}$. Approximate two-sided q confidence intervals for u or b are obtained as the sets of values u_0 or b_0 for which $\Lambda_1(u_0)$ or $\Lambda_2(b_0)$ are $\le \chi^2_{(1),q}$. One-sided confidence intervals are

obtained as the sets of values satisfying inequalities for the signed square roots of $\Lambda_1(u_0)$ and $\Lambda_2(b_0)$ (see (C24) of Appendix C). This gives, for example, lower q confidence limits for u consisting of values

$$\{u_0 : I(u_0 \le \hat{u})\Lambda_1(u_0) \le \chi^2_{(1),2q-1}\}, \tag{5.1.17}$$

and upper q confidence limits consisting of values

$$\{u_0 : I(u_0 \ge \hat{u})\Lambda_1(u_0) \le \chi^2_{(1),2q-1}\}. \tag{5.1.18}$$

Likelihood ratio inferences about quantiles or survival probabilities are also straightforward. For the pth quantile we consider hypotheses $H: y_p = y_{p0}$; the likelihood ratio statistic for testing H is

$$\Lambda(y_{p0}) = 2\ell(\hat{u}, \hat{b}) - 2\ell(\tilde{u}, \tilde{b}), \tag{5.1.19}$$

where $(\tilde{u}, \tilde{b})$ maximizes $\ell(u, b)$ under H. Since $y_p = u + w_p b$, to find $\tilde{u}, \tilde{b}$ we can just maximize

$$\ell_1(b) = \ell(y_{p0} - w_p b, b) \tag{5.1.20}$$

to get $\tilde{b}$, and then $\tilde{u} = y_{p0} - w_p \tilde{b}$. Confidence intervals for y_p require the determination of values y_{p0} such that $\Lambda(y_{p0}) \le \chi^2_{(1),q}$. This involves the same degree of computation as likelihood ratio confidence intervals for u or b and, indeed, u is just the quantile for which $w_p = 0$.

Confidence intervals for $S(y_0)$ are also readily found. The likelihood ratio statistic for testing $H: S(y_0) = s_0$ is $\Lambda(s_0)$, of the same form as in (5.1.19), but with $(\tilde{u}, \tilde{b})$ maximizing $\ell(u, b)$ under the restriction H. The condition $S(y_0) = s_0$ implies by (5.1.1) that $u + S_0^{-1}(s_0)b = y_0$, so $(\tilde{u}, \tilde{b})$ can be obtained by maximizing

$$\ell_2(b) = \ell(y_0 - S_0^{-1}(s_0)b, b) \tag{5.1.21}$$

to get $\tilde{b}$, and then $\tilde{u} = y_0 - S_0^{-1}(s_0)\tilde{b}$. This is precisely the same as the maximization of (5.1.20) in the estimation of quantiles.

The likelihood ratio procedures tend to give quite accurate confidence intervals when the number of failures is about 20 or bigger, with two-sided intervals giving closer to nominal coverage than one-sided. The accuracy can be made very good even for fairly small samples by approximating the distribution of the signed square-root statistic by bootstrapping, or by using an analytic correction for mean and variance. Given that any parametric model is only an approximation to reality and that a moderate number of observations is needed both to assess models effectively and to provide estimates with a reasonable degree of precision, even the unadjusted likelihood ratio procedures are satisfactory for most practical purposes.

5.1.2 Exact Procedures Under Type 2 Censoring

Although the methods of Section 5.1.1 can be used quite generally, when lifetime data are complete or Type 2 censored, there exist exact pivotal quantities upon which inferences about parameters in models (5.1.1) can be based. This topic is discussed in Appendix E; we employ the results here.

Let $\hat{u}$ and $\hat{b}$ be the m.l.e.'s of u and b, or any other pair of equivariant estimators (see Appendix E). The following theorem follows immediately from Theorems E1 and E2 of Appendix E:

THEOREM 5.1.1. Let $\hat{u}$ and $\hat{b}$ be the m.l.e.'s of u and b, based on a Type 2 censored sample from (5.1.1), which for convenience is labeled $y_1 \le y_2 \le \cdots y_r$ with $r \le n$. Then

i. $Z_1 = (\hat{u} - u)/\hat{b}$, $Z_2 = \hat{b}/b$, and $Z_p = (\hat{u} - y_p)/\hat{b}$ are pivotal quantities.

ii. The quantities $a_i = (y_i - \hat{u})/\hat{b}$ form a set of ancillary statistics (i.e., statistics whose distribution does not depend on u or b), of which only $r - 2$ are functionally independent.

The pivotals Z_1, Z_2, and Z_p can be used to construct confidence intervals or tests for u, b, y_p, and $S(y_0)$; note that these are not the same pivotals as in (5.1.11) and (5.1.13). A practical difficulty is that the distributions of the pivotals, whose general form is given in Theorem E3 of Appendix E, is complicated. However, for any given values of r and n it is easy to approximate the distributions to a high degree of accuracy by simulation, as follows. Since the distributions of Z_1, Z_2, and Z_p do not depend on the values of u and b in the underlying model (5.1.1), we can set $u = 0$ and $b = 1$, generate a pseudorandom sample $y_1 \le y_2 \le \cdots \le y_r$, then obtain $\hat{u}$, $\hat{b}$, and thus values $z_1^* = \hat{u}/\hat{b}$, $z_2^* = \hat{b}$, $z_p^* = (\hat{u} - w_p)/\hat{b}$, where $w_p = F_0^{-1}(p)$ is the pth quantile of $F_0(z)$. These are random values from the distributions of Z_1, Z_2, and Z_p, and by repeating the process a large number of times we can obtain a precise estimate of their distributions.

Some tables of percentage points (quantiles) for Z_1, Z_2, and Z_p have been obtained by simulation in the case of the extreme value and normal distributions. Similar tables for pivotal quantities based on linear estimators, rather than m.l.e.'s of u and b, have also been given. Information on this is provided in Sections 5.2.2, 5.3.1, and the Bibliographic Notes for this chapter.

There is a second way to use the pivotal quantities Z_1, Z_2, and Z_p for inference, and that is through their conditional distributions, given the set of ancillary statistics $a_1, \ldots, a_r$ defined in Theorem 5.1.1. Except for the case of uncensored data from the normal distribution or the general case $r = n = 2$, no pair of estimators $\hat{u}$ and $\hat{b}$ can be sufficient for u and b in a location-scale model (5.1.1). It has been convincingly argued that inferences about parameters in (5.1.1) should be made conditional on the observed value of $\mathbf{a} = (a_1, \ldots, a_r)$; this is discussed in Appendix E. By these arguments, confidence intervals should be based on the conditional distributions of Z_1, Z_2, and Z_p given $\mathbf{a}$. To get a q confidence interval for u, for example, we need

to obtain values ℓ_1 and ℓ_2 such that

$$Pr(\ell_1 \le Z_1 \le \ell_2 | \mathbf{a}) = q. \tag{5.1.22}$$

This may seem more complicated than considering the unconditional (marginal) distribution of Z_1, but it turns out to be easier to calculate probabilities like (5.1.21) numerically than unconditional probabilities such as $Pr(\ell_1 \le Z_1 \le \ell_2)$. The conditional distributions of Z_1 and Z_2 given $\mathbf{a}$ are given in Theorem E3 of Appendix E, and the calculation of probabilities for Z_1, Z_2, and Z_p is discussed in Section 5.2.2 for the case of the extreme value distribution.

The m.l.e.'s $\hat{u}$ and $\hat{b}$ are asymptotically sufficient, and in even moderate-sized samples there is usually little difference between conditional and unconditional confidence intervals based on Z_1, Z_2, and Z_p. In addition, although Z_1, Z_2, and Z_p are strictly pivotal only for the case of Type 2 censoring, their distributions often do not depend much on u and b for other censoring schemes, assuming the expected number of observed failures is moderately large. An alternative to the inference procedures of Section 5.1.1 is to use Z_1, Z_2, and Z_p of Theorem 5.1.1, approximating their distributions via bootstrap simulation.

5.2 WEIBULL AND EXTREME-VALUE DISTRIBUTIONS

This section considers inference procedures for the Weibull distribution (1.3.5) with p.d.f. written in the form

$$f(t; \alpha, \beta) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left[-(t/\alpha)^{\beta}\right], \qquad t \ge 0 \tag{5.2.1}$$

where $\alpha > 0$ and $\beta > 0$ are the scale and shape parameters. Instead of working directly with (5.2.1), we often find it convenient to work with the equivalent extreme value distribution for $Y = \log T$, which is of location-scale form (5.1.1) with p.d.f.

$$f(y; u, b) = \frac{1}{b} e^{(y-u)/b} \exp\left[-e^{(y-u)/b}\right], \qquad -\infty < y < \infty, \tag{5.2.2}$$

where $u = \log \alpha$ and $b = \beta^{-1}$. The methods of Section 5.1 can be applied to the extreme value distribution, and inferences about Weibull parameters obtained by transformation from the extreme value parameters.

5.2.1 Likelihood-Based Inference Procedures

Let (t_i, δ_i), $i = 1, \ldots, n$ be a censored random sample of lifetimes from (5.2.1) and (y_i, δ_i), $i = 1, \ldots, n$, with $y_i = \log t_i$, be the corresponding censored sample from (5.2.2). To apply the general expressions (5.1.4)–(5.1.9) of Section 5.1.1 to the extreme value distribution, note that for (5.2.2) expressed in the general form (5.1.1)

$$S_0(z) = \exp(-e^z), \qquad f_0(z) = -S_0'(z) = e^z \exp(-e^z). \tag{5.2.3}$$

This gives the log-likelihood from (5.1.4) as

$$\ell(u, b) = -r \log b + \sum_{i=1}^{n} (\delta_i z_i - e^{z_i}), \tag{5.2.4}$$

where $z_i = (y_i - u)/b$ and $r = \sum \delta_i$. It follows from (5.2.3) that

$$\frac{\partial \log f_0(z)}{\partial z} = 1 - e^z, \qquad \frac{\partial^2 \log f_0(z)}{\partial z^2} = -e^z,$$

$$\frac{\partial \log S_0(z)}{\partial z} = -e^z, \qquad \frac{\partial^2 \log S_0(z)}{\partial z^2} = -e^z,$$

giving straightforward expressions for the first and second derivatives of $\ell(u, b)$ from (5.1.5)–(5.1.9). The log-likelihood function $\ell(u, b)$ is easily maximized to give $\hat{u}$, $\hat{b}$; many software packages handle this and the inference procedures in the next paragraph.

Using the fact that (5.1.5) and (5.1.6) equal 0 at $(\hat{u}, \hat{b})$, we find the observed information matrix at $(\hat{u}, \hat{b})$ for the extreme value model to be

$$I(\hat{u}, \hat{b}) = \frac{1}{\hat{b}^2} \begin{pmatrix} r & \sum_{i=1}^{n} \hat{z}_i e^{\hat{z}_i} \\ \sum_{i=1}^{n} \hat{z}_i e^{\hat{z}_i} & r + \sum_{i=1}^{n} \hat{z}_i^2 e^{\hat{z}_i} \end{pmatrix}. \tag{5.2.5}$$

Approximate confidence intervals or tests for u, b, y_p, or other quantities can be found by treating $(\hat{u}, \hat{b})$ as bivariate normal with mean vector (u, b) and covariance matrix $I(\hat{u}, \hat{b})^{-1}$, and the corresponding approximate pivotal quantities in (5.1.11) and (5.1.13) as $N(0, 1)$, as described in Section 5.1.1 and Appendix C. Inferences for the Weibull parameters $\alpha = \log u$, $\beta = b^{-1}$, and $t_p = \log y_p$ are easily obtained by transformation.

Likelihood ratio procedures are also easy to apply, though less commonly available in software packages. As discussed in Section 5.1.1, their accuracy is generally superior to that of the Wald-type procedures. Inference for b is especially straightforward, since under the hypothesis $H: b = b_0$, the likelihood equation $\partial \ell(u, b_0)/\partial u = 0$ has the closed-form solution

$$\hat{u}(b_0) = b_0 \log\left(\frac{1}{r} \sum_{i=1}^{n} e^{y_i/b_0}\right).$$

The likelihood ratio statistic

$$\Lambda(b_0) = 2\ell(\hat{u}, \hat{b}) - 2\ell(\hat{u}(b_0), b_0) \tag{5.2.6}$$

and confidence intervals $\{b_0 : \Lambda(b_0) \le \chi^2_{(1),q}\}$ are therefore easy to obtain.

Let us also consider confidence intervals or tests for quantiles $y_p = u + w_p b$, where from (5.2.3) we have that $w_p = S_0^{-1}(1-p) = \log[-\log(1-p)]$. The location parameter u is just $y_{.632}$, and so is a special case. The likelihood ratio statistic $\Lambda(y_{p0})$ for testing $H: y_p = y_{p0}$ is given by (5.1.19), and $(\tilde{u}, \tilde{b})$ in (5.1.19) can be obtained by maximizing (5.1.20), which here becomes

$$\ell_1(b) = -r \log b + \sum_{i=1}^{n} \left[\delta_i \left(\frac{y_i - y_{p0} + w_p b}{b} \right) - \exp\left(\frac{y_i - y_{p0} + w_p b}{b} \right) \right]. \tag{5.2.7}$$

This is easily maximized with standard software to give $\tilde{b}$ and then $\tilde{u} = y_{p0} - w_p \tilde{b}$, and

$$\Lambda(y_{p0}) = 2\ell(\hat{u}, \hat{b}) - 2\ell(\tilde{u}, \tilde{b}). \tag{5.2.8}$$

Two-sided confidence intervals $\{y_{p0} : \Lambda(y_{p0}) \le \chi^2_{(1),q}\}$ or analogous one-sided intervals are found by iterating this process over a range of values y_{p0}. A close approximation to $\Lambda(y_{p0})$ can usually be obtained by calculating it at a small number of values y_{p0} on either side of $\hat{y}_p = \hat{u} + w_p \hat{b}$ and then passing an interpolating spline through the points $(y_{p0}, \Lambda(y_{p0}))$.

To obtain confidence intervals for survival probabilities $S(y_0) = \exp[-\exp((y_0 - u)/b)]$ we consider the likelihood ratio procedure described in Section 5.1.1. This is based on the statistic

$$\Lambda(s_0) = 2\ell(\hat{u}, \hat{b}) - 2\ell(\tilde{u}, \tilde{b}), \tag{5.2.9}$$

where $(\tilde{u}, \tilde{b})$ maximizes $\ell(u, b)$ subject to $H: S(y_0) = s_0$. By (5.1.20) and (5.2.4), $\tilde{b}$ is found by maximizing (5.2.7), with y_0 replacing y_{p0} and $\log(-\log s_0)$ replacing w_p, and then $\tilde{u} = y_0 - \log(-\log s_0)\tilde{b}$.

Example 5.2.1. Leukemia remission time data were given in Example 1.1.7. The observations were remission or censoring times, in weeks, for two groups of patients, one given a treatment (drug 6-MP) and the other a placebo. Each group had 21 individuals; 12 times were censored in the 6-MP group and 0 in the Placebo group. Kaplan–Meier estimates of the survivor functions were given in Example 3.2.1, and probability plots suggest that Weibull models are consistent with the data. We consider here the results of fitting Weibull distributions to the two samples.

The results of several confidence interval procedures for estimating (for each treatment group) the Weibull shape parameters β, the median remission times $t_{.50}$, and $S(10)$, the probabilities remission lasts longer than 10 weeks, are shown in Table 5.1. The maximum likelihood calculations and Wald-type procedures were carried out using the extreme value parameters (u, b). That is, confidence intervals for u, b, and $y_{.50}$ were obtained using the approximate pivotals (5.1.11) and (5.1.13), and

Table 5.1. Approximate .95 Confidence Intervals by Several Methods

Parameter	Method	Drug 6-MP $\hat{\alpha} = 33.77, \hat{\beta} = 1.35$	Placebo $\hat{\alpha} = 9.48, \hat{\beta} = 1.37$
β	LR	(.72, 2.21)	(.98, 1.88)
	Wald ($\hat{b}$)	(.87, 3.00)	(1.02, 2.07)
	Wald ($\log \hat{b}$)	(.79, 2.33)	(.98, 1.93)
	Bootstrap	(.95, 2.04)	(1.00, 1.78)
$t_{.50}$	LR	(16.2, 51.6)	(4.75, 10.3)
	Wald ($\hat{y}_{.50}$)	(15.8, 42.1)	(5.02, 10.5)
	Bootstrap	(15.6, 46.2)	(4.72, 9.99)
$S(10)$	LR	(.637, .931)	(.197, .513)
	Wald ($\hat{\psi}$)	(.630, .923)	(.208, .524)
	Bootstrap	(.689, .915)	(.164, .511)

then transformed to give the confidence intervals for α, β, and $t_{.50}$ via $\alpha = \exp(u)$, $\beta = b^{-1}$, $t_{.50} = \exp(y_{.50})$. Confidence intervals obtained both by treating $\hat{b}$ as normal and $\log \hat{b}$ as normal are shown for comparison. The asymptotic covariance matrices $I(\hat{u}, \hat{b})^{-1}$ for the two groups are as follows:

$$\text{6–MP: } \begin{pmatrix} .07473 & .02442 \\ .02442 & .04229 \end{pmatrix} \qquad \text{Placebo: } \begin{pmatrix} .02816 & .00671 \\ .00671 & .01604 \end{pmatrix}.$$

Wald-type confidence intervals for $S(10) = \exp[-(10/\alpha)^\beta]$ were obtained by considering the parameter $\psi = \log(-\log S(10)) = (\log 10 - u)/b$. Confidence intervals for ψ were based on the approximate $N(0, 1)$ pivotal quantity $(\hat{\psi} - \psi)/se(\hat{\psi})$ of (5.1.14); these were then transformed to confidence intervals for $S(10) = \exp(-e^{\psi})$. Numerous software packages implement these procedures for Weibull and extreme value models; see the Computational Notes at the end of the chapter.

Likelihood ratio procedures do not depend on what parameterization is used for the models. The appropriate statistics, written in terms of u and b, are given in (5.2.6), (5.2.8), and (5.2.9). There is no need to employ derivatives of $\ell(u, b)$, and a simple but effective approach is to maximize constrained log-likelihoods such as (5.1.20) and (5.1.21) using general derivative-free optimization software (see Appendix D).

The bootstrap confidence intervals were obtained by the S-Plus 2000 implementation of the nonparametric BCa methods of Efron and Tibshirani (1993, Sec. 14.3).

Relative to the widths of the confidence intervals, the methods agree quite well, except for the Wald interval for β in the drug 6-MP group. This group has only nine failure times, and it is known that Wald procedures may not perform well in such cases. The approximate $N(0, 1)$ pivotal based on $\log \hat{b}$ gives intervals in much better agreement with the likelihood ratio (LR in the table) intervals. It would not be misleading to base conclusions on either the likelihood ratio or bootstrap intervals. It is known that the likelihood ratio procedures tend to be accurate with the sample

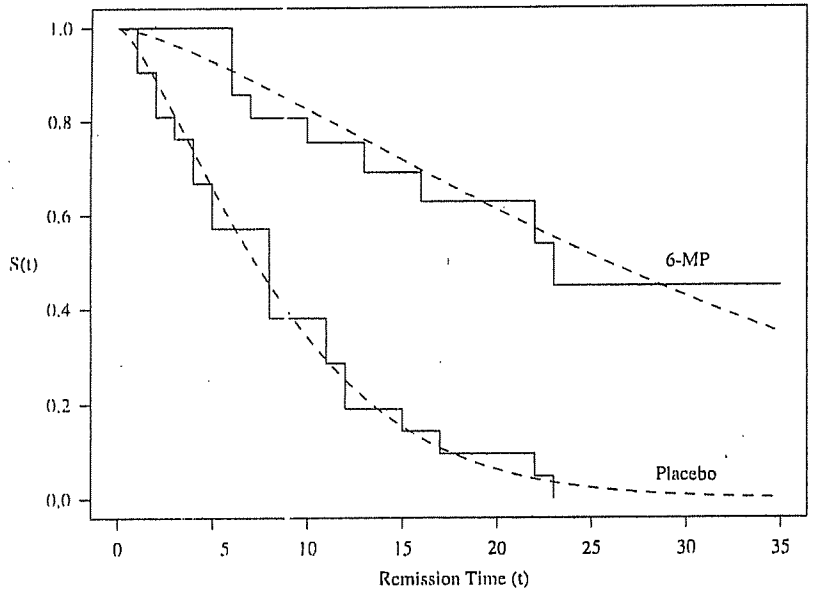

Figure 5.1. Weibull and KM estimates of survival (leukemia remission times).

sizes here, but if desired they can be improved slightly by making adjustments to the signed likelihood ratio statistic (e.g., Doganaksoy and Schmee 2000).

Confidence intervals for $S(10)$ were obtained by nonparametric methods in Example 3.2.3; likelihood ratio intervals were (.54, .90) for 6-MP and (.20, .59) for Placebo patients. The shorter intervals in Table 5.1 reflect the additional information that assumption of a parametric model invokes. The Weibull model fits well, so there is some comfort level in using it for inference. Another advantage of the parametric model is that it provides a confidence interval for the median for each group, whereas the nonparametric method cannot do this for the 6-MP group, because of the degree of censoring. In any case, both the parametric and nonparametric analyses make it clear that there is a substantial difference in the remission time distributions for the drug 6-MP and Placebo patients.

Figure 5.1 shows the Weibull and Kaplan–Meier estimates of $S(t)$ for each treatment group.

Example 5.2.2. These data are based on Type 2 censored lifetimes with $r = 28$ and $n = 40$, and were discussed by Lawless (1975, p. 258). They will also be considered later in Example 5.2.3, because it is possible to construct exact confidence intervals when data are Type 2 censored. Here we illustrate the large-sample procedures of the current section, and compare the results from them with the exact intervals.

Table 5.2. .90 Confidence Intervals Obtained by Three Methods

Method	b	u	$y_{.10}$	$y_{.01}$
Wald	(.66, 1.16)	(−.13, .44)	(−2.50, −1.29)	(−5.19, −2.87)
LR	(.70, 1.22)	(−.12, .48)	(−2.62, −1.37)	(−5.44, −3.05)
Exact	(.72, 1.28)	(−.11, .51)	(−2.71, −1.41)	(−5.67, −3.15)

The ordered log failure times $y_i = \log t_i$ are −2.982, −2.849, −2.546, −2.350, −1.983, −1.492, −1.443, −1.394, −1.386, −1.269, −1.195, −1.174, −.845, −.620, −.576, −.548, −.247, −.195, −.056, −.013, .006, .033, .037, .046, .084, .221, .245, .296. A probability plot of the data indicates that the data are consistent with an extreme value model for Y.

The m.l.e.'s and asymptotic covariance matrix for the extreme value parameters u and b are $\hat{u} = .1563$, $\hat{b} = .9104$, and

$$\hat{V} = I(\hat{u}, \hat{b})^{-1} = \begin{pmatrix} .02994 & .00282 \\ .00282 & .02336 \end{pmatrix}.$$

Lower quantiles of a lifetime distribution are often of interest in engineering or reliability settings, because of a desire to identify times before which few failures are likely to occur. Table 5.2 shows .90 Wald-type confidence intervals for u, b, and quantiles $y_{.10}$ and $y_{.01}$, obtained using the approximate $N(0, 1)$ pivotal quantities (5.1.11) and (5.1.13); note that $y_p = u + \log(-\log(1-p))b$, so that $y_{.10} = u - 2.25b$ and $y_{.01} = u - 4.60b$. Table 5.2 also shows likelihood ratio (LR in the table) confidence intervals for each parameter. Finally, exact conditional confidence intervals are shown; they are obtained using methods in Section 5.2.2.1 and their calculation is outlined in Example 5.2.3.

The likelihood ratio intervals agree well with the exact intervals, the only discrepancy of any size being for the extreme quantile, $y_{.01}$. The Wald intervals also agree quite well relative to the widths of the various intervals. It would not be misleading to use intervals based on either of the approximate methods, but the likelihood ratio procedures are preferred. Bootstrap intervals, not shown here but similar to those used in Example 5.2.1, agree well with the likelihood ratio intervals.

5.2.2 Exact Confidence Intervals Under Type 2 Censoring

It was noted in Section 5.1.2 that when data are complete or Type 2 censored, the quantities $Z_1 = (\hat{u} - u)/\hat{b}$, $Z_2 = \hat{b}/b$, and $Z_p = (\hat{u} - y_p)/\hat{b}$ are pivotal quantities. Exact conditional tests or confidence intervals for parameters are rather easily obtained; we consider them first and then discuss exact unconditional intervals.

5.2.2.1 Conditional Confidence Intervals

The conditional distributions for Z_1 and Z_2, given the vector $\mathbf{a} = (a_1, \ldots, a_r)$ of ancillary statistics defined in Theorem 5.1.1, is derived in Theorem E3 of Appendix E. This shows that for the general location-scale model (5.1.1), the joint

p.d.f. of Z_1, Z_2 given $\mathbf{a}$ is of the form

$$k(\mathbf{a}, r, n) z_2^{r-1} \left(\prod_{i=1}^{r} f_0(a_i z_2 + z_1 z_2) \right) S_0(a_r z_2 + z_1 z_2)^{n-r}, \tag{5.2.10}$$

where $k(\mathbf{a}, r, n)$ is a function of $a_1, \ldots, a_{r-2}, r$ and n. For the extreme value distribution, the density $f_0(z)$ and survivor function $S_0(z)$ are as given in (5.2.3), so that (5.2.10) becomes

$$k(\mathbf{a}, r, n) z_2^{r-1} \exp\left(\sum_{i=1}^{r} (a_i z_2 + z_1 z_2) - \sum_{i=1}^{r^*} e^{a_i z_2 + z_1 z_2} \right), \tag{5.2.11}$$

where we employ the notation $\sum_{i=1}^{r^*} w_i = \sum_{i=1}^{r} w_i + (n-r) w_r$. From (5.2.11) we can derive the marginal distributions for each of Z_1, Z_2, and Z_p, conditional on $\mathbf{a}$. The results are given in the following theorem (see Lawless 1978).

THEOREM 5.2.1. Let $Z_p = (\tilde{u} - y_p)/\tilde{b} = (\tilde{u} - u - w_p b)/\tilde{b}$ and $Z_2 = \tilde{b}/b$, where $w_p = \log[-\log(1-p)]$, $\tilde{u}$ and $\tilde{b}$ are equivariant estimators of u and b based on a Type 2 censored sample $y_1 \le \cdots \le y_r$ from the extreme value distribution (5.2.2), and $a_i = (y_i - \tilde{u})/\tilde{b}$, $i = 1, \ldots, r$. Then

i. The conditional p.d.f. of Z_2, given $\mathbf{a}$, is of the form

$$h_2(z|\mathbf{a}) = \frac{k'(\mathbf{a}, r, n) z^{r-2} \exp\left((z-1) \sum_{i=1}^{r} a_i \right)}{\left(\frac{1}{r} \sum_{i=1}^{r^*} e^{a_i z} \right)^r} \qquad z \ge 0 \tag{5.2.12}$$

ii. The conditional distribution function (d.f.) of Z_p, given $\mathbf{a}$, is

$$Pr(Z_p \le t|\mathbf{a}) = \int_0^\infty h_2(z|\mathbf{a}) I\left(r, e^{w_p + tz} \sum_{i=1}^{r^*} e^{a_i z} \right) dz \tag{5.2.13}$$

where $I(r, s)$ is the incomplete gamma function (B12). The d.f. of Z_1, given $\mathbf{a}$, is given by (5.2.13) with $w_p = 0$.

Proof.

i. To obtain (5.2.12) we integrate z_1 out of (5.2.11). This takes the form

$$h_2(z_2|\mathbf{a}) = k(\mathbf{a}, r, n) z_2^{r-1} \int_{-\infty}^{\infty} \exp\left(z_2 \sum_{i=1}^{r} a_i + r z_1 z_2 - e^{z_1 z_2} \sum_{i=1}^{r^*} e^{a_i z_2} \right) dz_1.$$

Letting

$$y = \left(\sum_{i=1}^{r^*} e^{a_i z_2} \right) e^{z_1 z_2},$$

one finds

$$h_2(z_2|\mathbf{a}) = \left[k(\mathbf{a}, r, n) \exp\left(\sum_{i=1}^{r} a_i z_2 \right) z_2^{r-2} / \left(\sum_{i=1}^{r^*} e^{a_i z_2} \right)^r \right] \Gamma(r).$$

This is essentially (5.2.12), where for convenience several terms are grouped together to form the new constant

$$k'(\mathbf{a}, r, n) = \frac{k(\mathbf{a}, r, n) \Gamma(r) \exp(\sum a_i)}{r^r}.$$

ii. It is not possible to integrate z_2 out of (5.2.11) analytically, so we work with the distribution function (d.f.) of Z_p, given $\mathbf{a}$. The joint p.d.f. of Z_2 and $Z_p = Z_1 - w_p Z_2^{-1}$, given $\mathbf{a}$, is easily found from (5.2.11) to be

$$h(z, z_2|\mathbf{a}) = k(\mathbf{a}, r, n) z_2^{r-1} \exp\left(\sum_{i=1}^{r} (a_i z_2 + z_2 z + w_p) - \sum_{i=1}^{r^*} \exp(a_i z_2 + z_2 z + w_p) \right)$$

where $z_2 > 0$, $-\infty < z < \infty$. The d.f. of Z_p, given $\mathbf{a}$, is

$$Pr(Z_p \le t|\mathbf{a}) = \int_0^\infty \int_{-\infty}^{t} h(z, z_2|\mathbf{a}) \, dz \, dz_2.$$

Making the change of variables

$$y = e^{z z_2} \sum_{i=1}^{r^*} e^{a_i z_2 + w_p}, \qquad z_2 = z_2,$$

we get

$$Pr(Z_p \le t|\mathbf{a}) = \int_0^\infty \left[\frac{k(\mathbf{a}, r, n) z_2^{r-2} \exp(\sum_{i=1}^{r} a_i z_2 + r w_p)}{(\sum_{i=1}^{r^*} e^{a_i z_2 + w_p})^r} \times \int_0^{t^*} y^{r-1} e^{-y} \, dy \right] dz_2$$

where

$$i^* = \left(\sum_1^{r^*} e^{a_i z_2}\right) \exp(t z_2 + w_p).$$

After a little rearrangement, this gives (5.2.13).

To construct confidence intervals for u, b, or y_p one needs to obtain percentage points for Z_1, Z_2, or Z_p. These are readily found from the results of Theorem 5.2.1, though numerical integration is necessary to integrate (5.2.12) and to evaluate (5.2.13). This is described in Example 5.2.3. It can be shown (see Appendix E) that different equivariant estimators yield the same confidence intervals for a given sample, and therefore it is immaterial what estimators are used to form the pivotals and $\mathbf{a}$, but we use the m.l.e.'s Note that although $h_2(z_2|\mathbf{a})$ involves an unknown constant $k'(\mathbf{a}, r, n)$, this can be evaluated by using the fact that $h_2(z_2|\mathbf{a})$ must integrate to one.

The mechanics of the method will be made clear in the following example.

Example 5.2.3. (Example 5.2.2 revisited). We will get confidence intervals by working with pivotal quantities and ancillaries based on the m.l.e.'s. The m.l.e.'s, from Example 5.2.2, are $\hat{u} = .1563$ and $\hat{b} = .9104$, and the ancillaries are $a_i = (y_i - .1563)/.9104$, $i = 1, \ldots, 28$. Before obtaining the confidence intervals, let us consider the type of calculations required. The constant $k'(\mathbf{a}, r, n)$ in (5.2.12) can be evaluated from the fact that

$$\int_0^\infty h_2(z|\mathbf{a})\, dz = 1,$$

which implies that

$$k'(\mathbf{a}, r, n) = \left[\int_0^\infty z^{r-2} \exp\left((z-1)\sum_{i=1}^r a_i\right) \Big/ \left(\frac{1}{r}\sum_{i=1}^{r^*} e^{a_i z}\right)^r dz\right]^{-1}. \quad (5.2.14)$$

The integrand in (5.2.14) is well behaved and the integral is easily evaluated numerically with standard software The integrand is generally very close to 0 for z outside of the range 0 to 10.

Having obtained $k'(\mathbf{a}, r, n)$, we can easily determine percentage points for Z_2, using

$$Pr(Z_2 \le l|\mathbf{a}) = \int_0^l h_2(z|\mathbf{a})\, dz. \quad (5.2.15)$$

Exact percentage points $l = z_{2,\gamma}$ making (5.2.15) equal to γ can be obtained iteratively.

Getting percentage points for Z_p involves similar computations with (5.2.13). This integral behaves in much the same way as integrals (5.2.15). Since $I(r, s) \le 1$, the integrand in (5.2.13) is in fact always less than or equal to the integrand in (5.2.14). Finally, note that if a significance test for some specified value of b or y_p is wanted, one needs to calculate only a single probability for Z_2 or Z_p.

Let us now consider the data in the example. By numerical integration, the integral in (5.2.14) is found to be .4355, and hence $k'(\mathbf{a}, r, n) = 2.2961$. We now have the complete p.d.f. $h_2(z|\mathbf{a})$, and can calculate any desired probabilities for Z_2, Z_1, or Z_p. Suppose, for example, we want a two-sided .90 confidence interval for b. Integrating $h_2(z|\mathbf{a})$ numerically, we determine that $P(Z_2 \le .713|\mathbf{a}) = .05$ and $Pr(Z_2 \le 1.257|\mathbf{a}) = .95$. Thus

$$Pr(.713 \le \hat{b}/b \le 1.257|\mathbf{a}) = .90,$$

and this yields $.724 \le b \le 1.277$ as a .90 confidence interval for b from the observed value $\hat{b} = .9104$.

Suppose we also want a lower .95 confidence limit for $y_{.10}$. From (5.2.13), we find that

$$Pr(Z_{.10} \le 3.153|\mathbf{a}) = .95$$

where $Z_{.10} = (\hat{u} - y_{.10})/\hat{b}$. Thus $y_{.10} \ge \hat{u} - 3.153\hat{b}$ is the desired confidence interval, which gives $y_{.10} \ge -2.714$ as the realized interval.

Let us also consider confidence limits for the survivor function, continuing to work with the extreme value distribution for which $S(y_0) = \exp(-e^{(y_0-u)/b})$. Confidence limits or tests for $S(y_0)$ can be obtained from the pivotal Z_p used to get confidence limits for quantiles, because of the relationship between the survivor function and quantiles. Specifically, suppose that $\ell(\mathbf{y})$ is a lower q confidence limit on y_p based on data $\mathbf{y}$; that is, $Pr(\ell(\mathbf{y}) \le y_p) = q$. This is true if and only if $Pr[S(\ell(\mathbf{y})) \ge 1 - p] = q$, so if we determine $p = p(\mathbf{y})$ such that $\ell(\mathbf{y}) = y_0$, then $1 - p$ is a lower q confidence limit for $S(y_0)$.

To obtain a lower q confidence limit for $S(y_0)$ using Z_p, one can therefore proceed as follows: since the lower q confidence limit for y_p is of the form $\ell(\mathbf{y}) = \hat{u} - \hat{b}z_{p,q}$, where $z_{p,q}$ is the qth quantile of Z_p, it is merely necessary to determine p such that

$$-z_{p,q} = (y_0 - \hat{u})/\hat{b}. \quad (5.2.16)$$

The lower q confidence limit on $S(y_0)$ is then $1 - p$.

Example 5.2.3. (continued) Suppose that a lower .95 confidence limit on $S(-1.0) = \exp(-e^{(-1-u)/b})$ is wanted. According to (5.2.16), we need to determine p such that

$$-z_{p,.95} = \frac{-1 - \hat{u}}{\hat{b}} = -1.207,$$

recalling that $\hat{u} = .1563$ and $\hat{b} = .9104$. That is, we must find p such that $Pr(Z_p \leq 1.207|a) = .95$, using (5.2.13). Noting that $\hat{S}(-1.0) = \exp(-e^{(-1-\hat{u})/\hat{b}}) = .755$ and starting with $1 - p$ in the neighborhood of .70, we easily find after a few iterations with (5.2.13) that the required value of p is .353. The .95 confidence interval for $S(-1.0)$ is therefore $S(-1.0) \geq .647$.

The conditional method can be used to obtain confidence intervals whatever the size of the Type 2 censored sample. It can also be used with Type 2 progressively censored data (see Problem 5.6). In even moderate-size samples the unconditional methods based on the same pivotals Z_1, Z_2, and Z_p tend to agree closely with the conditional methods, and provide a viable alternative. We discuss these methods next.

5.2.2.2 *Unconditional Confidence Intervals*

As noted in Section 5.1.2, confidence intervals can also be based on the marginal distributions of the pivotal quantities Z_1, Z_2, and Z_p. It is generally impossible to obtain these distributions in a simple form suitable for numerical evaluation; for the case of the extreme value distribution, in particular, $a_1, \ldots, a_{r-2}$ must be integrated out of (5.2.11). However, the distributions of Z_1, Z_2, and Z_p can for specified values of r and n be estimated very accurately by simulation, as discussed in Section 5.1.2. Some tables of quantiles obtained via large-scale simulations for the extreme value distribution have been published (e.g., Thoman et al. 1969, 1970; Billmann et al. 1972; McCool 1970, 1974). The amount of computation required to estimate the distributions of Z_1, Z_2, or Z_p accurately enough for practical purposes, for given r and n, is of the same order as for approximation of the distributions using naive nonparametric or parametric bootstrap methods; about 2000 random samples is sufficient. It is therefore feasible to obtain quantiles as needed, though tables are useful if available.

A very simple chi-squared approximation has been found for the distribution of $Z_2 = \hat{b}/b$. It was developed empirically and is accurate enough for virtually all practical situations involving Type 2 censoring. We provide it here since it offers a convenient way to estimate or compare Weibull shape parameters β, and it can provide a check on the accuracy of Wald-type large-sample procedures for b or β.

The approximation is of the form

$$g(r, n)\left(\frac{\hat{b}}{b}\right) \sim \chi^2_{(h(r,n))}, \tag{5.2.17}$$

where $g(r, n)$ and $h(r, n)$ are constants specified below. Their values are determined by matching the first two moments of $g(r, n)(\hat{b}/b)$ to those of $\chi^2_{(h(r,n))}$. Since the moments of $\hat{b}/b$ are not known exactly, they have been estimated by simulation and approximated in various ways. Table 5.3 was prepared by combining results given by Harter and Moore (1968), McCool (1975b), and Lawless and Mann (1976). Values of $h(r, n)$ are given for various (r, n) combinations; it follows from (5.2.17) that

$$g(r, n) = h(r, n) + 2.$$

Table 5.3. Values $h(r, n)$ for Use in Approximation (5.2.17)[a]

		n							
		5	10	20	40	60	80	100	∞
	.1	—	—	2.0	6.0	10.0	14.1	18.1	$0.205n$
	.2	—	2.0	6.2	14.6	23.0	31.5	39.9	$0.420n$
	.3	—	4.3	10.9	24.0	37.0	50.1	63.2	$0.652n$
	.4	2.2	6.7	15.8	33.8	51.8	69.9	87.9	$0.899n$
r/n	.5	3.5	9.1	20.7	44.0	67.3	90.6	113.9	$1.165n$
	.6	4.7	11.4	25.8	54.7	83.5	112.3	141.1	$1.457n$
	.7	6.0	14.8	32.6	68.1	103.8	139.5	175.0	$1.782n$
	.8	7.8	18.5	40.0	83.3	126.4	169.5	212.5	$2.155n$
	.9	10.3	23.0	49.0	100.9	153.0	204.9	256.9	$2.607n$
	1.0	12.9	29.3	62.4	128.2	194.8	257.6	325.5	$3.290n$

[a] $g(r, n) = h(r, n) + 2$.

Comparison of percentage points of $\hat{b}/b$ obtained by using (5.2.17) and these values of $h(r, n)$ with the essentially exact percentage points given by McCool (1974, 1975b) and Billmann et al. (1972) shows that the approximation is adequate for virtually all situations. The approximation tends to improve as n or r/n increases and is exact in the limit as $n \to \infty$, for fixed r/n. For values of n and r/n not covered in Table 5.3, suitable values of $h(r, n)$ can be obtained by linear interpolation.

Example 5.2.4. As a comparison of the exact conditional and unconditional confidence interval procedures, consider the following artificial Type 2 censored sample $y_1 < \cdots < y_{10}$ from the standard extreme value distribution, with $n = 20$, $r = 10$: -3.57, -2.55, -2.02, 1.66, -1.36, -1.15, $-.95$, $-.77$, $-.61$, $-.45$. Table 5.4 shows .90 confidence intervals based on pivotals Z_1, Z_2, and Z_p of (5.1.11) and (5.1.13). The m.l.e.'s of u and b are $\hat{u} = -.112$ and $\hat{b} = .907$; the conditional intervals were obtained as described in Section 5.2.2.1, and the unconditional intervals were based on percentage points given by McCool (1974). It is clear that there is no practical difference in the two sets of intervals. We remark in passing that the use of approximation (5.2.17) and Table 5.3 gives the same unconditional interval for b (to two decimals) as in Table 5.4.

Table 5.4. Two-Sided .90 Conditional and Unconditional Confidence Intervals

Parameter	Unconditional Method	Conditional Method
b	(.64, 1.81)	(.64, 1.82)
$u(y_{.632})$	(−.51, .90)	(−.51, .89)
$y_{.10}$	(−3.74, −1.49)	(−3.76, −1.51)

5.3 LOG-NORMAL AND LOG-LOGISTIC DISTRIBUTIONS

5.3.1 Inferences for Log-Normal and Normal Models

The log-normal and log-logistic distributions are similar in many respects, and are discussed together in this section.

The log-normal distribution (1.3.10) for lifetime T is equivalent to $Y = \log T$ having a normal distribution $N(\mu, \sigma^2)$. The normal model is of location-scale form (5.1.1), with $u = \mu$, $b = \sigma$ and standard survivor and density functions

$$S_0(z) = 1 - \Phi(z), \qquad f_0(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \tag{5.3.1}$$

where $\Phi(z)$ and $\phi(z)$ are the standard normal cumulative distribution function (c.d.f.) and p.d.f. This and (5.1.4) give the log-likelihood function for a censored random sample as

$$\ell(\mu, \sigma) = -r \log \sigma - \frac{1}{2} \sum_{i=1}^{n} \delta_i z_i^2 + \sum_{i=1}^{n} (1 - \delta_i) \log S_0(z_i), \tag{5.3.2}$$

where $z_i = (y_i - \mu)/\sigma$ and $r = \sum \delta_i$. The first and second derivatives of $\ell(\mu, \sigma)$ are given by (5.1.5)–(5.1.9) with

$$\frac{\partial \log f_0(z)}{\partial z} = -z, \qquad \frac{\partial^2 \log f_0(z)}{\partial z^2} = -1$$

$$\frac{\partial \log S_0(z)}{\partial z} = -\frac{f_0(z)}{S_0(z)}, \qquad \frac{\partial^2 \log S_0(z)}{\partial z^2} = \frac{z f_0(z)}{S_0(z)} - \frac{f_0(z)^2}{S_0(z)^2}.$$

The log-likelihood (5.3.2) is easily maximized, and many software packages handle the procedures below. The observed information matrix $I(\hat{\mu}, \hat{\sigma})$ and asymptotic covariance matrix $\hat{V} = I(\hat{\mu}, \hat{\sigma})^{-1}$ for $(\hat{\mu}, \hat{\sigma})$ are calculated using (5.1.7)–(5.1.9), with (μ, σ) replacing (u, b). Confidence intervals or tests for μ, σ, quantiles $y_p = \mu + \Phi^{-1}(p)\sigma$ or probabilities $S_0[(y_0 - \mu)/\sigma]$ can be based on pivotals (5.1.11) and (5.1.13) or on likelihood ratio procedures described in Section 5.1.1.

5.3.1.1 Exact Methods

With complete (uncensored) samples, exact test and confidence interval procedures for μ and σ are well-known and discussed in elementary texts. In particular, $\hat{\mu} = \bar{y}$ and $\hat{\sigma} = [(n-1)s^2/n]^{1/2}$, where $\bar{y} = \sum y_i/n$ and $s^2 = \sum (y_i - \bar{y})^2/(n-1)$ are the sample mean and variance, and the pivotal quantities $Z_1 = \sqrt{n}(\bar{y} - \mu)/s$ and $Z_2 = (n-1)s^2/\sigma^2$ have student-t and chi-squared distributions with $n-1$ degrees of freedom, respectively. Note that these pivotal quantities are written in a different form than those used for general location-scale models in Theorem 5.1.1.

Estimation of quantiles or the survivor function of the normal distribution is not discussed in most elementary texts, so we consider it for the case of complete samples. This can be approached through the pivotal quantity

$$Z_p = \frac{\tilde{y}_p - y_p}{s}, \tag{5.3.3}$$

where $\tilde{y}_p = \bar{y} + w_p s = \bar{y} + \Phi^{-1}(p)s$ estimates y_p. Probabilities for Z_p can be obtained from those for the noncentral t distribution, which is defined as follows. A noncentral t random variable with v degrees of freedom and noncentrality parameter λ, denoted $t'_{(v)}(\lambda)$, arises by considering

$$t = \frac{Z + \lambda}{(W/v)^{1/2}},$$

where $Z \sim N(0, 1)$ and $W \sim \chi^2_{(v)}$ are independent random variables and λ is a constant. The noncentral t distribution is discussed in detail by Johnson et al. (1995, Ch. 31). From (5.3.3) we now have that

$$\begin{aligned} Pr(Z_p \le z) &= Pr\left[\frac{\sqrt{n}(\bar{y} - \mu)/\sigma - \sqrt{n}w_p}{s/\sigma} \le \sqrt{n}(z - w_p)\right] \\ &= Pr\left[t'_{(n-1)}(-\sqrt{n}w_p) \le \sqrt{n}(z - w_p)\right]. \end{aligned} \tag{5.3.4}$$

Tables and software exist for noncentral t probabilities and quantiles, as discussed by Johnson et al. (1995, Ch. 31); an illustration of their use is given in Example 5.3.1.

Exact tests and confidence interval procedures are also available when data are Type 2 censored, according to the methods discussed in Section 5.1.2 and Appendix E. The distributions of the exact pivotal quantities Z_1, Z_2, and Z_p of Theorem 5.1.1 are analytically intractable for $r < n$, but can be approximated closely by simulation. Schmee and Nelson (1977) give tables and charts for obtaining confidence intervals that were determined in this way. Conditional distributions of the pivotal quantities Z_1, Z_2, and Z_p given the observed value of the ancillary statistic $\mathbf{a}$ defined in Theorem 5.1.1 can also be used for inference. The distribution of Z_1 and Z_2 given $\mathbf{a}$ is given in Theorem E3, and it can be seen that numerical double integration is necessary to obtain probabilities or quantiles for any of Z_1, Z_2, or Z_p. As a result, and because unconditional and conditional probabilities for Z_1, Z_2, and Z_p tend to agree closely except for very small samples, this is not generally used.

Examples of the methodology in this section are deferred until Section 5.3.3, where both log-normal and log-logistic models are fitted to data.

5.3.2 Inferences for Log-Logistic and Logistic Models

The log-logistic distribution (1.3.12) for a lifetime variable T and the corresponding logistic distribution for $Y = \log T$ were discussed in Section 1.3.4. The logistic distribution is of location-scale form (5.1.1) with standard survivor and density

functions

$$S_0(z) = \frac{1}{(1+e^z)}, \qquad f_0(z) = \frac{e^z}{(1+e^z)^2} \tag{5.3.5}$$

respectively. The log-likelihood function for a censored random sample is, from (5.1.4),

$$\ell(u, b) = -r \log b + \sum_{i=1}^{n} \delta_i [z_i - 2\log(1+e^{z_i})] - \sum_{i=1}^{n} (1-\delta_i) \log(1+e^{z_i}), \tag{5.3.6}$$

where $z_i = (y_i - u)/b$ and $r = \sum \delta_i$. The first and second derivatives of $\ell(u, b)$ are given by (5.1.5)–(5.1.9) with

$$\frac{\partial \log f_0(z)}{\partial z} = 1 - \frac{2e^z}{1+e^z}, \qquad \frac{\partial^2 \log f_0(z)}{\partial z^2} = \frac{-2e^z}{(1+e^z)^2}$$

$$\frac{\partial \log S_0(z)}{\partial z} = \frac{-e^z}{1+e^z}, \qquad \frac{\partial^2 \log S_0(z)}{\partial z^2} = \frac{-e^z}{(1+e^z)^2}.$$

The m.l.e.'s $\hat{u}$, $\hat{b}$, and associated large-sample procedures for obtaining tests or confidence intervals are available from several common software packages. The observed information matrix $I(\hat{u}, \hat{b})$ and asymptotic covariance matrix $\hat{V} = I(\hat{u}, \hat{b})^{-1}$ for $(\hat{u}, \hat{b})$ are calculated using (5.1.7)–(5.1.9). Confidence intervals or tests for u, b, quantiles $y_p = u + [\log(p/(1-p))]b$ or probabilities $S_0[(y_0 - u)/b]$ can be based on the usual pivotals (5.1.11) and (5.1.13) or on likelihood ratio statistics, as described in Section 5.1.1.

Exact tests and confidence intervals based on the pivotals Z_1, Z_2, and Z_p of Theorem 5.1.1 are available when the data are Type 2 censored. Probabilities and quantiles for the pivotals can be obtained by simulation as described in Section 5.1.2. Unlike the cases of the extreme value and normal distributions, no tables seem to have been published for the logistic model. Conditional probabilities for the pivotal quantities, given the ancillary statistic $\mathbf{a}$ of Theorem 5.1.1, require numerical double integration.

5.3.3 Examples

The examples below illustrate applications of the log-normal and log-logistic distributions. The two distributions have similarly shaped density and hazard functions, and it is usually difficult to discriminate between them unless the data contain a reasonably large number of failure times.

Example 5.3.1. The data below were discussed by Schmee and Nelson (1977) and show the number of thousand miles at which different locomotive controls failed, in a life test involving 96 controls. The test was terminated after 135,000 miles, by which time 37 failures had occurred. The failure times for the 37 failed units are

22.5, 37.5, 46.0, 48.5, 51.5, 53.0, 54.5, 57.5, 66.5, 68.0, 69.5, 76.5, 77.0, 78.5, 8.0, 81.5, 82.0, 83.0, 84.0, 91.5, 93.5, 102.5, 107.0, 108.5, 112.5, 113.5, 116.0, 117.0, 118.5, 119.0, 120.0, 122.5, 123.0, 127.5, 131.0, 132.5, 134.0. In addition, there are 59 censoring times, all equal to 135.0.

Probability plots suggest that either a log-normal or log-logistic model fits the data rather well. The log-normal and log-logistic models will be written here with distribution functions of the respective forms

$$F_{LN}(t; u, b) = \Phi\left(\frac{\log t - u}{b}\right), \qquad t \geq 0$$

$$F_{LL}(t; u, b) = \frac{e^{(\log t - u)/b}}{1 + e^{(\log t - u)/b}}, \qquad t \geq 0$$

to emphasize the location-scale parameters (u, b), which it is convenient to use. The m.l.e.'s and asymptotic covariance matrices $\hat{V} = I(\hat{u}, \hat{b})^{-1}$ are obtained from standard software as

$$\text{log-normal:} \quad \hat{u} = 5.117, \quad \hat{b} = .706, \quad \hat{V}_N = \begin{pmatrix} .01085 & .005729 \\ .005729 & .008686 \end{pmatrix}$$

$$\text{log-logistic:} \quad \hat{u} = 5.083, \quad \hat{b} = .384, \quad \hat{V}_L = \begin{pmatrix} .008018 & .002550 \\ .002550 & .003244 \end{pmatrix}.$$

The maximum log-likelihoods for the models are almost equal; the log-normal's is .14 larger. Note that although u is the median for both models, the interpretation of b differs: for the log-normal model the variance of $\log T$ is b^2, whereas for the log-logistic model, it is $\pi^2 b^2/3$. Figure 5.2 shows plots of the two fitted survivor functions $1 - F_{LN}(t; \hat{u}, \hat{b})$ and $1 - F_{LL}(t; \hat{u}, \hat{b})$, along with the Kaplan–Meier estimate of $S(t)$.

An objective of many life tests is to determine times before which the fraction of units failing is small, or to estimate the reliability at certain times. We will use the two fitted models to estimate the reliability at 80,000 miles, that is, $S(80)$, which was of interest because of warranty coverage.

Let us consider the procedure based on the approximate standard normal pivotal quantity (5.1.14). Following the argument leading to (5.1.16), we find $\hat{\psi} = (y_0 - \hat{u})/\hat{b}$ for either model, and

$$se(\hat{\psi}) = \frac{1}{\hat{b}}(\hat{V}_{11} + 2\hat{\psi}\hat{V}_{12} + \hat{\psi}^2 \hat{V}_{22})^{1/2},$$

where $y_0 = \log(80)$. Estimates for $S(80)$ are then $\hat{S}_{LN}(80) = 1 - \Phi(\hat{\psi})$ and $\hat{S}_{LL}(80) = (1 + \exp(\hat{\psi}))^{-1}$ for the log-normal and log-logistic models, respectively. Two-sided .95 confidence intervals for ψ are computed as $\hat{\psi} \pm 1.96 se(\hat{\psi})$, and are transformed to confidence intervals for $S(80)$ by the relationships $S_{LN}(80) = 1 - \Phi(\psi)$ and $S_{LL}(80) = (1 + \exp(\psi))^{-1}$. This yields the following estimates and

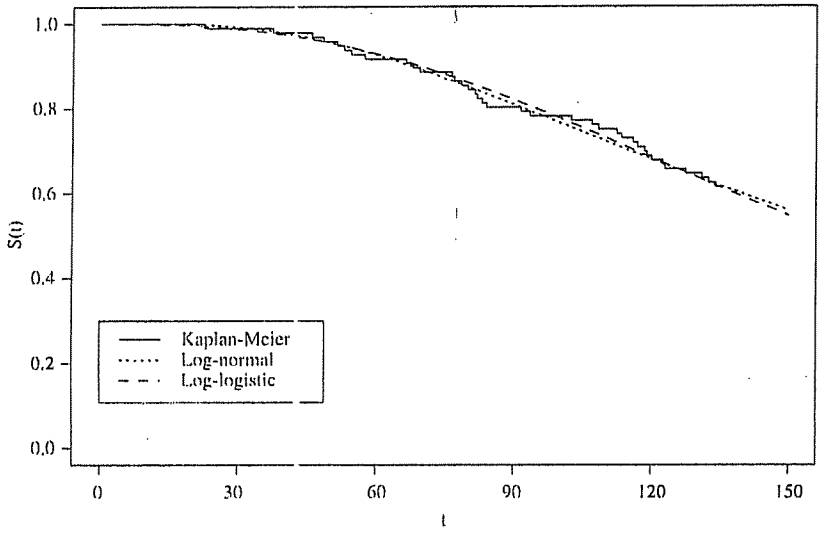

Figure 5.2. Parametric and KM estimates of failure time s.f. for locomotive controls.

.95 confidence intervals for $S(80)$:

$$\text{log-normal:} \quad \hat{S}(80) = .851, \qquad .785 \le S(80) \le .902$$

$$\text{log-logistic:} \quad \hat{S}(80) = .861, \qquad .791 \le S(80) \le .911.$$

The confidence intervals for $S(80)$ agree well, as they should since the survivor functions (s.f.'s) for both models agree closely with each other and with the Kaplan–Meier estimate of $S(t)$ over the range 0–135. By way of comparison with confidence intervals constructed using the likelihood ratio procedures described in Section 5.1.1, we used (5.1.19) and (5.1.20) with the log-normal model to find a two-sided .95 confidence interval for $S(80)$. This gave $.785 \le S(80) \le .903$, in very close agreement with the interval obtained using the approximate standard normal pivotals.

The log-normal and log-logistic distributions have hazard functions that first increase and then decrease at higher times. However, the points at which the hazard functions begin to decrease are well beyond the censoring time of 135,000 miles. It is impossible to draw any conclusions about the right half of the distribution from these data. Indeed, a Weibull model for which the hazard function is monotone increasing also fits these data well; we consider this in Example 5.5.1.

The following example compares exact confidence intervals for a log-normal quantile with intervals based on maximum likelihood large-sample theory.

Example 5.3.2. Example 3.3.1 and 3.3.2 discussed uncensored data on the number of millions of cycles to failure for 23 ball bearings in a life test. The data

are given in Example 3.3.1 and appear consistent with a log-normal distribution for time (i.e., number of million cycles) to failure, T. Lower confidence limits on quantiles such as $t_{.10}$ and $t_{.01}$ are often used to rate bearings with respect to endurance or reliability. We will compare limits obtained using the large-sample approximations of Section 5.3.1 with exact confidence limits, which are in this case available because the data are uncensored. We revert to using μ and σ to represent the normal distribution location and scale parameters, instead of u, b, as in the preceding example.

The sample mean and variance for the 23 log failure times $y_1, \ldots, y_{23}$ are $\bar{y} = 4.150$ and $s^2 = .2841$ ($s = .534$). The m.l.e.'s of μ and σ are consequently $\hat{\mu} = 4.150$ and $\hat{\sigma} = (22/23)^{1/2}s = .522$. The .01 and .10 quantiles of Y are $y_{.01} = \mu - 2.326\sigma$ and $y_{.10} = \mu - 1.282\sigma$, respectively. Using the m.l.e.'s, the asymptotic covariance matrix $\hat{V} = I(\hat{\mu}, \hat{\sigma})^{-1}$ and (5.1.12), we find $\hat{y}_{.10} = 3.481$ ($se = .175$) and $\hat{y}_{.01} = 2.937$ ($se = .271$). Approximate .95 lower confidence limits for y_p based on (5.1.13) are given by $\hat{y}_p - 1.645se(\hat{y}_p)$; this gives $y_{.10} \ge 3.194$ and $y_{.01} \ge 2.492$ and corresponding intervals $t_{.10} \ge 24.4$ and $t_{.01} \ge 12.1$ for the quantiles of $T = \exp Y$.

Exact confidence intervals for the quantiles are based on (5.3.3) and (5.3.4). For a .95 lower confidence limit for y_p, (5.3.3) and (5.3.4) indicate that we need the .95 quantile of the noncentral t variable $t'_{(22)}(-\sqrt{23}w_p)$. For $w_{.10} = -1.282$ and $w_{.01} = -2.326$ this gives $t'_{(22)}(6.14)$ and $t'_{(22)}(11.17)$. Tables in Owen (1968) give the .95 quantiles as 8.966 and 15.398, and yield the confidence intervals $y_{.10} \ge 3.152$, $y_{.01} \ge 2.435$ and $t_{.10} \ge 23.4$, $t_{.01} \ge 11.4$. The intervals based on the approximate $N(0, 1)$ pivotals (5.1.13) are in good agreement with the exact intervals.

5.4 COMPARISON OF DISTRIBUTIONS

The comparison of distributions is often important, for example, the comparison of failure time distributions for products manufactured at different sites, or the comparison of times to response for subjects in different arms of a clinical trial. Chapters 7 and 8 consider comparisons based on semiparametric models. In this section, comparisons based on parametric location-scale models are discussed. In particular, if lifetimes in m populations are distributed according to the same log-location-scale family (5.1.2), with survivor functions

$$S_j(t) = S_0^*[(t/\alpha_j)^{\beta_j}] \qquad j = 1, \ldots, m, \tag{5.4.1}$$

then comparison of the distributions merely involves a comparison of the parameters (α_j, β_j). In terms of log-lifetime $Y = \log T$, the survivor functions are

$$S_j(y) = S_0\left(\frac{y - u_j}{b_j}\right) \qquad j = 1, \ldots, m, \tag{5.4.2}$$

where $u_j = \log \alpha_j$, $b_j = \beta_j^{-1}$, and $S_0^*(w) = S_0(\log w)$.

The comparison of distributions (5.4.2) when $S_0(z)$ is normal (Gaussian) is a well-known problem. Comparisons are especially simple when the standard deviations σ_j $(j = 1, \ldots, m)$ are equal. The same is true in the general case: if $b_1 = b_2 = \cdots = b_m$, then the pth quantile of the jth distribution is $y_{jp} = u_j + w_p b$, where $w_p = S_0^{-1}(1-p)$, and so for any two distributions (say 1 and 2)

$$y_{1p} - y_{2p} = u_1 - u_2 \tag{5.4.3}$$

is a constant for all $0 < p < 1$. If the means of the two distributions exist, their difference is also equal to $u_1 - u_2$. Furthermore, testing equality of the distributions amounts to a test of $H: u_1 = u_2$.

In terms of lifetime T, (5.4.3) implies that the ratio of quantiles $t_{1p}/t_{2p} = \alpha_1/\alpha_2$ is constant for all p in (0, 1). There are also simple consequences of (5.4.3) for the survivor functions or the c.d.f.'s of Y and T:

$$S_2(y) = S_1[y + (u_1 - u_2)], \qquad S_2(t) = S_1[t(\alpha_1/\alpha_2)]. \tag{5.4.4}$$

In particular, the survivor or distribution functions for Y are translations of one another by an amount $u_1 - u_2$ along the y-axis.

One normally begins by comparing the scale parameters $b_1, \ldots, b_m$ in (5.4.2), or the shape parameters $\beta_1, \ldots, \beta_m$ in (5.4.1). If it can safely be assumed that they are equal, then comparison of quantiles or means reduces to a comparison of $u_1, \ldots, u_m$ or $\alpha_1, \ldots, \alpha_m$. If the parameters $b_1, \ldots, b_m$ are not equal, then (5.4.3) and (5.4.4) do not hold and the difference $y_{1p} - y_{2p} = (u_1 - u_2) + w_p(b_1 - b_2)$ depends on b_1, b_2, and p.

We will consider the comparison of arbitrary distributions (5.4.1) or (5.4.2), and then specific families of models. The comparisons in this section are based on independent samples from the distributions in question, and studies that use pairing are not discussed. This topic is considered in Problem 5.15, and further in Chapter 11.

5.4.1 General Methods for Comparing (Log-) Location-Scale Distributions

For convenience we continue to consider problems in terms of the location-scale parameters (u_j, b_j). It is assumed that there are independent censored random samples from the m distributions in question; the sample of lifetimes from the jth distribution will be denoted (t_{ji}, δ_{ji}), $i = 1, \ldots, n_j$, with $y_{ji} = \log(t_{ji})$.

One way to compare, say, pth quantiles y_{1p} and y_{2p} of two distributions is simply to examine confidence intervals for each. If a confidence interval or test for $y_{1p} - y_{2p}$ is wanted, then the approximate standard normal pivotal quantity

$$Z = \frac{(\hat{y}_{1p} - \hat{y}_{2p}) - (y_{1p} - y_{2p})}{[se(\hat{y}_{1p})^2 + se(\hat{y}_{2p})^2]^{1/2}} \tag{5.4.5}$$

can be used, where $\hat{y}_{1p}$, $\hat{y}_{2p}$, $se(\hat{y}_{1p})$, $se(\hat{y}_{2p})$ are the m.l.e.'s and standard errors from the two individual samples. Similarly, to compare b_1 and b_2 we can use the approximate standard normal pivotal

$$Z = \frac{(\log \hat{b}_1 - \log \hat{b}_2) - (\log b_1 - \log b_2)}{[se(\log \hat{b}_1)^2 + se(\log \hat{b}_2)^2]^{1/2}} \tag{5.4.6}$$

to obtain confidence intervals or tests for b_1/b_2.

5.4.1.1 *Likelihood Ratio Procedures*

Likelihood ratio procedures tend to be slightly more accurate than those based on (5.4.5) or (5.4.6) for small- or medium-size samples, as discussed in the preceding sections. It is also convenient to test equality of parameters for two or more distributions through likelihood ratio methods. First consider a test of the hypothesis $H: b_1 = \cdots = b_m$ for m distributions (5.4.2). The log-likelihood function for the combined sample from all m distributions is

$$\ell(u_1, \ldots, u_m, b_1, \ldots, b_m) = \sum_{j=1}^{m} \ell_j(u_j, b_j),$$

where $\ell_j(u_j, b_j)$ is the log-likelihood (5.1.4) for distribution j. To test H, we use the likelihood ratio statistic

$$\Lambda = 2\ell(\hat{u}_1, \ldots, \hat{u}_m, \hat{b}_1, \ldots, \hat{b}_m) - 2\ell(\tilde{u}_1, \ldots, \tilde{u}_m, \tilde{b}, \ldots, \tilde{b}), \tag{5.4.7}$$

where $\hat{u}_j, \hat{b}_j$ are the m.l.e.'s obtained from $\ell_j(u_j, b_j)$ and $\tilde{u}_1, \ldots, \tilde{u}_m, \tilde{b}$ are the values obtained by maximizing the constrained log-likelihood function $\ell(u_1, \ldots, u_m, b, \ldots, b)$. The distribution of Λ when H is true is asymptotically $\chi^2_{(m-1)}$ as the n_j become large.

Confidence intervals for the ratio of two scale parameters, say b_1/b_2, can be obtained by considering the likelihood ratio statistic $\Lambda(a)$ for testing $H: b_1 = ab_2$, where $a > 0$ is some specified value. Note that

$$\Lambda(a) = 2\ell(\hat{u}_1, \hat{u}_2, \hat{b}_1, \hat{b}_2) - 2\ell(\tilde{u}_1, \tilde{u}_2, a\tilde{b}_2, \tilde{b}_2), \tag{5.4.8}$$

where $\tilde{u}_1, \tilde{u}_2, \tilde{b}_2$ maximize $\ell(u_1, u_2, ab_2, b_2)$. An approximate q confidence interval for b_1/b_2 consists of values $\{a : \Lambda(a) \le \chi^2_{(1),q}\}$.

A test of equality for location parameters $u_1, \ldots, u_m$ is of interest mainly when the scale parameters $b_1, \ldots, b_m$ are equal. A likelihood ratio test of $H: u_1 = \cdots = u_m, b_1 = \cdots = b_m$ versus the alternative hypothesis $H_1: u_j$ not all equal; $b_1 = \cdots = b_m$ uses the statistic

$$\Lambda = 2\ell(\tilde{u}_1, \ldots, \tilde{u}_m, \tilde{b}, \ldots, \tilde{b}) - 2\ell(u^*, \ldots, u^*, b^*, \ldots, b^*), \tag{5.4.9}$$

where $\tilde{u}_1, \ldots, \tilde{u}_m$ and $\tilde{b}$ are as in (5.4.7), and u^*, b^* maximize $\ell(u, \ldots, u, b, \ldots, b)$. The distribution of Λ is close to $\chi^2_{(m-1)}$ for large samples, if H is true.

If $b_1 = b_2$, confidence intervals for $u_1 - u_2$ can be obtained by testing hypotheses $H_0: u_1 - u_2 = \delta, b_1 = b_2$ against the alternatives $H_1: u_1, u_2$ are unrestricted; $b_1 = b_2$.

The likelihood ratio statistic for doing this is

$$\Lambda(\delta) = 2\ell(\tilde{u}_1, \tilde{u}_2, \tilde{b}, \tilde{b}) - 2\ell(u_2^* + \delta, u_2^*, b^*, b^*), \tag{5.4.10}$$

where $\tilde{u}_1, \tilde{u}_2, \tilde{b}$ maximize $\ell(u_1, u_2, b, b)$ and u_2^*, b^* maximize $\ell(u_2 + \delta, u_2, b, b)$. An approximate q confidence interval for $u_1 - u_2$ is given by $\{\delta : \Lambda(\delta) \le \chi^2_{(1),q}\}$. Recall from (5.4.3) that $u_1 - u_2$ is also the difference $y_{1p} - y_{2p}$ for $0 < p < 1$, when $b_1 = b_2$.

If $b_1 \neq b_2$, then confidence intervals for $y_{1p} - y_{2p}$ can be obtained through the likelihood ratio statistic for tests of hypotheses $H: y_{1p} - y_{2p} = \Delta$, which is equivalent to $H: u_1 - u_2 = w_p(b_2 - b_1) + \Delta$, where $w_p = S_0^{-1}(1-p)$, with $S_0(z)$ as in (5.4.2). The likelihood ratio statistic $\Lambda(\Delta) = 2\ell(\hat{u}_1, \hat{u}_2, \hat{b}_1, \hat{b}_2) - 2\ell(\tilde{u}_1, \tilde{u}_2, \tilde{b}_1, \tilde{b}_2)$ requires that we maximize $\ell(u_2 + \Delta + w_p(b_2 - b_1), u_2, b_1, b_2)$ to obtain $\tilde{u}_2, \tilde{b}_1, \tilde{b}_2$ and $\tilde{u}_1 = \tilde{u}_2 + \Delta + w_p(\tilde{b}_2 - \tilde{b}_1)$. This is easily handled with general-purpose optimization software.

When the samples from the m distributions are all either complete or Type 2 censored, then it is in principle possible to develop exact tests, for example, of $b_1 = b_2$ or $u_1 = u_2$ (assuming $b_1 = b_2$), by using the results of Section 5.1.2 and Appendix E. Except for the case of normally distributed complete data, which lead to well-known F and t tests, these procedures are not analytically tractable. It is possible, however, to obtain p-values to a close approximation by simulation. Sections 5.4.2 and 5.4.3 consider this further.

5.4.2 Comparison of Weibull or Extreme Value Distributions

Weibull distributions with survivor functions

$$S_j(t) = \exp[-(t/\alpha_j)^{\beta_j}] \qquad j = 1, \ldots, m \tag{5.4.11}$$

can be compared using the procedures of Section 5.4.1, with the log-likelihoods $\ell_j(u_j, b_j)$ for $u_j = \log \alpha_j$, $b_j = \beta_j^{-1}$ of the form (5.2.4) in Section 5.2.1. Illustrations are presented in the examples that follow, but first we mention some additional simple methods for complete or Type 2 censored samples.

To test $H: b_1 = b_2$ ($\beta_1 = \beta_2$) or obtain confidence intervals for b_1/b_2, we could use the fact that $Z_{2j} = \hat{b}_j/b_j$ ($j = 1, 2$) are pivotal quantities; see Section 5.2.2. Thus

$$W_1 = \frac{Z_{22}}{Z_{21}} = \left(\frac{b_1}{b_2}\right)\frac{\hat{b}_2}{\hat{b}_1} \tag{5.4.12}$$

is also a pivotal quantity, and confidence intervals for b_1/b_2 can be found from values ℓ_1, ℓ_2 such that $Pr(\ell_1 \le W_1 \le \ell_2) = q$; this gives $[\ell_1(\hat{b}_1/\hat{b}_2), \ell_2(\hat{b}_1/\hat{b}_2)]$ as the confidence interval. The values of ℓ_1 and ℓ_2 can be obtained by simulation along the same lines described in Section 5.2.2; some tables generated by simulation are provided by Thoman and Bain (1969) and McCool (1974). Tests of $H: b_1 = b_2$ can likewise be based on W_1 with $b_1 = b_2$, values that are too small or too large providing evidence against H.

An approach that avoids simulation and is satisfactory in virtually all situations is to use the chi-squared approximation (5.2.17), which gives $g_j\hat{b}_j/b_j \sim \chi^2_{(h_j)}$ for constants g_j, h_j, which can be determined from Table 5.3. This implies the approximation

$$\frac{g_2 h_1}{g_1 h_2} W_1 \sim F_{(h_2, h_1)}. \tag{5.4.13}$$

If $b_1 = b_2$, then it is also possible to obtain exact confidence intervals or tests for $u_1 - u_2$ or α_1/α_2 through the pivotal quantity

$$W_2 = \frac{(\tilde{u}_1 - \tilde{u}_2) - (u_1 - u_2)}{\tilde{b}}, \tag{5.4.14}$$

where $\tilde{u}_1$, $\tilde{u}_2$, $\tilde{b} = \tilde{b}_1 = \tilde{b}_2$ are the m.l.e.'s under the restriction $b_1 = b_2$. The distribution of W_2 is intractable, but can be determined for specified values r_1, n_1, r_2, n_2 by simulation; Schafer and Sheffield (1976) provide some tables generated this way.

The exact intervals just discussed are unconditional, in the sense of Section 5.2.2. Conditional procedures are based on the distributions of pivotals conditional on the observed values of ancillary statistics; these are rather intractable and not pursued here. As discussed in Section 5.2.2, the conditional and unconditional approaches generally give close to the same results, except possibly for very small samples.

Example 5.4.1. Example 1.1.5 introduced some data on the time to breakdown of electrical insulating fluid subject to a constant voltage stress in a life test experiment. Specimens of insulating fluid were tested at seven voltage levels v_j ($j = 1, \ldots, 7$), with samples sizes ranging from 3 to 19. The data are given in Table 1.1.

A model suggested by engineering considerations is that, for a fixed voltage level, time, T, to breakdown has a Weibull distribution, and that the Weibull shape parameters β_j are the same for different voltage levels. These assumptions can be assessed informally through Weibull probability plots; if they are true, plots (see Section 3.3.1) of the samples should be roughly linear and parallel. Example 6.2.2 in Section 6.2.2 considers this. We consider here the assumption of equal β_j through a hypothesis test.

Table 5.5 shows values of the m.l.e.'s $\hat{b}_j = \hat{\beta}_j^{-1}$, $\hat{u}_j$, and $\hat{\alpha}_j = \exp(\hat{u}_j)$ for each sample. The $\hat{b}_j$ are not too different, but the $\hat{u}_j$ and $\hat{\alpha}_j$, which represent the .632 quantiles of Y and T, respectively, differ considerably. To test the hypothesis $H: b_1 = \cdots = b_7$, we consider the likelihood ratio test based on (5.4.7); for this we need the values $\hat{\ell}_j = \ell_j(\hat{u}_j, \hat{b}_j)$ and the m.l.e.'s $\tilde{u}_1, \ldots, \tilde{u}_7, \tilde{b}$ under H, and they are also given in Table 5.5. The estimates $\tilde{u}_1, \ldots, \tilde{u}_7, \tilde{b}$ can be obtained using Weibull–extreme value regression software or by direct maximization of

Table 5.5. Estimates from Voltage Breakdown Data

Sample (kV)	n_j	$\hat{b}_j$	$\hat{u}_j$	$\hat{\alpha}_j$	$\tilde{u}_j$	$\hat{\ell}_j$
26	3	1.834	6.86	1177.3	7.07	−23.72
28	5	1.022	5.87	321.4	5.77	−34.38
30	11	.944	4.35	69.5	4.24	−58.58
32	15	1.781	3.26	34.3	3.53	−65.74
34	19	1.297	2.50	12.5	2.53	−68.39
36	15	1.124	1.46	4.05	1.40	−37.69
38	8	.734	.001	.87	−.14	−6.76

Under $H: b_1 = \cdots = b_7$, $\tilde{b} = 1.251$, $\tilde{\ell} = -299.65$.

$\ell(u_1, \ldots, u_7, b, \ldots, b)$, where

$$\ell(u_1, \ldots, u_7, b_1, \ldots, b_7) = \sum_{j=1}^{7} \ell_j(u_j, b_j),$$

with $\ell_j(u_j, b_j)$ the extreme value log-likelihood (5.2.4) for sample j.

The values $\ell(\hat{u}_1, \ldots, \hat{u}_7, \hat{b}_1, \ldots, \hat{b}_7) = \sum \hat{\ell}_j = -295.26$ and $\ell(\tilde{u}_1, \ldots, \tilde{u}_7, \tilde{b}, \ldots, \tilde{b}) = -299.65$ give the observed value of (5.4.7) as $\Lambda = 8.78$. Treating Λ as $\chi^2_{(6)}$ under H, we get a p-value of $Pr(\chi^2_{(6)} \geq 8.78) = .187$, which does not provide evidence of a difference in values for $b_1, \ldots, b_7$ or $\beta_1, \ldots, \beta_7$.

Example 5.4.2. The data below are the voltage levels at which failures occurred in two types of electrical cable insulation when specimens were subjected to an increasing voltage stress in a laboratory test. Twenty specimens of each type were involved, and the failure voltages (in kilovolts per millimeter) were, in increasing order, as follows.

Type I Insulation	32.0, 35.4, 36.2, 39.8, 41.2, 43.3, 45.5, 46.0, 46.2, 46.4, 46.5, 46.8, 47.3, 47.3, 47.6, 49.2, 50.4, 50.9, 52.4, 56.3.
Type II Insulation	39.4, 45.3, 49.2, 49.4, 51.3, 52.0, 53.2, 53.2, 54.9, 55.5, 57.1, 57.2, 57.5, 59.2, 61.0, 62.4, 63.8, 64.3, 67.3, 67.7.

Engineering experience (Stone and Lawless 1979) suggests that failure voltages for the two types of cable are adequately represented by Weibull distributions (5.4.11) with a common shape parameter β. Weibull probability plots (Section 3.3.1) of the two samples are shown in Figure 5.3, and the fact that the points lie roughly along two parallel lines indicate that these assumptions are reasonable.

The m.l.e.'s for the extreme value parameters (u_j, b_j), $j = 1, 2$, are shown below, with standard errors in parentheses:

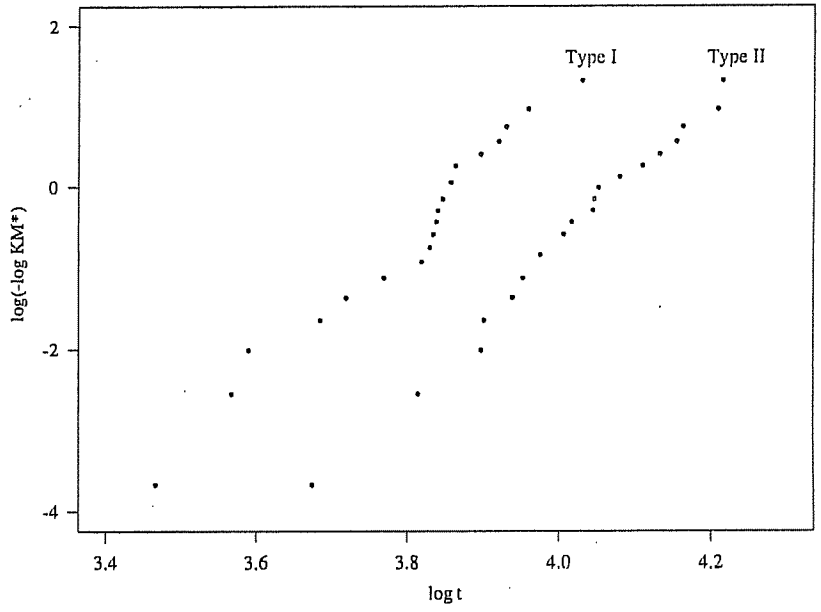

Figure 5.3. Weibull probability plots of cable-insulation failure voltages.

Type I: $\hat{u}_1 = 3.867(.025)$ $\hat{b}_1 = .107(.018)$

Type II: $\hat{u}_2 = 4.080(.026)$ $\hat{b}_2 = .109(.019)$.

There is clearly no evidence of a difference in the values of b_1 and b_2 (or of β_1 and β_2).

Under the assumption that $b_1 = b_2$, the difference $\delta = u_1 - u_2$ equals the difference $y_{1p} - y_{2p}$ in quantiles of the two log-lifetime distributions, so let us get a confidence interval for it. The likelihood ratio procedure based on the statistic $\Lambda(\delta)$ of (5.4.10) gives an approximate .90 interval through the determination of values such that $\Lambda(\delta) \leq \chi^2_{(1),.90} = 2.706$. Values $\Lambda(\delta)$ for a specified value δ are given by software for Weibull and extreme value regression models, which are discussed in Chapter 6. This approach gives the interval $-.272 \leq \delta \leq -.157$, which yields a .90 confidence interval for $\exp(u_1 - u_2) = \alpha_1/\alpha_2$ of $.762 \leq \alpha_1/\alpha_2 \leq .855$. Note that this is also a confidence interval for the ratio t_{1p}/t_{2p} of the quantiles for the two failure distributions. A slightly simpler but perhaps less accurate procedure is to use the approximate standard normal pivotal quantity

$$W = \frac{(\hat{u}_1 - \hat{u}_2) - \delta}{[se(\hat{u}_1)^2 + se(\hat{u}_2)^2]^{1/2}} \tag{5.4.15}$$

to obtain confidence intervals. The approximation $Pr(-1.645 \leq W \leq 1.645) = .90$ yields the .90 confidence interval $-.272 \leq \delta \leq -.154$, and the interval $.762 \leq$

$\alpha_1/\alpha_2 \leq .857$, virtually identical in this case to the likelihood ratio-based intervals. "Exact" intervals could also be obtained using the pivotal quantity (5.4.14), since the samples are uncensored here. Simulation or use of tables of Schafer and Sheffield (1976) produces intervals close to those given by the approximations above.

5.4.3 Comparison of Log-Normal or Log-Logistic Distributions

The comparison of parameters in log-normal and log-logistic models can be carried out in the same ways as for the Weibull distribution. In the case of log-normal distributions there are well-known methods based on normal models that can be applied when the data are uncensored. In particular, if $y_{11}, \ldots, y_{1n_1}$ and $y_{21}, \ldots, y_{2n_2}$ are independent random samples from two normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, and if $\bar{y}_1, s_1^2, \bar{y}_2, s_2^2$ are the sample means and variances, then confidence intervals for σ_1/σ_2 or tests of $H: \sigma_1 = \sigma_2$ can be based on the pivotal quantity

$$\left(\frac{\sigma_2}{\sigma_1}\right)^2 \left(\frac{s_1}{s_2}\right)^2 \sim F_{(n_1-1, n_2-1)}. \tag{5.4.16}$$

If $\sigma_1 = \sigma_2$, then confidence intervals or tests concerning $\delta = \mu_1 - \mu_2$ can be based on the pivotal quantity

$$\frac{(\bar{y}_1 - \bar{y}_2) - \delta}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n_1+n_2-2)}, \tag{5.4.17}$$

where $s^2 = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]/(n_1 + n_2 - 2)$ estimates the common value of σ^2.

The following example compares large-sample and exact intervals for δ, the data being uncensored.

Example 5.4.3. (**Example 5.4.2 revisited**). The cable insulation failure voltages in Example 5.4.2 are consistent with both log-normal and Weibull models. Let us consider a comparison of quantiles for Types I and II insulation under a log-normal model.

There is no censoring and we find the following estimates of the parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ under the assumption that log failure voltages have distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ for Types I and II insulation:

$$\hat{\mu}_1 = \bar{y}_1 = 3.805 \qquad \hat{\sigma}_1 = .136 \qquad (s_1^2 = .01936)$$
$$\hat{\mu}_2 = \bar{y}_2 = 4.018 \qquad \hat{\sigma}_2 = .132 \qquad (s_2^2 = .01833).$$

The sample variances s_j^2 are related to the m.l.e.'s $\hat{\sigma}_j$ by $\hat{\sigma}_j = (19s_j^2/20)^{1/2}$.

There is clearly no evidence of a difference in the standard deviations σ_1, σ_2, so we obtain a .90 confidence interval for $\delta = \mu_1 - \mu_2$ based on the exact pivotal quantity (5.4.17). We find that $\hat{\mu}_1 - \hat{\mu}_2 = -.213$, $s = .137$, and $Pr(-1.686 \leq t_{(38)} \leq 1.686) = .90$, giving a .90 confidence interval as $-.213 \pm 1.686(.1373)(.1)^{1/2}$, or $-.286 \leq \delta \leq -.140$. The large-sample procedure based on the approximate pivotal quantity (5.4.15) yields the approximate .90 confidence interval $-.213 \pm (1.645)(\hat{\sigma}_1^2/20 + \hat{\sigma}_2^2/20)^{1/2}$, or $-.283 \leq \delta \leq -.143$, which is in close agreement with the exact interval.

Recall from Example 5.4.2 that the extreme value model gave the confidence interval $-.272 \leq \delta \leq -.154$. One would expect the estimates to be similar, given that both distributions fit the data well.

5.5 MODELS WITH ADDITIONAL SHAPE PARAMETERS

5.5.1 Introduction

As discussed in Section 1.3.6, location-scale models for which the baseline or error distribution involves one or more shape parameters are sometimes useful. Besides additional flexibility for fitting data, such models can provide comparisons of Weibull, log-normal, and log-logistic models, and can be used to examine the robustness of conclusions to plausible variations in the model.

For simplicity, and because it is usually sufficient for practical applications, we consider models that have a single shape parameter k. In this case, the distribution (5.1.1) for $Y = \log T$ takes the form

$$S(y; u, b, k) = S_0\left(\frac{y-u}{b}; k\right), \qquad -\infty < y < \infty, \tag{5.5.1}$$

where $S_0(z; k)$ is a survivor function on $(-\infty, \infty)$ for k in some set of allowable values. The corresponding survivor function for lifetime T is

$$S(t; \alpha, \beta, k) = S_0^*[(t/\alpha)^\beta; k] \qquad t > 0, \tag{5.5.2}$$

where $S_0^*(w; k) = S_0(\log w; k)$, $\alpha = \exp(u)$, and $\beta = b^{-1}$.

Two useful models are the log-Burr and generalized log-gamma families represented by (1.3.20) and (1.3.22); the former includes the extreme value and logistic distributions, and the latter the extreme value and normal distributions. We consider them in Sections 5.5.2 and 5.5.3, but first discuss some general points.

Maximum likelihood estimates $\hat{u}, \hat{b}, \hat{k}$ from censored data for a model (5.5.1) are conveniently found by first maximizing the log-likelihood $\ell(u, b, k)$, with k held fixed at various values, to obtain estimates $\tilde{u}(k), \tilde{b}(k)$, and the profile log-likelihood function $\ell_p(k) = \ell(\tilde{u}(k), \tilde{b}(k), k)$. The profile $\ell_p(k)$ often has flat regions, and in some models can have more than one stationary point, but it is generally easy to obtain the m.l.e. $\hat{k}$ (which maximizes $\ell_p(k)$) and $\hat{u} = \tilde{u}(\hat{k})$, $\hat{b} = \tilde{b}(\hat{k})$. The profile log-likelihood function, relative likelihood function, or equivalent likelihood ratio statistic

$$\Lambda(k) = 2\ell(\hat{u}, \hat{b}, \hat{k}) - 2\ell(\tilde{u}(k), \tilde{b}(k), k) \tag{5.5.3}$$

can be used for an assessment of plausible k-values. For interior values k_0 in the parameter space the statistic $\Lambda(k_0)$ is asymptotically $\chi^2_{(1)}$ when $k = k_0$.

To obtain interval estimates or tests about parameters, one possibility is to treat all three parameters u, b, k as unknown and employ large-sample procedures based on likelihood ratio statistics or normal approximations for $(\hat{u}, \hat{b}, \hat{k})$. This approach requires care, because k is often imprecisely estimated and normal approximations can be poor unless the number of failures is large. An alternative procedure is to make inferences with k held fixed, but to vary k across a plausible range of values in order to see its effect. This has the advantage that large-sample methods with k fixed tend to be quite accurate, as described for extreme value, normal, and logistic models earlier in this chapter. In the case of complete or Type 2 censored data, exact methods are also available. Moreover, with k fixed there are straightforward interpretations of the parameters u and b in any model (5.5.1). The pth quantile of Y is

$$y_p(k) = u + w_p(k)b, \tag{5.5.4}$$

where $w_p(k) = S_0^{-1}(1 - p; k)$ is the pth quantile of $S_0(z; k)$, and u corresponds to the quantile for which $w_p(k) = 0$.

In some situations the parameters u, b, and k may all have meaningful physical interpretations; see, for example, Problem 1.14 concerning a derivation of the generalized Burr and log-Burr models. In this case there is a stronger argument for formally treating k as unknown in the construction of tests or estimates.

5.5.2 The Generalized Log-Burr Distribution

The generalized log-Burr model (1.3.20) has s.f. of the form (5.5.1) with

$$S_0(z; k) = \left(1 + \frac{e^z}{k}\right)^{-k} \qquad -\infty < z < \infty \tag{5.5.5}$$

and corresponding p.d.f.

$$f_0(z; k) = e^z \left(1 + \frac{e^z}{k}\right)^{-k-1} \qquad -\infty < z < \infty. \tag{5.5.6}$$

It includes the logistic distribution ($k = 1$) and extreme value distribution (limit as $k \to \infty$) as special cases. The log-likelihood function for (u, b, k) based on a censored random sample (t_i, δ_i), $i = 1, \ldots, n$ takes the form

$$\ell(u, b, k) = -r \log b + \sum_{i=1}^{n} \delta_i \log f_0(z_i; k) + \sum_{i=1}^{n} (1 - \delta_i) \log S_0(z_i; k), \tag{5.5.7}$$

where $z_i = (y_i - u)/b$, $y_i = \log t_i$ and $r = \sum \delta_i$.

It can be useful to parameterize the model in terms of $\lambda = k^{-1}$; the logistic and extreme value models then correspond to $\lambda = 1$ and 0. The log-likelihood (5.5.7) can be maximized for fixed λ or k, giving $\tilde{u}(k)$, $\tilde{b}(k)$, and the profile log-likelihood

value $\ell_p(k) = \ell(\tilde{u}(k), \tilde{b}(k), k)$. The profile log-likelihood can have more than one stationary point. From empirical studies (Gould 1986) it appears always to have a local minimum at some value k^*, below which it is strictly increasing. The model degenerates to an improper uniform distribution as k approaches 0, and in practice we restrict attention to values above some minimal $k_0 > 0$. For values $k \geq k_0$, the maximum may occur at $k = \infty$ or a finite k. The likelihood ratio statistic $\Lambda(k)$ of (5.5.3) identifies plausible values and for finite k can be used for tests or confidence intervals by reference to the usual $\chi^2_{(1)}$ limiting distribution. The extreme value case $k = \infty$ ($\lambda = 0$) is a boundary point and under $H: k = \infty$, the likelihood ratio statistic $\Lambda(\infty)$ has a limiting distribution with $Pr(\Lambda(\infty) \leq a) = .5 + .5Pr(\chi^2_{(1)} \leq a)$.

Example 5.5.1. (Example 5.3.1 revisited). Failure times from a life test involving 96 locomotive controls were examined in Example 5.3.1, where log-normal and log-logistic distributions were found to describe the data well. Although the sample is fairly large, 59 of the 96 failure times were censored at $t = 135.0$ (thousand miles), and no discrimination between the two models was possible; the log-normal's maximized log-likelihood was only .14 larger than the log-logistic's. We will now fit a log-Burr model, showing that the Weibull model is also consistent with the data.

The log-likelihood function (5.5.7) is easily maximized with standard software when k is a fixed value. Table 5.6 shows the m.l.e.'s $\tilde{u}(k)$, $\tilde{b}(k)$, and values $\ell_p(k) = \ell(\tilde{u}(k), \tilde{b}(k), k)$ for several k values. The profile log-likelihood function $\ell_p(k)$ has a local maximum at $\hat{k} = .598$, and Table 5.6 also shows the likelihood ratio statistic $\Lambda(k) = 2\ell_p(\hat{k}) - 2\ell_p(k)$. Figure 5.4 shows $\Lambda(k)$ and also the same statistic as a function of $\lambda = k^{-1}$. We see that $\Lambda(k)$ increases very little for $k > \hat{k}$, with both the log-logistic ($k = 1$) and Weibull ($k = \infty$) failure time models being very plausible. The profile $\ell_p(k)$ increases rapidly for $k < .1$; in Figure 5.4 the plot is truncated at $k = .1$.

The parameters u and b have different interpretations for models with different k values, but alternative plausible models give very similar estimates of survival probabilities or quantiles in the lower half of the distribution. For example, for the logistic model ($k = 1$) the estimate of the median for Y is $\hat{y}_{.50} = \hat{u} = 5.083$, and for the Weibull ($k = \infty$) it is $\hat{y}_{.50} = \hat{u} - .3665\hat{b} = 5.054$.

Table 5.6. m.l.e.'s and Likelihood Values (Log-Burr Models)

k	$\tilde{u}(k)$	$\tilde{b}(k)$	$\ell_p(k)$	$\Lambda(k)$
0.1	4.516	.195	−75.47	3.79
0.5	4.978	.346	−73.58	.01
0.598	5.009	.358	−73.57	.00
1.0	5.083	.384	−73.60	.05
10.0	5.198	.424	−73.73	.31
∞	5.212	.429	−73.75	.35

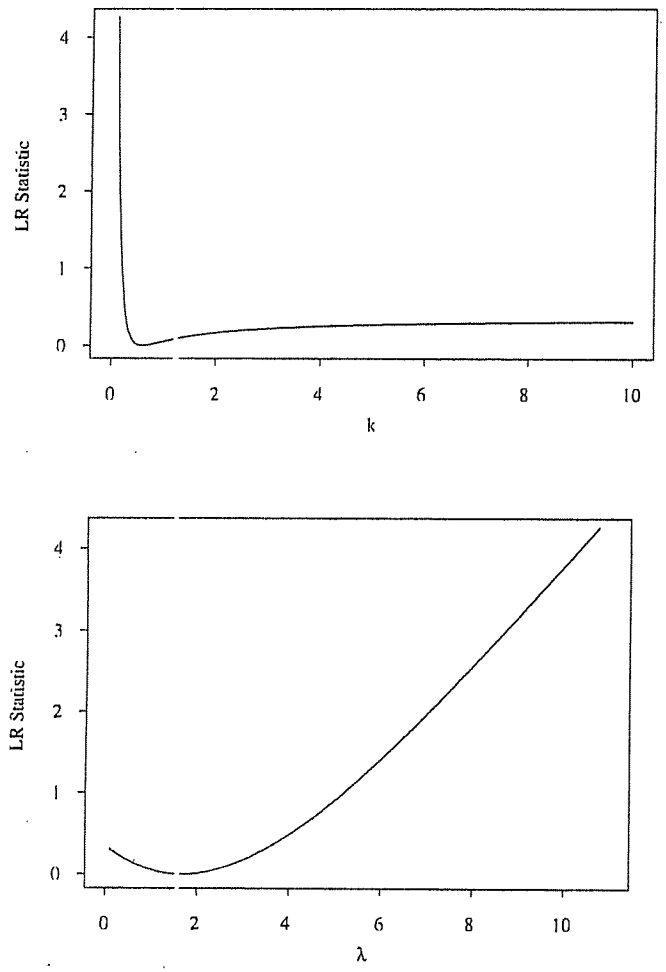

Figure 5.4. Likelihood ratio statistics $\Lambda(k)$ and $\Lambda(\lambda)$ for log-Burr model (locomotive controls data).

Plots portraying confidence intervals with different confidence coefficients, under alternative models or methods of estimation are often useful. An example is Figure 4.1 of Example 4.1.2, where confidence intervals for an exponential mean are shown. Such plots are sometimes termed confidence distribution plots. Log-Burr models with a wide range of k values fit the data well in the present example, so let us demonstrate the effect of model choice on estimation of the .10 quantiles $t_{.10}$ or $y_{.10}$. The .10 quantile for $Y = \log T$ is, for given k,

$$y_{.10}(k) = u + b \log[\log k + \log(.9^{-1/k} - 1)].$$

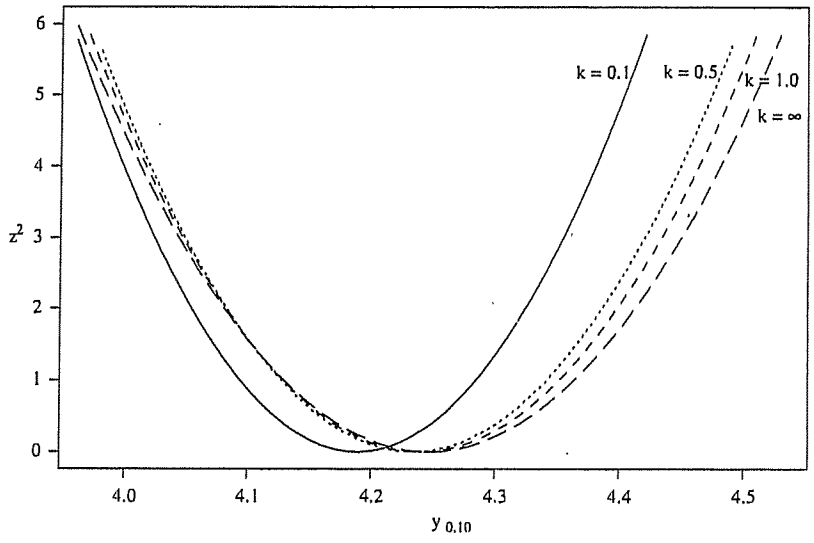

Figure 5.5. Confidence distribution plots for $y_{.10}$ (locomotive controls data).

Confidence intervals, treating k as fixed, may be based on the approximate $N(0, 1)$ pivotal quantity (5.1.13),

$$Z = \frac{\hat{y}_{.10}(k) - y_{.10}(k)}{se(\hat{y}_{.10}(k))}.$$

A two-sided approximate q confidence interval consists of values $y_{.10}(k)$ that satisfy $Z^2 \le \chi^2_{(1),q}$. Figure 5.5 shows a plot of Z^2 versus $y_{.10}(k)$ for $k = .1, .5, 1.0$, and ∞; note from Table 5.6 that all four values are within a .95 confidence interval for k, though .1 is barely so. Approximate confidence intervals with any confidence coefficient are evident from the plot; for example, $Z^2 \le 2.706$ and $Z^2 \le 3.84$ give .90 and .95 intervals, respectively. The three models with $k = .5, 1.0$, and ∞ give similar lower confidence limits, but upper limits that are a little different, relative to the widths of the confidence intervals. The model with $k = .1$ gives a quite different upper limit.

5.5.3 The Generalized Log-Gamma Distribution

The generalized log-gamma distribution (1.3.22) has s.f. of the form (5.5.1) for $0 < k < \infty$, with

$$S_0(z; k) = 1 - I[k, k \exp(zk^{-1/2})], \qquad -\infty < z < \infty, \qquad (5.5.8)$$

where

$$I(k; x) = \frac{1}{\Gamma(k)} \int_0^x u^{k-1} e^{-u}\, du \tag{5.5.9}$$

is the incomplete gamma integral defined in (1.3.16). In the limit as $k \to \infty$, $S_0(z; k)$ approaches the survivor function for the standard normal distribution, $1 - \Phi(z)$. The p.d.f. of $Z = (Y - u)/b$ for $0 < k < \infty$ is

$$f_0(z; k) = \frac{k^{k-.5}}{\Gamma(k)} \exp[k^{.5}z - k\exp(zk^{-.5})], \qquad -\infty < z < \infty, \tag{5.5.10}$$

and as $k \to \infty$, this approaches the standard normal p.d.f. $\phi(z)$.

As discussed in Section 1 3.6, the cases $k = 1$ and $k = \infty$ give the extreme value and normal distributions, corresponding to Weibull and log-normal distributions for T. The p.d.f. $f_0(z; k)$ changes relatively little as k increases from 1 to ∞, and unless a large amount of data is available, it is not usually possible to estimate k very precisely. For many data sets both Weibull and log-normal models provide satisfactory fits; see Example 3.3.1 and Example 5.5.2 below.

It is frequently useful to parameterize the model (5.5.10) in terms of $\lambda = k^{-1/2}$ instead of k. The normal and extreme value distributions then correspond to $\lambda = 0$ and 1. Prentice (1974) showed that models with $\lambda < 0$ are also allowed; distributions with $\lambda < 0$ correspond to situations where $-Z$ has p.d.f. (5.5.10) with $k = |\lambda|^{-2}$. Extending the model in this way not only provides a wider class but also alleviates some technical difficulties that arise in (5.5.10) because the important case $\lambda = 0$ is a boundary point.

The log-likelihood function for (u, b, k) based on a censored random sample $(t_i; \delta_i)$, $i = 1, \ldots, n$ takes the form

$$\ell(u, b, k) = -r \log b + \sum_{i=1}^{n} \delta_i \log f_0(z_i; k) + \sum_{i=1}^{n} (1 - \delta_i) \log S_0(z_i; k), \tag{5.5.11}$$

where $z_i = (y_i - u)/b$, $y_i = \log t_i$, and $r = \sum \delta_i$. It is straightforward to maximize (5.5.11) with k fixed to obtain $\tilde{u}(k)$, $\tilde{b}(k)$, and the profile log-likelihood function for k. For the values $k = 0$ and ∞, software for extreme value–Weibull and normal-log-normal models can be used: general optimization software handles other k values quite easily. For $0 < k < \infty$ the functions $f_0(z; k)$ and $S_0(z; k)$ are given by (5.5.10) and (5.5.8), respectively. The information matrix $I_k(u, b)$ for (u, b) with k fixed can be obtained by numerical differentiation of $\ell(u, b, k)$, which is available in much optimization software, or from analytic expressions that are messy but straightforward to obtain.

It is usually sufficient to consider about 8–10 different k values in obtaining the profile log-likelihood $\ell(\tilde{u}(k), \tilde{b}(k), k) = \ell_p(k)$ and a few more to locate the m.l.e. $\hat{k}$

precisely. The parameter $\lambda = k^{-1/2}$ is useful for computation and portrayal of $\ell_p(k)$. Values of λ close to zero are common in lifetime applications, and it is possible to have $\hat{\lambda} = 0$ ($\hat{k} = \infty$). This happens when $\hat{\lambda}$ in the extended model, which allows $\lambda < 0$, is less than zero. In situations where $\lambda = 0$ is a plausible value, it is a good idea to consider the extended model rather than just (5.5.10).

Tests of the hypothesis $H: k = k_0$ can be based on the likelihood ratio statistic $\Lambda(k_0)$ of (5.5.3). For finite k_0 the distribution of $\Lambda(k_0)$ is asymptotically $\chi^2_{(1)}$ under H. A slight technical difficulty arises in testing the normal-log-normal model, since $k_0 = \infty$ ($\lambda_0 = 0$) is on the boundary of the parameter space. This can be overcome by working with the extended family, in which case $\lambda = 0$ becomes an interior point and the likelihood ratio statistic is asymptotically $\chi^2_{(1)}$ at $\lambda = 0$. If the parameter space $0 < k < \infty$ ($\lambda \geq 0$) is retained, then $\Lambda(k_0)$ has an asymptotic distribution when $k_0 = \infty$ ($\lambda_0 = 0$) such that for $a \geq 0$, $Pr(\Lambda(\infty) \leq a) = .5 + .5Pr(\chi^2_{(1)} \leq a)$.

The generalized log-gamma family provides tests of the Weibull and log-normal models against parametric alternatives. It can also be used to examine the effect of departures from an assumed Weibull, log-normal or other model on inferences or to conduct sensitivity analyses of model choice. Example 5.5.2 illustrates this.

5.5.3.1 *Exact Methods for Uncensored Samples*

If k is assumed known in (5.5.10), then, as discussed in Section 5.1.2, exact inference procedures based on pivotal quantities

$$Z_1 = \frac{\hat{u} - u}{\hat{b}}, \qquad Z_2 = \frac{\hat{b}}{b}, \qquad Z_p = \frac{\hat{u} - y_p}{\hat{b}} \tag{5.5.12}$$

are available when the data are Type 2 censored. The distributions of Z_1, Z_2, and Z_p are analytically intractable, but can be obtained to a desired degree of accuracy by simulation. An alternative approach is to consider the conditional distributions of the pivotals, given the observed values of the ancillary statistic $\mathbf{a}$ in Theorem 5.1.1. Lawless (1980) showed that when the data are uncensored, this approach is computationally feasible, and we summarize the relevant results here.

In (5.5.10) the m.l.e.'s with k held fixed are $\hat{u} = \tilde{u}(k)$ and $\hat{b} = \tilde{b}(k)$; similarly, $y_p = y_p(k) = u + w_p(k)b$, as in Section 5.5.1. The ancillary statistics are $a_i = (y_i - \hat{u})/\hat{b}$, where $y_1, \ldots, y_n$ is the uncensored sample of log-lifetimes, and where for notational convenience we write $\hat{u}$, $\hat{b}$ for $\tilde{u}(k)$, $\tilde{b}(k)$. The distributions of Z_2 and Z_p given $\mathbf{a} = (a_1, \ldots, a_n)$ can be obtained from Theorem E3 of Appendix E and are as follows:

1. The marginal p.d.f. of Z_2, given $\mathbf{a}$, is

$$g_2(z|\mathbf{a}; k) = \frac{C(\mathbf{a}, n, k) z^{n-2} \exp[k^{1/2}(z - 1)\sum_{i=1}^n a_i]}{\left[\frac{1}{n}\sum_{i=1}^n \exp(a_i z k^{-1/2})\right]^n}, \qquad z > 0. \tag{5.5.13}$$

2. The marginal distribution function of Z_p, given $\mathbf{a}$, is

$$Pr(Z_p \le t|\mathbf{a}; k) = \int_0^\infty g_2(z|\mathbf{a}; k) I\left(nk, ke^{k^{-1/2}(tz+w)} \sum_{i=1}^{n} e^{k^{-1/2}a_i z}\right) dz, \tag{5.5.14}$$

where $I(k, x)$ is given by (5.5.9) and we write w for $w_k(p)$.

Both (5.5.13) and the integral in (5.5.14) can be integrated numerically; the normalizing constant $C(\mathbf{a}, n, k)$ can be determined from the fact that (5.5.13) integrates to one. The integrals involved are of the same type as those encountered in Section 5.2.2.1. For the special case $k = \infty$ the distributions of Z_2 and Z_p are given by well-known results for the normal distribution, as discussed in Section 5.3.1.

Example 5.5.2. (Example 5.3.2 revisited). Let us reconsider the data on failure times for 23 ball bearings in a life test, which were examined using a log-normal distribution in Example 5.3.2, but which earlier probability plots suggested were consistent with either a log-normal or Weibull model. Table 5.7 shows maximum likelihood estimates $\tilde{u}(k)$, $\tilde{b}(k)$, and profile likelihood values $\ell_p(k) = \ell(\tilde{u}(k), \tilde{b}(k), k)$, based on maximizing (5.5.11) with k fixed at the values shown. The m.l.e. $\hat{k} = 10.6$ ($\hat{\lambda} = .31$) is easily located; this gives $\hat{u} = 4.23$, $\hat{b} = .510$. The table also gives the profile relative likelihood function, $R_p(k) = L(\tilde{u}(k), \tilde{b}(k), k)/L(\hat{u}, \hat{b}, \hat{k})$. As discussed in Example 4.5.1, a plot of the function $R_p(k)$ is a nice alternative to a plot of $\Lambda(k)$ or of $\ell_p(k)$, since it shows the likelihood function directly.

The values $k = \infty$ ($\lambda = 0$) and $k = 1$ ($\lambda = 1$) are both highly plausible; with likelihood ratio statistic values $\Lambda(\infty) = .35$ and $\Lambda(1) = 1.45$, there is clearly no evidence to contradict either the log-normal or Weibull model. An approximate .95 confidence interval for k, based on values such that $\Lambda(k) \le \chi^2_{(1),.95} = 3.84$, consists of all k values greater than about .40. There are obviously members of the extended family of models with $\lambda < 0$ that are also very well supported by the data, though parameter and likelihood values have not been shown for these models.

To examine the effects of model choice on inferences about distribution characteristics, we determine confidence intervals for the quantiles $y_{.01}$, $y_{.10}$, and $y_{.50}$ of the

Table 5.7. m.l.e.'s and Likelihood Values (Log-Gamma Models)

k	$\lambda = k^{-1/2}$	$\tilde{u}(k)$	$\tilde{b}(k)$	$\ell_p(k)$	$R_p(k)$
.3	1.826	4.604	.417	−20.14	.072
.5	1.414	4.507	.449	−19.14	.216
1.0	1.000	4.405	.476	−18.23	.485
4.0	.500	4.279	.502	−17.57	.942
9.0	.333	4.237	.509	−17.509	.999
10.6	.307	4.230	.510	−17.5075	1.000
12.0	.289	4.225	.511	−17.508	1.000
400	.050	4.164	.520	−17.62	.895
∞	0.0	4.150	.522	−17.68	.839

log-lifetime distribution for several values of k. Because exact confidence intervals based on the pivotals Z_p in (5.5.12) are available here by working with (5.5.14), we give them. Confidence intervals for the normal distribution case ($k = \infty$) were considered previously in Example 5.3.2. Note that with a given value k, the pth quantile is $y_p(k) = u + w_p(k)b$, where $w_p(k)$ is the pth quantile of the standard distribution (5.5.10), and satisfies

$$I[k, k\exp(w_p(k)k^{-1/2})] = p.$$

The incomplete gamma integral $I(k, x)$, given by (5.5.9), is the distribution function for a scaled chi-squared or single-parameter gamma model (1.3.17), and quantiles are provided in many software packages. In particular, let $x_p(k)$ be the gamma pth quantile, satisfying $I(k, x) = p$. Then $k\exp[w_p(k)k^{-1/2}] = x_p(k)$, or

$$w_p(k) = k^{1/2}\log(x_p(k)/k).$$

For example, we find with $k = 10$, $p = .50$ that $x_{.5}(10) = 9.667$, and so $w_{50}(10) = -.107$.

Table 5.8 shows m.l.e.'s and .90 intervals (lower confidence limit, LCL; upper confidence limit, UCL) for $k = .5, 1, 10, \infty$. Lawless (1980) describes the calculation of the intervals in some detail. As would be expected, there is good agreement in the confidence intervals for the median $y_{.50}$ at different values of k. However, as we move to the tail of the distribution, the different models give somewhat different results. The left tail of a distribution is of much interest in lifetime applications, so this is important. For example, the .10 quantile is used in rating certain types of manufactured items. If the lifetimes in this example are assumed to come from a log-normal distribution, the lower .95 confidence limit on $y_{.10}$ is 3.15 (corresponding to a limit of 23.34 million revolutions for $t_{.10}$), whereas it is 2.83 (corresponding to a limit of 16.95 million revolutions for $t_{.10}$) if a Weibull distribution is assumed. This difference may or may not be of practical importance, but is substantial relative to the lengths of the two-sided .90 confidence intervals for $y_{.10}$ or $t_{.10}$. The differences

Table 5.8. m.l.e.'s and .90 Confidence Intervals for Quantiles

		$k = .5$	$k = 1$	$k = 10$	$k = \infty$
$y_{.50}$	UCL	4.45	4.42	4.36	4.34
	$\hat{y}_{.50}$	4.26	4.23	4.18	4.15
	LCL	4.01	4.01	3.98	3.90
$y_{.10}$	UCL	3.56	3.63	3.69	3.68
	$\hat{y}_{.10}$	3.19	3.33	3.46	3.48
	LCL	2.56	2.83	3.10	3.15
$y_{.01}$	UCL	2.36	2.71	3.12	3.20
	$\hat{y}_{.01}$	1.73	2.22	2.81	2.94
	LCL	.54	1.29	2.21	2.44

in the estimates of $y_{.01}$ are larger still, to the extent that the m.l.e. for $y_{.01}$ under the normal model is not even inside the .90 interval for $y_{.01}$ under the extreme value model.

When the effect of model choice is accounted for, the real precision with which low (or high) quantiles are estimated is often considerably less than is implied by the estimates of an individual model; this is seen clearly here. Example 5.5.1 also showed, in Figure 5.5, the dependence of confidence intervals for $y_{.10}$ on which of the log-Burr models was used for analysis.

5.6 PLANNING EXPERIMENTS OR LIFE TESTS

The selection of an experimental plan for studying lifetimes involves a consideration of costs, time, and other constraints. Some general issues were discussed in Section 2.5, and the design of plans for an underlying exponential distribution was considered in Section 4.1.4. Experimental plans for extreme value, logistic, and normal distributions have also been studied, and simulation may be used to help with one-time designs for them or other location-scale models. A brief discussion of this area follows.

Experiments or life tests can be of many forms, as discussed in Sections 2.5 and 4.1.4. In particular, they can involve staggered entry of items into the study, including in some cases the replacement of items. The large-sample inference methods discussed in this chapter are generally valid in such settings, but numerical calculation or simulation is generally needed to investigate frequency properties and thus compare plans. In addition, with a location-scale model we typically need guesses at the values of parameters u and b in order to plan a study.

Simulation is an important method of investigating plans. Given tentative parameter values, data can be generated from the distribution in question according to a specified experimental plan, and point and interval estimates of parameters of interest can be obtained. By replicating this process we can assess the precision with which parameters are estimated. This can be repeated with different experimental plans, thus allowing a comparison and guidance in the choice of a plan. Since this procedure is dependent on the values of u and b, it may be advisable to repeat it with alternative values of u and b. Time and effort can be saved in such investigations by using principles of experimental design for the simulation study itself.

An example involving simulation is considered in Section 5.6.3. Some plans involving Type 1 or 2 censoring can also be explored by direct numerical calculation; we consider this next.

5.6.1 Information Calculations Under Type 1 or Type 2 Censoring

Experiments in which n items or individuals are placed on study at the same time and followed until some common censoring time, C, are referred to as Type 1 censored; the same term applies to studies in which individuals may enter the study at different times, so that individual i has a specified potential censoring time, C_i, equal to the

available follow-up time for the individual. Under a location-scale model (5.1.1) for log-lifetime Y, it is possible to evaluate the Fisher or expected information matrix $\mathcal{I}(u, b)$ and corresponding asymptotic covariance matrix $\mathcal{I}(u, b)^{-1}$, as described in Section 5.1.1, though numerical integration is needed to do this. Let us consider this briefly.

Consider the case where all censoring times are the same: $C_i = C$. The discussion surrounding expressions (5.1.7)–(5.1.10) in Section 5.1.1 shows that the Fisher information matrix $\mathcal{I}(u, b)$ is then of the form

$$\mathcal{I}(u, b) = \frac{n}{b^2}\mathcal{I}_0(u, b). \tag{5.6.1}$$

$\mathcal{I}(u, b)$ is the 2×2 matrix whose entries are the negative expected values of (5.1.7)–(5.1.9), and since the expected values of the terms in (5.1.5) and (5.1.6) equal 0, we have the following as the entries in $\mathcal{I}_0(u, b)$:

$$\mathcal{I}_{0,11} = -E\left[\delta_i \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1 - \delta_i)\frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right] \tag{5.6.2}$$

$$\mathcal{I}_{0,12} = \mathcal{I}_{0,21} = -E\left[\delta_i z_i \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1 - \delta_i) z_i \frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right] \tag{5.6.3}$$

$$\mathcal{I}_{0,22} = -E\left[\delta_i z_i^2 \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1 - \delta_i) z_i^2 \frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right] + E(\delta_i). \tag{5.6.4}$$

These expressions can be evaluated for any specific location-scale model with the use of (5.1.10) and the comments surrounding it. Note that the values of u and b affect $\mathcal{I}_0(u, b)$ through the standardized log censoring time $R_i = R = (\log C - u)/b$. Problem 5.4 gives the resulting information matrix expressions for the extreme value model.

Escobar and Meeker (1994) provide an algorithm to compute $\mathcal{I}_0(u, b)$ and the asymptotic covariance matrix $\mathcal{I}(u, b)^{-1}$ for extreme value, normal, and logistic distributions. Several sources contain tables or figures that give asymptotic variances for estimates of quantiles y_p or other parameters. Meeker and Escobar (1998, Sec. 10.5) provide references as well as a table and charts for extreme value and normal distributions. These are discussed further in Section 5.6.3.

Type 2 censoring yields Fisher information matrices that are formally equivalent to those for Type 1 censoring, so the preceding comments can be applied to this case as well. Extensions to deal with fixed censoring times C_i, which vary across individuals, can also be made; this merely means that $\mathcal{I}(u, b)$ in (5.6.1) is a sum of different matrices $\mathcal{I}_{0i}(u, b)/b^2$, given by (5.6.2)–(5.6.4).

5.6.2 Formal Tests and Acceptance Procedures

An important problem in acceptance sampling or reliability demonstration is the development of formal hypothesis tests for quantiles or survival probabilities. For

example, consider the hypotheses

$$H_0 : y_p = y_{p0} \quad \text{versus} \quad H_1 : y_p < y_{p0}$$

for some particular location-scale distribution. With arbitrarily censored data we would typically resort to a test based on the likelihood ratio statistic $\Lambda(y_{p0})$, as in (5.1.19), or on the Wald-type approximate pivotal quantity (5.1.13). For the sake of discussion, suppose that (5.1.13) is used, so that a formal size α test of H_0 versus H_1 would seek a critical value z_0 such that

$$Pr(Z_p \le z_0; y_p = y_{p0}) = Pr\left(\frac{\hat{y}_p - y_{p0}}{se(\hat{y}_p)} \le z_0; y_p = y_{p0}\right) = \alpha. \qquad (5.6.5)$$

The usual procedure is to treat Z_p as standard normal, giving $z_0 = -N_\alpha$, where N_α is the α quantile for $N(0, 1)$.

We usually want to select the sample size and experimental plan so that the power of the test at some specified alternative value of the parameter in question is satisfactorily large. The condition $y_p = y_{p1}$ is a composite hypothesis, just as $y_p = y_{p0}$ is, since $y_p = u + w_p b$. A reasonable procedure in many instances is to assume that b is fixed and that changes in y_p are due to changes in u. One can then consider probabilities

$$Pr\left(\frac{\hat{y}_p - y_{p0}}{se(\hat{y}_p)} \le z_0; u, b\right), \qquad (5.6.6)$$

which can be closely approximated by simulation for given values of u and b. Using plausible values u, b, for which $u + w_p b = y_{p0}$, simulation provides a check on the normal approximation used to obtain z_0. The power at y_{p1} can be estimated by generating data using values u_1 and b such that $u_1 + w_p b = y_{p1}$. In each case we generate samples according to the experimental plan, and for each sample determine whether $(\hat{y}_p - y_{p0})/se(\hat{y}_p) \le z_0$. The fraction of samples for which this is true provides an estimate of the probability of rejecting the hypothesis $y_p = y_{p0}$, given the parameter values in question.

In the case of Type 2 censoring we can base a test of H_0 versus H_1 on the exact pivotal quantity $Z_p = (\hat{u} - y_p)/\hat{b}$ in Theorem 5.1.1. For given values of r and n, probabilities $Pr(Z_p \le z_0; y_p)$ can be obtained by simulation, as discussed in Example 5.1.2. In particular, samples can be generated from $EV(0, 1)$, for which $y_p = w_p = \log(-\log(1 - p))$, and the values of $(\hat{u} - w_p)/\hat{b}$ are then realizations of Z_p. The power of the test is given by

$$Pr(Z_p \le z_0; y_p = y_{p1}) = Pr\left(\frac{\hat{u} - y_{p1}}{\hat{b}} + \frac{y_{p1} - y_{p0}}{b}\left(\frac{b}{\hat{b}}\right) \le z_0; y_{p1}\right).$$

Note that if $y_p = y_{p1}$, then $Z_p' = (\hat{u} - y_{p1})/\hat{b}$ and $Z_2 = \hat{b}/b$ are pivotal quantities, so this probability depends on $(y_{p1} - y_{p0})/b$ and thus on b. It can be estimated by

simulation, for any specified value of $(y_{p1} - y_{p0})/b$; Z_p' and Z_2 are precisely the pivotals in Theorem 5.1.1, and realizations (Z_p', Z_2) can be obtained by generating samples from the distribution with $u = 0$, $b = 1$ and computing $((\hat{u} - w_p)/\hat{b}, \hat{b})$.

A few tables have been determined by simulation in the case of the extreme value distribution: Thoman and Bain (1969) consider tests for $u = y_{.632}$ and McCool (1974) considers $y_{.10}$ and $y_{.50}$.

Two sample tests, as discussed in Section 5.4, can also be subjected to power considerations. Pivotal quantities such as (5.4.5) and (5.4.6) can be addressed in the same way as Z_p was earlier. More generally, the power for any of the tests discussed in Section 5.4 can be approximated by simulation. For normal distributions with uncensored data, power calculations based on noncentral t, χ^2 and F distributions are of course available for various one, two, and m sample tests (e.g., Johnson et al. 1995, Chs. 29–31).

5.6.3 An Example

The reliability of electronic units is often assessed in accelerated life tests in which the units are subjected to thermal cycling over a wide temperature range. Let us consider the design of a test for which the primary objective is a demonstration that the 10th percentile $t_{.10}$ for the distribution of time to failure in the extreme testing environment meets a certain minimum standard. It will be assumed that previous experience has shown the Weibull distribution to provide a reasonable model for time to failure.

Consider first the question of confidence intervals for $t_{.10}$, based on a life test in which n units are observed for a maximum of C hours; this will be the censoring time for any units that have not failed by the end of the test. Suppose further that we will be satisfied with a degree of precision for which the ratio of the upper and lower confidence limits $(UCL(t_{.10})/LCL(t_{.10}))$ for a .95 confidence interval for $t_{.10}$ is about 2. To select a plan we will have to decide on values for n and for C; to do this, it will be necessary to have preliminary estimates of the parameters for the Weibull failure time distribution, as we now discuss.

In estimating a pth quantile t_p or $y_p = \log(t_p)$, the gains in precision for an uncensored sample of size n over a censored sample with the same n are not very great, provided the fraction of uncensored failure times is a bit larger (.05–.10 is sufficient) than p. We can use tables and charts discussed by Meeker and Escobar (1998, Ch. 10), but we will take an alternative approach to illustrate how one can proceed when tables or charts are not available. As a first step, we consider the asymptotic covariance matrix $V = \text{Asvar}(\hat{u}, \hat{b})$ for the extreme value parameters u and b, based on the Fisher information matrix (5.6.1) for an uncensored sample (corresponding to $C = \infty$). This is derived in Problem 5.4 for the Weibull–extreme value model, which gives

$$V = \frac{b^2}{n}\begin{pmatrix} 1.1087 & -.1087 \\ -.1087 & .6079 \end{pmatrix}.$$

This depends on b, but not on u. The asymptotic standard error for $\hat{y}_{.10} = \hat{u} - 2.25\hat{b}$ is given by the square root of $Asvar(\hat{y}_{.10}) = (1, -2.25)V(1, -2.25)'$; this is $se(\hat{y}_{.10}) = 2.16b/\sqrt{n}$. An approximate .95 confidence interval for $y_{.10}$ is given by $\hat{y}_{.10} \pm 1.96se(\hat{y}_{.10})$, for which $UCL(y_{.10}) - LCL(y_{.10}) = 8.47b/\sqrt{n}$.

Since $t_{.10} = \exp(y_{.10})$, to give $UCL(t_{.10})/LCL(t_{.10}) = 2$, we need $UCL(y_{.10}) - LCL(y_{.10}) = \log 2 = .693$, making the sample size requirement approximately $n = 150b^2$. We therefore require an estimate of b. This may be available from previous experience; for example, the Weibull shape parameter $\beta = b^{-1}$ is roughly known in certain applications. Otherwise, we can estimate b (and u) by first estimating a pair of quantiles t_p. For the sake of illustration, suppose that we estimated $t_{.10} = 168$ hours and $t_{.50} = 672$ hours, based on previous experience. This gives $y_{.10} = 5.124$ and $y_{.50} = 6.510$, and since $y_p = u + b\log(-\log(1-p))$, we can determine that $u = 6.78$ and $b = .74$, approximately.

With the value $b = .74$, the sample size requirement is $n = 150b^2 = 82$, giving $se(\hat{y}_{.10}) = .17$. Now, we prefer to use a test with limited follow-up time C, for obvious practical reasons. Since our prior estimate of $t_{.10}$ was 168 hours, let us consider plans for which C is about 200–240 hours, assuming that test equipment can be made available for 8–10 days. At this point we could evaluate Fisher information matrices as given by (5.6.1)–(5.6.4) and corresponding asymptotic standard errors for $\hat{y}_{.10}$; Escobar and Meeker (1994) give algorithms. A preferable approach, since it provides a wider range of information, is to simulate some samples.

To be a little conservative, we considered a plan with $n = 100$ and $C = 240$ hours. We generated 100 censored random samples from the extreme value distribution with $u = 6.78$, $b = .74$, using the log censoring time log(240) for $Y = \log T$. For each sample we computed $\hat{y}_{.10}$ and $se(\hat{y}_{.10})$ as obtained from the observed information matrix $I(\hat{u}, \hat{b})^{-1}$ in (5.2.5). The sample standard deviation for $\hat{y}_{.10}$ across the 100 samples was .20, but the values of $se(\hat{y}_{.10})$ ranged from .11 to .41, with the widths of the .95 confidence intervals for $y_{.10}$ consequently ranging from .44 to 1.62, and $UCL(t_{.10})/LCL(t_{.10})$ ranging from 1.55 to 5.05. An important point to take from this is that confidence-interval widths vary substantially across repetitions of the experiment. The 10th, 50th, and 90th percentiles of $se(\hat{y}_{.10})$ across the 100 samples were .13, .20, and .27, which give $UCL(t_{.10})/LCL(t_{.10})$ ratios of 1.66, 2.21, and 2.88.

At this point, we might decide to investigate the effect of variations in the values of u and b on the precision of estimation for $t_{.10}$. We might also want to look at the effects of increasing the sample size, n, in order to increase the chance that the .95 confidence interval ratio $UCL(t_{.10})/LCL(t_{.10})$ is no greater than 2. We will not pursue this here.

It may also be desired to test a hypothesis concerning $t_{.10}$, in order to demonstrate reliability. For example, we might want a test of the hypothesis $H: t_{.10} = 168$ ($y_{.10} = 5.124$) to have high power against alternatives for which $t_{.10}$ is small. A one-sided test of H with approximate size 0.05 is to reject H if

$$Z = \frac{\hat{y}_{.10} - 5.124}{se(\hat{y}_{.10})} \leq -1.645. \qquad (5.6.7)$$

Suppose we consider the alternative hypothesis that $t_{.10} = 96$ hours ($y_{.10} = 4.564$). The power can be estimated by generating data from an extreme value model with the same value of b (.74) as before, but $u = 6.23$, chosen to make $y_{.10} = u - 2.25b = 4.564$. We generated 100 censored random samples with $n = 100$ and $C = 240$ as earlier, and found that 92 gave a Z-value satisfying (5.6.7); this estimates the power at $t_{.10} = 96$ as .92.

BIBLIOGRAPHIC NOTES

Statistical inference for location-scale parameter models has been widely studied. The exact conditional procedures of Section 5.1.2 and Appendix E for complete or Type 2 censored data were suggested (in the complete data case) by Fisher (1934) and Bartlett (1937), and subsequently studied by Fraser (1968) and others. The accuracy of large-sample procedures in Section 5.1.1, and adjustments to improve accuracy, have been discussed by various authors; Barndorff-Nielsen and Cox (1994) cover this area. Jeng and Meeker (2000), Doganaksoy and Schmee (2000), Wong and Wu (2000), and Appendix E provide additional references, emphasizing lifetime data applications. Inference for the Weibull distribution was studied extensively from the mid-1960s onward, particularly for the case of Type 2 censored data. Because of its connection with the extreme value location-scale model, it was possible to base procedures on the pivotals Z_1, Z_2, and Z_p defined in Theorem 5.2.1. Early work (e.g., Lieblein and Zelen 1956; Mann 1968; Mann and Fertig 1973; Mann et al. 1974) focussed a good deal on linear estimators of the extreme value parameters u and b, since they were more easily obtained than m.l.e.'s. Thoman et al. (1969, 1970), Billmann et al. (1972) and McCool (1970, 1974), however, considered pivotals based on m.l.e.'s. Lawless (1972, 1975, 1978) developed the conditional procedures of Section 5.2.2.1. In addition to methods based on the exact distributions of the pivotal quantities Z_1, Z_2, and Z_p, simple point estimators for u and b, and approximations for pivotal quantities were developed (e.g., Mann et al. 1974; Engelhardt 1975; Mann 1977).

Early researchers also proposed linear estimators of μ and σ for censored normal samples (e.g., Sarhan and Greenberg 1962; Persson and Rootzen 1977). Exact test and confidence interval procedures for complete data are well-known; Owen (1968) provides discussions of the noncentral t distribution and the calculation of confidence limits for quantiles of a normal distribution. Nelson and Schmee (1979) gave tables of quantiles for the pivotals Z_1, Z_2, and Z_p defined in Theorem 5.1.1 for Type 2 censored normal samples, using linear estimators of μ and σ. Schmee and Nelson (1977) gave similar tables for pivotals based on the m.l.e.'s. As with the extreme value distribution, various approximations to the distributions of pivotal quantities were developed in early work for the case of Type 2 censoring (e.g., Mann 1977).

For the case of Type 2 censoring, some tables for comparing extreme value distributions have been obtained by simulation (e.g., Thoman and Bain 1969; McCool 1970, 1974, 1975a,b; Schafer and Sheffield 1976). The comparison of normal distributions based on uncensored samples is a well-known topic (e.g., Box et al. 1978).

Efron and Tibshirani (1993) and Davison and Hinkley (1997) provide comprehensive treatments of bootstrap methodologies; Meeker and Escobar (1998, Ch. 9) consider lifetime data applications.

Location-scale models (5.5.1) with additional parameters have been considered by numerous authors. Prentice (1974) and Farewell and Prentice (1977) considered maximum likelihood methods for the generalized log-gamma model. Lawless (1980) considered exact conditional methods for Type 2 censored data; Balakrishnan and Chan (1995) give simulation-based tables for unconditional confidence intervals. Earlier work on generalized gamma models for lifetimes (e.g., Hager and Bain 1970) used versions of the model for which the connection with location-scale models was not obvious, and the parameterizations used often led to difficulties in obtaining m.l.e.'s. The generalized log-Burr model has been used to some extent (e.g., Dubey 1968; Lancaster and Nickell 1980), though the properties of maximum likelihood estimation do not seem to have been exhaustively studied (Gould 1986). Location-scale models with two or more shape parameters have also been used. One prominent family is that where $S_0(z; k)$ is a standardized log F distribution (e.g., Prentice 1975; Kalbfleisch and Prentice 1980, Sec. 3.9).

Experimental plans for log-location-scale models have been studied extensively; Meeker and Escobar (1998, Ch. 10) provide details and numerous references for estimation-based plans. Plans for reliability demonstration hypothesis tests have been considered by Fertig and Mann (1980), Schneider (1989), and Balasooriya et al. (2000). Many government and professional organizations maintain reliability standards that include life test plans for special purposes. Blischke and Murthy (2000, pp. 697–701) provide a listing of standards.

COMPUTATIONAL NOTES

Methodology for Weibull–extreme value, log-normal-normal, and log-logistic-logistic models is widely available in commercial software systems. In S-Plus see in particular, function censorReg, and in SAS, procedure LIFEREG. Some packages also handle log-Burr and log-gamma distributions. A variety of special-purpose packages exist, especially for Weibull analysis (e.g., Abernethy 1996).

S functions for bootstrap methodology are given in the books by Efron and Tibshirani (1993) and Davison and Hinkley (1997). Packages such as S-Plus implement some of the methods.

Meeker (2002) has constructed SPLIDA, a comprehensive package for the analysis of reliability data. It implements most of the methodology discussed in this chapter, and provides tools for planning as well as analyzing studies.

PROBLEMS AND SUPPLEMENTS

5.1 *Linear estimation of location and scale parameters.* Let $Y_{(1)} \le \cdots \le Y_{(n)}$ be the ordered observations in a random sample of n from a location-scale parameter distribution with p.d.f. of the form

$$f(y; u, b) = \frac{1}{b} f_0\left(\frac{y-u}{b}\right), \qquad -\infty < y < \infty. \tag{5.7.1}$$

Let $Z_{(i)} = (Y_{(i)} - u)/b$, $i = 1, \ldots, n$ be the standardized order statistics, and define

$$\alpha_i = E(Z_{(i)}) \qquad v_{ij} = \mathrm{Cov}(Z_{(i)}, Z_{(j)}) \qquad i, j = 1, \ldots, n.$$

(a) Show that if $\mathbf{Y} = (Y_{(1)}, \ldots, Y_{(n)})'$, then $E(\mathbf{Y}) = \mathbf{A}\boldsymbol{\theta}$ and $\mathrm{Var}(\mathbf{Y}) = b^2\mathbf{V}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha_1 \\ 1 & \alpha_2 \\ \vdots & \vdots \\ 1 & \alpha_n \end{bmatrix} \qquad \boldsymbol{\theta} = (u, b)' \qquad \mathbf{V} = (v_{ij})_{n\times n}.$$

Thus show that the linear unbiased estimators of u and b that have minimum variance are given by

$$\tilde{\boldsymbol{\theta}} = (\tilde{u}, \tilde{b})' = (\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V}^{-1}\mathbf{Y}$$

and that the covariance matrix for $(\tilde{u}, \tilde{b})'$ is $(\mathbf{A}'\mathbf{V}^{-1}\mathbf{A})^{-1}b^2$. Calculation of the best linear unbiased estimates (b.l.u.e.) of u and b for a given distribution therefore requires knowledge of the means, variances, and covariances of the standardized order statistics in samples from the distribution.

(b) Let $\phi = l_1u + l_2b$ and $\tilde{\phi} = l_1\tilde{u} + l_2\tilde{b}$ and suppose that $\mathrm{Var}(\tilde{\phi}) = Ab^2$, $\mathrm{Var}(\tilde{b}) = Cb^2$, and $\mathrm{Cov}(\tilde{\phi}, \tilde{b}) = Bb^2$. Define new estimators

$$b^* = \frac{\tilde{b}}{1+C} \qquad \phi^* = \tilde{\phi} - \frac{B}{1+C}\tilde{b}.$$

Prove that the mean-square errors of b^* and ϕ^* are less than those of $\tilde{b}$ and $\tilde{\phi}$, respectively. In fact, it can be shown that b^* and ϕ^* are the best linear invariant estimators of b and ϕ.

(Lloyd 1952; Mann, 1969)

5.2 *Equivariant estimators of location and scale parameters.* Consider a location-scale parameter model with p.d.f. (5.7.1).

(a) Consider a Type 2 censored sample from (5.7.1) and linear estimators of u and b of the form

$$\tilde{u} = \sum_{i=1}^{r} a_i(n, r)y_{(i)}, \qquad \tilde{b} = \sum_{i=1}^{r} b_i(n, r)y_{(i)}.$$

Determine necessary and sufficient conditions on the coefficients $a_i(n, r)$ and $b_i(n, r)$ so that $\tilde{u}$ and $\tilde{b}$ are equivariant, that is, they satisfy (E2) and (E3) of Appendix E.

(b) Show that the b.l.u.e.'s of u and b from Problem 5.1 are equivariant. (*Hint:* Show that $\sum_{i=}^{r} \alpha_i = 0$.)

(c) Let $\hat{u}$ and $\hat{b}$ be the m.l.e.'s of u and b from a Type 2 censored sample. Show that $\hat{u}$ and $\hat{b}$ are equivariant.

(d) Show that the m.l.e.'s $\hat{u}$ and $\hat{b}$ from a Type 1 censored sample where each individual has a prespecified censoring time C_i $(i = 1, \ldots, n)$ are not equivariant. What are the ramifications of this for the quantities $Z_1 = (\hat{u} - u)/\hat{b}$ and $Z_2 = \hat{b}/b$ defined in Theorem 5.1.1?

(Section 5.1, App. E)

5.3 *Bayesian inference with an improper prior.* Consider a censored sample from a location-scale parameter distribution (5.7.1) and an improper prior distribution

$$p(u, b) = \frac{1}{b} \qquad -\infty < u < \infty, \qquad b > 0.$$

(a) Determine the form of the marginal posterior distributions for u and b.

(b) If the data are Type 2 censored, show that the posterior probability intervals in part (a) are numerically identical to conditional confidence intervals for u and b obtained by using the pivotal quantities $Z_1 = (\tilde{u} - u)/\tilde{b}$ and $Z_2 = \tilde{b}/b$ in Theorem E2, along with the results of Theorem E3.

(Section 5.1.2; Bogdanoff and Pierce 1973)

5.4 Let $\ell_i(u, b)$ be the contribution to the log-likelihood function (5.2.4) from an individual with potential censoring time C_i under a Type 1 censoring scheme and an assumed extreme value model (5.2.2).

(a) Use the discussion leading to (5.1.10) to show that the Fisher (expected) information matrix $\mathcal{I}(u, b)$ has entries that are the sums over $i = 1, \ldots, n$ of

$$\mathcal{I}_{uu,i} = E\left[\frac{-\partial^2 \ell_i(u, b)}{\partial u^2}\right] = \frac{1}{b^2}\left[1 - \exp(-e^{R_i})\right]$$

$$\mathcal{I}_{bb,i} = E\left[\frac{-\partial^2 \ell_i(u, b)}{\partial b^2}\right]$$

$$= \frac{1}{b^2}\left[\int_{-\infty}^{R_i} (1 + z^2 e^z) \exp(z - e^z)\, dz + R_i^2 \exp(R_i - e^{R_i})\right]$$

$$\mathcal{I}_{ub,i} = E\left[\frac{-\partial^2 \ell_i(u, b)}{\partial u \partial b}\right]$$

$$= \frac{1}{b^2}\left[\int_{-\infty}^{R_i} z \exp(2z - e^z)\, dz + R_i \exp(R_i - e^{R_i})\right],$$

where $R_i = (\log C_i - u)/b$.

(b) Determine $\mathcal{I}(u, b)$ for an uncensored sample of n observations by letting $C_i \to \infty$ in part (a). Thus show that the covariance matrix for the asymptotic normal distribution of $\sqrt{n}(\hat{u} - u, \hat{b} - b)$ is

$$n\mathcal{I}(u, b)^{-1} = b^2 \begin{pmatrix} 1 + \frac{6}{\pi^2}(1 - \gamma)^2 & -\frac{6}{\pi^2}(1 - \gamma) \\ -\frac{6}{\pi^2}(1 - \gamma) & \frac{6}{\pi^2} \end{pmatrix}, \qquad (5.7.2)$$

where $\gamma = .5772$ is Euler's constant (see (B11) of Appendix B).

(Section 5.2; Meeker and Nelson 1977)

5.5 Consider complete samples of size n from the extreme value distribution (5.2.2).

(a) Compare exact quantiles of $\hat{b}/b$ given in Table 1 of Thoman et al. (1969) with approximate quantiles obtained from (1) the asymptotic normal approximation $\hat{b}/b \sim N(1, 6/(\pi^2 n))$, and (2) the approximation $\log(\hat{b}/b) \sim N(0, 6/(\pi^2 n))$, both obtained from (5.7.2). Also make comparisons with percentage points given by the χ^2 approximation (5.2.17).

(b) McCool (1974) gives the following quantiles for $Z_p = (\hat{y}_p - y_p)/\hat{b}$, determined by simulation, for the case $n = 30$, $p = .10$:

.01	.05	.10	.90	.95	.99
−.790	−.567	−.442	.706	.915	1.389

Compare these with approximate quantiles of Z_p derived from the normal approximation $(\hat{u}, \hat{b}) \sim N_2[(u, b), \mathcal{I}(u, b)^{-1}]$, using (5.7.2).

(Section 5.2)

5.6 Show that the conditional methods developed in Section 5.1.2 apply to the case of progressively Type 2 censored data, defined in Section 2.2.1. In particular, show that under the sampling distribution (2.2.8) that applies to progressive Type 2 censoring with two stages, the results in Theorems E1 and E2 of Appendix E are still valid.

(Section 5.1.2; Viveros and Balakrishnan 1994)

5.7 *Generating a Type 2 Censored Sample.* Let $T_{(1)} < \cdots < T_{(r)}$ be a Type 2 censored sample based on n lifetimes from a distribution with cumulative hazard function $H(t)$, and let $E_{(i)} = H(T_{(i)})$, $i = 1, \ldots, r$.

(a) Show that $E_{(1)}, \ldots, E_{(r)}$ are the first r order statistics in a sample of size n from a standard exponential distribution.

(b) Use the representation in Theorem 4.1.1, that is,

$$E_{(i)} = \sum_{j=1}^{i} \frac{W_j}{(n-j+1)}, \qquad i = 1, \ldots, r,$$

where $W_1, \ldots, W_r$ are independent standard exponential random variables, to suggest a way of generating $T_{(1)}, \ldots, T_{(r)}$ for a distribution with easily invertible $H(T)$.

(c) Apply this method to the Weibull distribution (5.2.1).

(Section 5.2)

5.8 Table 5.9 shows results of an experiment designed to compare the performances of high-speed turbine engine bearings made out of five different compounds (McCool, 1979). The experiment tested 10 bearings of each type; the times to fatigue failure are given in units of millions of cycles.

(a) Assuming that the failure times in each sample came from a Weibull distribution (5.2.1), obtain m.l.e.'s for α and β and find confidence intervals for the tenth percentile of each distribution ($t_{.10}$ is used as a rating life).

(b) Carry out a comparison of the five failure time distributions and, in particular, the tenth percentiles of the distributions.

(c) Investigate whether a log-normal distribution also fits the data by considering a log-gamma model. Compare confidence intervals for $t_{.10}$ under a log-normal model with those in part (a).

(Sections 5.2, 5.3, 5.5)

5.9 In Example 4.5.1 estimates of $t_{.50}$ were given for a lifetime distribution assumed to be three-parameter Weibull, with different values $\gamma = 60$, 100, and 140 assumed for the threshold parameter. Obtain .95 confidence intervals for

Table 5.9. Failure Times of Bearing Specimens

Type of Compound				
I	II	III	IV	V
3.03	3.19	3.46	5.88	6.43
5.53	4.26	5.22	6.74	9.97
5.60	4.47	5.69	6.90	10.39
9.30	4.53	6.54	6.98	13.55
9.92	4.67	9.16	7.21	14.45
12.51	4.69	9.40	8.14	14.72
12.95	5.78	10.19	8.59	16.81
15.21	6.79	10.71	9.80	18.39
16.04	9.37	12.58	12.28	20.84
16.84	12.75	13.41	25.46	21.51

$t_{.50}$ in each of these situations via the likelihood ratio method of Section 5.2.1. Comment on the extent to which the intervals depend on γ.

(Sections 4.5, 5.2.1)

5.10 The data below show the number of cycles to failure for twenty-five 100-cm. specimens of yarn, tested at a particular strain level:

86, 146, 251, 653, 98, 175, 176, 76, 264, 15, 157, 220, 42, 321, 180, 198, 38, 20, 61, 121, 282, 224, 149, 180, 325.

Determine .95 lower confidence limits for the .01 and .10 quantiles of the failure time distribution, assuming a log-normal distribution. Check on the adequacy of the model. Repeat the analysis using a log-logistic model.

(Section 5.3)

5.11 The observations below are survival times (in weeks) of male mice exposed to a 240-roentgen dose of gamma radiation (Furth et al. 1959; Kimball 1960). Brackets after a value indicate the number of observations with that value.

40	48	50	54	56	59
62	63	67(2)	69	70	71
73(2)	76	77	80	81(2)	82
83	84	86(2)	87	88(5)	89
90(2)	91	93	94	95	96
97(2)	98	99(2)	100(4)	101(3)	102(2)
103(5)	104(3)	105(2)	106(3)	107	108
109(2)	110(3)	111(3)	11	113(2)	114(2)
115	116(2)	117	118(3)	119(2)	120(3)
121(2)	123(2)	124(3)	125(2)	126(5)	127(4)
128(4)	129(6)	130(4)	131(2)	132	133(3)
134(4)	135(3)	136(4)	137(3)	138	139(2)
140(2)	141(5)	142	144(5)	145(2)	146(4)
147(4)	148(4)	149	150	151(4)	152(2)
153	155	156	157	158(2)	160
161	162(2)	163(2)	164	165(2)	166
168	169	171(2)	172(2)	174	177(2)

Assess whether any of the log-logistic, log-normal, or Weibull models fit these data. What do you conclude about the survival distribution and the shape of the hazard function?

(Sections 5.2, 5.3, 5.5, 3.3)

5.12 Prove that (5.4.14) is a pivotal quantity when the data are complete or Type 2 censored.

(Section 5.4.2)

5.13 Suppose that two Weibull distributions have the same shape parameter β, but possibly different scale parameters α_1 and α_2. The ratio of the hazard func-

tions then depends on $\delta = (\alpha_1/\alpha_2)^\beta$, and the survivor functions are related by $S_1(t) = S_2(t)^\delta$.

(a) Develop likelihood ratio tests for δ and show how to use these to obtain confidence intervals for δ. Apply this to the data on leukemia remission times in Example 5.2.1, assuming that the lifetimes are mutually independent.

(b) Develop an alternative procedure based on the approximate normality of the m.l.e.'s for u_1, b_1, u_2, b_2, and apply it to the leukemia data.

(Section 5.4)

5.14 The data below are failure times for two types of polyethylene cable insulation, obtained from an accelerated life test. Of the 10 specimens of each type tested, 9 failed. Ordered failure times, in hours, are given below. The last time in each case is a censoring time.

Type I 5.1, 9.2, 9.3, 11.8, 17.7, 19.4, 22.1, 26.7, 37.3, 60.0*

Type II 11.0, 15.1, 18.3, 24.0, 29.1, 38.6, 44.2, 45.1, 50.9, 70.0*

Assuming that failure times for each type have a Weibull distribution, compare the two failure time distributions and assess the possible superiority of Type II insulation.

(Section 5.4)

5.15 *Paired data.* When response data are paired so that the two units in a pair received different treatments A and B, a common method of analysis is to consider the differences in responses for each pair. However, this is problematic when the responses are survival times subject to censoring. The data in Table 5.10 show log (base 10) survival times Y of rats poisoned with carbon tetrachloride in a laboratory experiment (Sampford and Taylor 1959); time was measured in minutes. Pairing was used to study whether injecting a rat with vitamin B_{12} had an effect on survival time. In the experiment, one rat from a pair of litter mates was injected with vitamin B_{12} and the other received no vitamin. Observation was suspended after 16 hours ($y = 2.98$), and three survival times were consequently censored.

(a) Suppose that, given an effect α_i specific to the ith pair of rats, the log survival times Y_{1i} and Y_{2i} for the vitamin B_{12} and control animals are independent, with $Y_{ji} \sim N(\mu_j + \alpha_i, \sigma_1^2)$, $j = 1, 2$. Hence $Z_i = Y_{1i} - Y_{2i} \sim N(\delta, \sigma^2)$, where $\delta = u_1 - u_2$ and $\sigma^2 = 2\sigma_1^2$. Note that if Y_{1i} and Y_{2i} are both censored, then Z_i is unknown. Explain why it would be improper to analyze the data by simply dropping doubly censored pairs and treating the rest as a random sample from $N(\delta, \sigma^2)$.

(b) For the data in Table 5.10 only 3 times are censored, producing 2 left-censored and 1 right-censored Z_i value. Treating the Z_i as arising from an independent censoring mechanism, estimate δ and thus the difference in median log survival, using the model in part (a). State any reservations you have about this approach.

Table 5.10. Log_{10} Survival Times for Rats

$y_1(B_{12})$	y_2 (Control)	$z = y_1 - y_2$
2.73	> 2.98	< −.25
2.80	> 2.98	< −.18
2.01	2.84	−.83
2.19	2.76	−.57
2.34	2.83	−.49
2.61	2.73	−.12
2.51	2.62	−.11
2.65	2.70	−.05
2.72	2.76	−.04
2.79	2.82	−.03
2.90	2.79	.11
2.78	2.64	.14
2.78	2.48	.30
2.97	2.64	.33
2.74	2.31	.43
2.96	2.51	.45
> 2.98	2.68	> .30

(c) An extreme value model also turns out to be tractable. Suppose that independent Y_{ji} have extreme value distributions (5.2.2) with location parameters $u_{ji} = u_j + \alpha_i$ and scale parameters $b_{ji} = b$. Show that the distribution of $Z_i = Y_{1i} - Y_{2i}$ is logistic, with location parameter $\delta = u_1 - u_2$, scale parameter b, and survivor function

$$S(z) = \left[1 + \exp\left(\frac{z - \delta}{b}\right)\right]^{-1} \qquad -\infty < z < \infty.$$

(d) Estimate δ using the model in part (c) under the same assumption about censoring as in part (b). Compare confidence intervals with those in part (b), under the same assumption about censoring as in part (b). Why would you expect them to be similar?

Paired data are discussed further in Section 11.2.

(Sampford and Taylor 1959; Holt and Prentice 1974)

5.16 *Prediction of a future observation.* Suppose that $y_1 \le \cdots \le y_r$ are the r smallest observations in a sample of size n from the extreme value distribution (5.2.2) and let $\tilde{u}$ and $\tilde{b}$ be equivariant estimators of the parameters in the model, based on $y_1, \ldots, y_r$. If Y_1^* is the smallest observation in a future sample of size m from the same distribution, then prediction intervals for Y_1^* can be based on the quantity

$$U = \frac{Y_1^* - \tilde{u}}{\tilde{b}}.$$

(a) Show that U is a pivotal quantity.

(b) Determine the joint distribution of $W_1 = (Y_1^* - u)/b$, $Z_2 = \tilde{b}/b$, and $Z_3 = (\tilde{u} - u)/b$, conditional on the ancillary statistics $a_i = (y_i - \tilde{u})/\tilde{b}$ $(i = 1, \ldots, r)$. Noting that $U = (W_1 - Z_3)/Z_2$, show that the conditional survivor function for U, given $\mathbf{a}$, can be written as

$$Pr(U \geq u|\mathbf{a}) = \int_0^\infty \left[t^{r-2} \exp\left(t \sum_{i=1}^{r} a_i \right) \Big/ \left(me^{ut} + \sum_{i=1}^{r^*} e^{a_i t} \right)^r \right] dt,$$

where $\sum_1^{r^*}$ is defined as in Section 5.2.2. This allows conditional confidence intervals to be obtained for Y_1^*. Mann (1977) discusses an approximation to the unconditional distribution of U, which can also be approximated very accurately by simulation.

(Sections 4.6, 5.2.2; Lawless 1973)

5.17 *Prediction of a future observation (continued).* Let (y_i, δ_i), $i = 1, \ldots, n$ be a censored random sample of log-lifetimes, where $Y = \log T$ has a location-scale distribution (5.1.1). Let $\hat{u}$, $\hat{b}$ be the m.l.e.'s for u and b, and let Y^* be an independent future observation from the same distribution.

(a) Argue that if the censoring is Type 2, then

$$W = (Y^* - \hat{u})/\hat{b}$$

is a pivotal quantity, and that in general, W is pivotal in the limit as n increases.

(b) The distribution of W can be approximated by simulation, as discussed in Section 4.6. A parametric bootstrap approach is to generate values W_j^* $(j = 1, \ldots, B)$ by simulating y^* and samples (y_i^*, δ_i^*), $i = 1, \ldots, n$, from $EV(\hat{u}, \hat{b})$, giving m.l.e.'s $\hat{u}^*$, $\hat{b}^*$ and a realized W-value $w^* = (y^* - \hat{u}^*)/\hat{b}^*$. The values $w_1^*, \ldots, w_B^*$ provide an empirical estimate of the distribution of W. When the censoring mechanism is not sufficiently specified that this approach can be used (see Appendix D.2 and Section 4.6), the pseudosample (y_i^*, δ_i^*), $i = 1, \ldots, n$, can instead be generated by nonparametric bootstrap sampling with replacement from the n data pairs (y_i, δ_i), $i = 1, \ldots, n$. Use this approach with the ball bearing failure time data in Examples 5.3.2 and 5.5.2 to obtain a one-sided .95 prediction interval (y_L, ∞) for Y^*, based on both extreme value and normal models for Y. Compare the values of y_L with the naive or plug-in values obtained by treating $\hat{u}$, $\hat{b}$ as the true values of u, b and solving $S(y_L; \hat{u}, \hat{b}) = .95$ to get y_L.

(c) Based on the same data, use simulation to obtain a lower .95 prediction limit for the smallest observation Y_1^* in a future sample of 10 log-lifetimes.

(Sections 4.6, 5.2, 5.3)

5.18 *Bayesian prediction.* Suppose that Y has a location-scale distribution (5.1.1) with parameters u and b. Let D represent the observed data from a censored random sample (y_i, δ_i), $i = 1, \ldots, n$, and $\text{post}(u, b|D)$ the posterior distribution obtained from D and a prior distribution for (u, b). The Bayesian predictive p.d.f. for a future observation Y^* is then given by

$$p(y|D) = \int f(y|u, b)\text{post}(u, b|D)\, du\, db,$$

where $f(y|u, b) = f(y; u, b)$ is the p.d.f. of Y^*.

Using the improper prior b^{-1} for u and b in Problem 5.3, obtain $p(y|D)$ in the case of an extreme value model $EV(u, b)$ for Y.

(Sections 4.6, 5.2; Bogdanoff and Pierce 1973)

5.19 *Estimation of median residual lifetime.* If lifetimes T have survivor function $S(t; \boldsymbol{\theta})$, then the median conditional lifetime at time t_0 is the value $t_0(\boldsymbol{\theta})$ satisfying

$$Pr(T \geq t_0(\boldsymbol{\theta})|T \geq t_0) = \frac{S(t_0(\boldsymbol{\theta}))}{S(t_0)} = .5.$$

The median residual lifetime is $t_0(\boldsymbol{\theta}) - t_0$.

(a) For a Weibull model (5.2.1) show that $y_0(\boldsymbol{\theta}) = \log t_0(\boldsymbol{\theta})$ is given by

$$y_0(u, b) = u + b \log[-\log(.5) + e^{(y_0 - u)/b}],$$

where $y_0 = \log t_0$ and u, b are the extreme value parameters in (5.2.2).

(b) Investigate methods of obtaining confidence limits for $y_0(u, b)$, for a given y_0; assume that a censored random sample (y_i, δ_i), $i = 1, \ldots, n$ from the distribution of $Y = \log T$ is available.

(c) Consider the leukemia remission time data in Example 5.2.1 for the group of patients receiving the drug 6-MP. Assuming a Weibull distribution for remission times, obtain a .95 lower confidence limit for the median residual remission time, given $T \geq 15$ weeks.

(Section 5.2)

5.20 Failures can occur in microcircuits because of the movement of atoms in the conductors in the circuit; this is referred to as electromigration. The data below are from an accelerated life test of 59 conductors (Schafft et al. 1987; Nelson and Doganaksoy 1995). Failure times are in hours, and there are no censored observations.

6.545	9.289	7.543	6.956	6.492	5.459	8.120	4.706
8.687	2.997	8.591	6.129	11.038	5.381	6.958	4.288
6.522	4.137	7.459	7.495	6.573	6.538	5.589	6.087
5.807	6.725	8.532	9.663	6.369	7.024	8.336	9.218
7.945	6.869	6.352	4.700	6.948	9.254	5.009	7.489
7.398	6.033	10.092	7.496	4.531	7.974	8.799	7.683
7.224	7.365	6.923	5.640	5.434	7.937	6.515	6.476
6.071	10.491	5.923					

Fit log-Burr and log-gamma distributions to these data. The log-normal model has often been found satisfactory with such data; consider the support for it and also for Weibull and log-logistic failure time distributions in the data here.

(Sections 5.2, 5.3, 5.5)

5.21 *Planning a comparative experiment.* Suppose that the distributions of lifetimes for two groups of individuals are adequately represented by Weibull distributions (5.2.1). The shape parameter β can be assumed to be roughly 1.5 for each distribution. An experiment is to be planned to test the hypothesis $H: \alpha_1 = \alpha_2$ or equivalently in terms of the extreme value parameters in (5.2.2), $H: u_1 = u_2$. The study will include an equal number of individuals from each group. We will consider two-sided tests of H based on the asymptotically normal statistic

$$Z = \frac{(\hat{u}_1 - \hat{u}_2)}{[se(\hat{u}_1)^2 + se(\hat{u}_2)^2]^{1/2}},$$

where $\hat{u}_1$ and $\hat{u}_2$ are the m.l.e.'s of u_1 and u_2 from the experimental data.

(a) If there is no censoring, use (5.7.2) from Problem 5.4 and the value $b = \beta^{-1} = .67$ to construct tests of size .95. Determine the value of n (total number of test individuals) needed to make the probability of rejecting H equal to .8, if $|u_1 - u_2| = \log 2$.

(b) What additional information is necessary, if the experiment must be terminated after time C? Discuss how you would choose n in this case.

(Section 5.6)

CHAPTER 6

Parametric Regression Models

6.1 INTRODUCTION

In most studies there are covariates or explanatory variables such as treatments, group indicators, individual characteristics, or environmental conditions, whose relationship to lifetime is of interest. This leads to a consideration of regression models.

Examples 1.1.5 and 1.1.7 to 1.1.9 in Chapter 1 described situations in which covariates were present. The following are two further examples.

Example 6.1.1. Krall et al. (1975) discussed a situation in which survival times for 65 multiple myeloma patients were examined in conjunction with 16 explanatory variables. The latter included physiological measures such as the white blood count of the individual at the time of diagnosis, qualitative factors such as the presence or absence of infection at diagnosis, and personal characteristics such as sex and age. A primary objective was to determine whether survival time was related to the explanatory variables. The data are discussed in Problem 6.9.

Example 6.1.2. Hamada (1995) described a multifactor experiment designed to improve the reliability of drill bits used in the manufacture of printed circuit boards. Small-diameter holes are wanted, but small-diameter drill bits may break; this is a serious problem because broken bits leave material embedded in the board, which then has to be scrapped at a cost of several hundred dollars. The experiment was planned to identify factors affecting bit breakage and lifetime (measured as the number of holes drilled before breakage of the bit), and to design a bit that would be reliable under varying manufacturing conditions. Eleven factors were investigated in a fractional factorial design; they included physical characteristics of the bit (e.g., moment of inertia, type of material, point type) as well as external factors such as the type of material in the printed circuit board.

Regression models were introduced in Section 1.4, where it was noted that both parametric and nonparametric methods are useful in the analysis of lifetime data. In addition, covariates may vary over time. This chapter concentrates on parametric

methods and mostly on fixed (nontime-varying) covariates; semiparametric methods are discussed in Chapters 7 and 8. Time-varying covariates are considered in Section 6.4.3 and in Chapters 7 and 8.

Regression analysis of lifetimes involves specifications for the distribution of a lifetime, T, given a vector of covariates $\mathbf{x}$. The most important parametric models are extensions of models in Chapters 4 and 5, to allow parameters to depend on $\mathbf{x}$. For example, consider the Weibull distribution (5.2.1) with scale parameter α and shape parameter δ: regression models for which either α or δ depend on $\mathbf{x}$ may be considered. Since α and δ are positive-valued, one pair of convenient specifications is $\alpha(\mathbf{x}) = \exp(\boldsymbol{\beta}'\mathbf{x})$ and $\delta(\mathbf{x}) = \exp(\boldsymbol{\gamma}'\mathbf{x})$, where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of regression coefficients of the same length as $\mathbf{x}$; in this case, $\alpha(\mathbf{x}) > 0$ and $\delta(\mathbf{x}) > 0$ without any restrictions on $\boldsymbol{\beta}$ or $\boldsymbol{\gamma}$.

A Weibull model that proves useful in many situations has only α depending on $\mathbf{x}$, so that the survivor function (s.f.) of T is

$$S(t|\mathbf{x}) = \exp[-(t/\alpha(\mathbf{x}))^{\delta}], \qquad t \geq 0. \tag{6.1.1}$$

The log-lifetime $Y = \log T$ in this case has s.f.

$$S(y|\mathbf{x}) = \exp\left[-\exp\left(\frac{y - u(\mathbf{x})}{b}\right)\right], \qquad -\infty < y < \infty, \tag{6.1.2}$$

where $u(\mathbf{x}) = \log \alpha(\mathbf{x})$ and $b = \delta^{-1}$. This is an extreme value location-scale parameter distribution (5.2.2), with $u = u(\mathbf{x})$. In terms of T, the model (6.1.1) is referred to as a log-location-scale or accelerated failure time model. These are the most widely used type of parametric regression model. Let us consider them in a bit more detail.

6.1.1 Log-Location-Scale (Accelerated Failure Time) Regression Models

Location-scale regression models take the distribution of Y given $\mathbf{x}$ to be of the form (5.1.1), with $u = u(\mathbf{x})$ and

$$S(y|\mathbf{x}) = S_0\left(\frac{y - u(\mathbf{x})}{b}\right), \qquad -\infty < y < \infty, \tag{6.1.3}$$

where $S_0(z)$ is independent of $\mathbf{x}$. Another way to express this is as

$$Y = u(\mathbf{x}) + bZ, \tag{6.1.4}$$

where Z is a random variable with s.f. $S_0(z)$. The family of models for which Z has a standard normal distribution is a frequent basis of regression analysis; with lifetime data the use of extreme value, logistic, and other distributions for Z is also common. As noted in Section 1.4, the s.f. for T given $\mathbf{x}$ corresponding to (6.1.3) is of the form

$$S(t|\mathbf{x}) = S_0^*[(t/\alpha(\mathbf{x}))^{\delta}], \qquad t \geq 0, \tag{6.1.5}$$

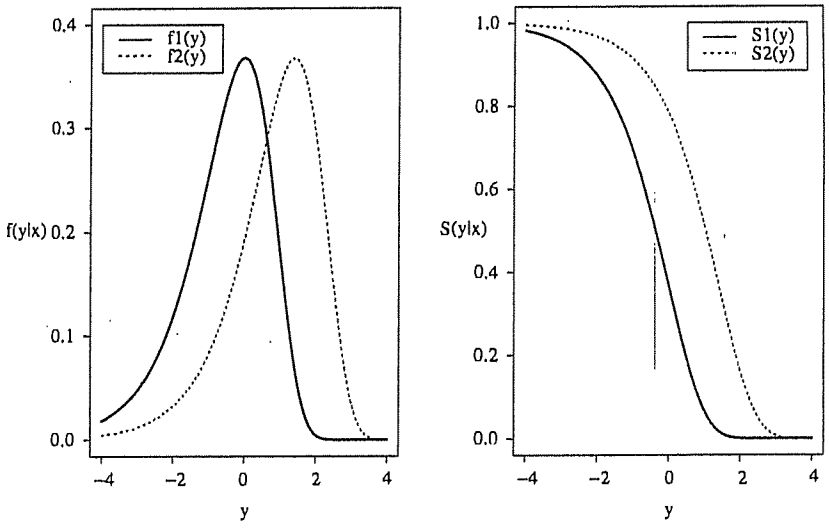

Figure 6.1. Density and survivor functions for location-scale regression models.

where $\alpha(\mathbf{x}) = \exp(u(\mathbf{x}))$, $\delta = b^{-1}$, and $S_0^*(t) = S_0(\log t)$. The covariates effectively alter the time scale and (6.1.5) is often referred to as an accelerated failure time (AFT) model. In particular, if $\alpha(\mathbf{x}) > 1$, the effect of the covariate vector is to decelerate time, and if $\alpha(\mathbf{x}) < 1$, the effect is to accelerate it.

Figure 6.1 shows the effects of covariates on the probability density and s.f.'s of log-lifetime Y. Different covariate vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ give functions that are translations of one another; they have the same shape but are separated by a distance $u(\mathbf{x}_1) - u(\mathbf{x}_2)$. Such models are especially useful when lifetimes for different individuals can vary by orders of magnitude, as for the electrical insulating fluid failure times in Example 5.4.1. Many engineering models in which failure is accelerated by thermal, voltage, or other stresses are of this type, and have linear specifications $u(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$. For example, the model with $u(x) = \beta_0 + \beta_1 x$, with $x = \log$ (stress) is often referred to as an inverse power law model, since β_1 is typically negative; this is often used with high-voltage stresses. For temperature as a stress factor, the Arrhenius model with $u(x) = \beta_0 + \beta_1 x$ and $x = d^{-1}$ is often used, where d is the temperature in degrees Kelvin (i.e., degrees Celsius plus 273.15).

Accelerated failure time, or log-location-scale models are also useful in other fields of application and, indeed, dominate many areas of regression analysis. Sections 6.3 and 6.4 deal with analysis based on (6.1.4), with $u(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$, and extensions of this model.

6.1.2 Proportional Hazards Regression Models

There are two main approaches to regression modeling for lifetimes. One uses time transformations, assuming that the effect of covariates is equivalent to altering the

rate at which time passes; the AFT models just discussed are of this type. The second approach adopts specifications of the way that the covariates affect the hazard function for T. The most common model of this type is the PH model, for which the hazard function for T, given $\mathbf{x}$, is of the form

$$h(t|\mathbf{x}) = h_0(t)r(\mathbf{x}), \tag{6.1.6}$$

where $r(\mathbf{x})$ and $h_0(t)$ are positive-valued functions. The function $h_0(t)$ is usually called the baseline hazard function; it is the hazard function for an individual whose covariate vector $\mathbf{x}$ is such that $r(\mathbf{x}) = 1$. A common specification for $r(\mathbf{x})$ is $\exp(\boldsymbol{\beta}'\mathbf{x})$, in which case $h_0(t)$ is the hazard function when $\mathbf{x} = \mathbf{0}$. The name proportional hazards (PH) comes from the fact that any two individuals have hazard functions that are constant multiples of one another.

Fully parametric PH models specify $h_0(t; \boldsymbol{\alpha})$ and $r(\mathbf{x}; \boldsymbol{\beta})$ in (6.1.6) parametrically. It follows from (6.1.6) and the relationship $S(t|\mathbf{x}) = \exp[-H(t|\mathbf{x})] = \exp[-\int_0^t h(u|\mathbf{x})\,du]$ that the s.f. for T, given $\mathbf{x}$, is of the form

$$S(t|\mathbf{x}) = S_0(t)^{r(\mathbf{x})}, \tag{6.1.7}$$

where $S_0(t) = \exp[-H_0(t)]$ is a baseline s.f. Figure 6.2 shows hazard and survivor functions for two different covariate vectors, $\mathbf{x}_1$ and $\mathbf{x}_2$. Note that one s.f. must lie completely above the other, by (6.1.7).

A feature of PH models is that if $S_0(t; \boldsymbol{\alpha})$ is in some family of parametric models, then $S(t|\mathbf{x})$ is not in general in the same family, though it is if $h_0(t)$ is of the form $\alpha_1 h_1(t; \boldsymbol{\alpha}_2)$ for parameters α_1 and $\boldsymbol{\alpha}_2$. This is in contrast to the situation for AFT

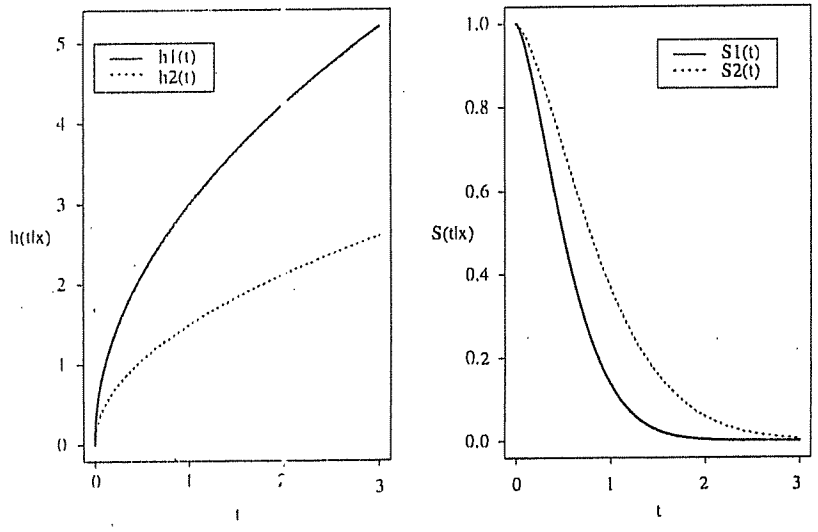

Figure 6.2. Hazard and survivor functions for PH regression models.

models, and is perhaps one reason why fully parametric PH models are used less than semiparametric models where $h_0(t)$ in (6.1.6) is left arbitrary. A more important reason is the existence of simple, flexible methods for the semiparametric PH model, which is the topic of Chapter 7.

One widely used parametric PH family is the Weibull; it is easily checked that (6.1.1) is a PH model since its hazard function is

$$h(t|\mathbf{x}) = \frac{\delta}{\alpha(\mathbf{x})}\left[\frac{t}{\alpha(\mathbf{x})}\right]^{\delta-1} = (\delta t^{\delta-1})\alpha(\mathbf{x})^{-\delta}. \tag{6.1.8}$$

This Weibull family is the only set of models that is in both the PH and AFT classes; see Problem 6.1.

An important point concerning the specification of covariate effects in Weibull AFT and PH models should be noted. It is common with PH models (6.1.6) to use the specification $r(\mathbf{x}) = \exp(\boldsymbol{\beta}'_{\text{PH}}\mathbf{x})$. For AFT models (6.1.1) the specification $\alpha(\mathbf{x}) = \exp(\boldsymbol{\beta}'_{\text{AFT}}\mathbf{x})$ is common; this corresponds to $u(\mathbf{x}) = \boldsymbol{\beta}'_{\text{AFT}}\mathbf{x}$ in the corresponding location-scale model (6.1.2) for log-lifetime. Now, if the model is Weibull, as in (6.1.1), then (6.1.8) gives the hazard function, and by comparison with (6.1.6) we have that $\exp(\boldsymbol{\beta}'_{\text{PH}}\mathbf{x}) = \exp(-\delta\boldsymbol{\beta}'_{\text{AFT}}\mathbf{x})$. Thus the two sets of regression coefficients are not the same, but related by

$$\boldsymbol{\beta}_{\text{PH}} = -\delta\boldsymbol{\beta}_{\text{AFT}} = -\frac{1}{b}\boldsymbol{\beta}_{\text{AFT}}, \tag{6.1.9}$$

where b is the scale parameter in (6.1.2).

It often occurs that the distributions of T or Y for different covariate values $\mathbf{x}_1$ and $\mathbf{x}_2$ are ordered in the sense that $S(t|\mathbf{x}_1) \geq S(t|\mathbf{x}_2)$ for all t, or $S(t|\mathbf{x}_1) \leq S(t|\mathbf{x}_2)$. In this case, we may seek a family of models for which the effect of covariates has a simple interpretation. The AFT and PH models are both of this type. For AFT models the effect of changes in covariates is to shift the distribution, as shown in Figure 6.1; this means, among other things, that the quantiles of Y are all translated by a constant amount and the quantiles of T are all multiplied by a constant amount. For PH models the effect of covariate changes is to multiply the hazard function by a constant amount.

6.1.3 Other Regression Models

The AFT and PH families accommodate settings where the distributions are ordered as in the preceding paragraph. Many other models also have this property. For example, additive hazards models with

$$h(t|\mathbf{x}) = h_0(t; \boldsymbol{\alpha}) + r(\mathbf{x}; \boldsymbol{\beta}) \tag{6.1.10}$$

are sometimes useful. The s.f.'s for (6.1.10) are ordered because the cumulative hazard functions are ordered: $H(t|\mathbf{x}_1) - H(t|\mathbf{x}_2) = [r(\mathbf{x}_1) - r(\mathbf{x}_2)]t$. This is also true

of any model for which $S(t|\mathbf{x})$ satisfies

$$\psi[S(t|\mathbf{x})] = \psi[S_0(t)] + w(\mathbf{x}; \boldsymbol{\beta}) \tag{6.1.11}$$

for some specified monotonic function $\psi(p)$. The PH family (6.1.7) satisfies (6.1.11) with $\psi(p) = \log(-\log p)$ and $w(\mathbf{x}; \boldsymbol{\beta}) = \log r(\mathbf{x})$. Another model of this type is the proportional odds family, for which $\psi(p) = \log((1-p)/p)$; see Problem 6.3.

In many applications the s.f.'s for different covariate values cross, and other models must be sought. Generalizations of AFT models that allow the scale parameter b in (6.1.3) to depend on $\mathbf{x}$ are capable of dealing with this. Extensions to PH and other of the models just cited can also be used.

Any of the models discussed in previous chapters can be extended to handle covariates. For example, in applications where long-term survivors are common, mixture models with s.f.'s

$$S(t|\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\alpha})S_0(t|\mathbf{x}; \boldsymbol{\beta}) + 1 - p(\mathbf{x}; \boldsymbol{\alpha}) \tag{6.1.12}$$

can be considered. These generalize the cure-rate models (4.4.1), with $1 - p(\mathbf{x}; \boldsymbol{\alpha})$ representing the fraction of long-term survivors among individuals with covariate vector $\mathbf{x}$. Some other models are discussed in Section 6.5.

Section 6.2 introduces some general techniques for exploring and checking models. The remaining sections of this chapter deal with the analysis of data under specific types of models. As in Chapter 5, we consider right-censored data only. Extensions to deal with interval censoring or truncation are straightforward, since parametric models and likelihood methods are employed for analysis.

6.2 GRAPHICAL METHODS AND MODEL ASSESSMENT

Graphical methods are useful for summarizing information and suggesting possible models. They also provide ways to check assumptions concerning the form of a lifetime distribution and its relationship to covariates. Some procedures that help in formulating and checking regression models for lifetime data are discussed in this section.

6.2.1 Looking for Models

One approach to the formulation of models is to fit certain canonical models and then to assess their suitability. This is reasonable when past experience or theory points toward certain types of models. However, it is often necessary to do some exploratory analysis in which the broad form of the lifetime distribution of T is considered, given covariates $\mathbf{x}$. As usual, the standard situation involves independent observations $(t_i, \delta_i, \mathbf{x}_i)$ on a random sample of n individuals, with t_i a lifetime or censoring time and $\delta_i = I(t_i \text{ is a lifetime})$. Censoring is assumed to be independent in the sense of Section 2.2.2; this means that it is permissible for the censoring mechanism to depend on the covariate values.

In situations involving a single quantitative covariate, x, plots of time, t_i, or log time, y_i, against x_i or functions of x_i may indicate the nature of any relationship between T and x. The presence of censoring is a problem, however. If the proportion of censored times is small, then it is usually satisfactory just to indicate which points in the plot correspond to censoring times, by using a separate symbol, but if there is substantial censoring, the opportunity to see a relationship between T and x may be destroyed. When there is more than one quantitative covariate and light censoring, plots of t_i or $\log t_i$ against the individual covariates are usually helpful, though if covariates display substantial association, then the individual plots may not indicate the sort of relationship found by fitting models with more than one covariate. The addition to a plot of smooth trend curves representing means or quantiles is often helpful.

Another useful device is to group individuals so that within groups they have similar values of important covariates. Except when covariates are discrete with a fairly small number of values, this entails judgment and some loss of information. Computation of mean or median lifetimes or log-lifetimes for each group is helpful; the use of medians is preferable in most cases, especially when censoring is present. Graphical tools such as box plots are also valuable, but need modification to deal with censoring (e.g., Gentleman and Crowley 1991). For box plots, the empirical quantiles for a group of individuals may be defined in terms of the Kaplan–Meier estimate of survival for that group, provided a quantile is not beyond the largest observation.

The distributions of lifetime within groups of individuals described in the preceding paragraph can be examined and compared in more detail through nonparametric estimation of survivor, density, or hazard functions, as discussed in Chapter 3, provided there is sufficient data. Plots of Kaplan–Meier estimates $\hat{S}_j(t)$ of the survivor function for each of J groups, $j = 1, \ldots, J$, are often useful. One good procedure is to use Weibull probability plots of $\log[-\log \hat{S}_j(t)]$ (vertical axis) versus $\log t$ (horizontal axis) for each group. Alternatively, plots of Nelson–Aalen estimates, $\log[\hat{H}_j(t)]$ versus $\log t$, can be used. The following points can be noted about such plots:

1. For an accelerated failure time model, it follows from (6.1.4) that

$$S(t|\mathbf{x}) = S_0\left[\frac{\log t - u(\mathbf{x})}{b}\right],$$

and thus

$$\log[-\log S(t|\mathbf{x})] = \log\left\{-\log S_0\left[\frac{\log t - u(\mathbf{x})}{b}\right]\right\}. \tag{6.2.1}$$

Thus if $u(\mathbf{x})$ is approximately constant for individuals within each group $j = 1, \ldots, J$, and if an AFT model is appropriate, then plots of $\log[-\log \hat{S}_j(t)]$ versus $\log t$ should be roughly parallel in the horizontal ($\log t$) direction. Of

course, plots of $\hat{S}_j$ versus $\log t$ should also be parallel, but because of points 2 and 3 below, we usually choose to plot $\log[-\log \hat{S}_j(t)]$.

2. For a proportional hazards model, it follows from (6.1.7) that

$$\log[-\log S(t|\mathbf{x})] = \log[-\log S_0(t)] + \log r(\mathbf{x}), \qquad (6.2.2)$$

so if $r(\mathbf{x})$ is approximately constant within groups $j = 1, \ldots, J$, then plots of $\log[-\log \hat{S}_j(t)]$ versus $\log t$ should be roughly parallel in the vertical direction.

3. If plots of $\log[-\log \hat{S}_j(t)]$ versus $\log t$ are roughly linear, then Weibull lifetime models are suggested. If the plots are also roughly parallel, then (6.1.1) with constant shape parameter δ is suggested; this is both an AFT and a PH model.

Plots of a similar nature can be used to suggest other regression specifications. For example, plots of $\psi[\hat{S}_j(t)]$ versus $\log t$ or t would be useful in connection with models (6.1.11). Some qualifications should be noted about the assessment of parallelism in such plots, however. Parallelism in the vertical direction can be hard or impossible to assess when differences in the location of log-lifetime distributions are large relative to the dispersion of the distributions; right censoring may rule out comparisons at large t values; there is considerable variability in probability plots, as discussed in Section 3.3: the variability of plots is most extreme for small and large t, where the visual impact may be considerable; splitting the data into several small groups leads to plots that are inherently very variable.

The analysis of data is an iterative process involving exploration, model fitting, and model assessment. Model assessment is considered in Section 6.2.2, and model fitting and inference in Sections 6.3–6.5. First, however, we consider some simple illustrations of graphical model exploration.

Example 6.2.1. (Leukemia Survival Times). In an early paper on regression analysis of lifetime data, Feigl and Zelen (1965) gave the data on survival times for 33 leukemia patients shown in Table 6.1. Survival times are in weeks from diagnosis, and there are two covariates: white blood cell count (WBC) at diagnosis and a binary variate AG that indicates a positive ($AG = 1$) or negative ($AG = 0$) test related to white blood cell characteristics. The original data had no censored survival times, but for illustrative purposes three of the lifetimes have here been replaced with censoring times.

A plot of $\log(t_i)$ versus $\log(wbc_i)$ is shown in Figure 6.3, with the symbols P and N denoting individuals with $AG = 1$ and $AG = 0$, respectively, and with the three censoring times (all with $AG = 1$) designated with lowercase p. Two AG-positive subjects have $wbc = 100$ and $t = 1$, so their symbols are overlaid. The decision to plot $\log(t_i)$ versus $\log(wbc_i)$ was based on a preliminary plot of $\log(t_i)$ versus wbc_i and previous experience that indicates that biological counts are frequently best treated as covariates on a log scale. The plot shows considerable variation in survival times for individuals with similar covariate values, but suggests that survival times tend to be shorter for individuals with higher WBC and with AG-negative tests.

Table 6.1. Leukemia Survival Data

t	AG	wbc	t	AG	wbc
65	1	2.3	56	0	4.4
140[a]	1	.75	65	0	3.0
100	1	4.3	17	0	4.0
134	1	2.6	7	0	1.5
16	1	6.0	16	0	9.0
106[a]	1	10.5	22	0	5.3
121	1	10.0	3	0	10.0
4	1	17.0	4	0	19.0
39	1	5.4	2	0	27.0
121[a]	1	7.0	3	0	28.0
56	1	9.4	8	0	31.0
26	1	32.0	4	0	26.0
22	1	35.0	3	0	21.0
1	1	100.0	30	0	79.0
1	1	100.0	4	0	100.0
5	1	52.0	43	0	100.0
65	1	100.0			

[a]Denotes a censoring time; wbc = WBC ÷ 1000.

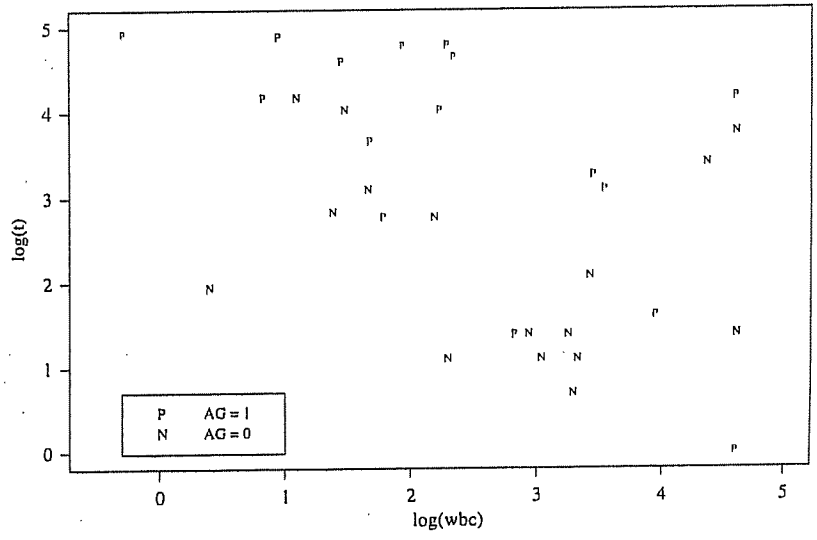

Figure 6.3. Scatter plot for leukemia survival data.

The degree of association between wbc and AG is not strong, and we could also examine graphically the distribution of survival times for the $AG = 0$ and $AG = 1$ groups. However, Figure 6.3 displays most of the salient information in this simple setting. Regression models are fitted to the data in the next section.

Example 6.2.2. (Insulating Fluid Failure Times). Example 1.1.5 gave data on the time-to-failure of electrical insulating fluid, subjected to various high-voltage levels in a life test experiment. Seventy-six insulating-fluid specimens were tested at seven voltage levels, v = 26, 28, 30, 32, 34, 36, and 38 kilovolts (kV); the data are given in Table 1.1. Primary objectives of analysis are to relate failure time to voltage and to obtain a model that could be used for extrapolation to lower voltages.

Engineering background for this problem suggests what is referred to as a power law model. This is an AFT model (6.1.5) where the scale parameter α is related to the voltage v by $\alpha(v) = cv^p$. In terms of log failure time Y and the equivalent location-scale model (6.1.3), we have $u(v) = \beta_0 + \beta_1 \log v$, where $\beta_0 = \log c$ and $\beta_1 = p$. The engineering background further suggests that a Weibull AFT model is appropriate. In Example 5.4.1 a test of constancy of shape parameters across the seven voltage levels was carried out, under the assumption of a Weibull model.

Figure 6.4 shows plots of $\log[-\log \hat{S}_j(t)]$ versus $\log t$, where $\hat{S}_1(t), \ldots, \hat{S}_7(t)$ are the Kaplan–Meier estimates at voltage levels 26, 28, ..., 38 kV, respectively. We have used the adjusted Kaplan–Meier values (3.3.1) in the plots, as described in Example 3.3.2. This shows a point for every failure time. To keep the plot uncrowded, only voltage levels 26, 30, 34, and 38 kV are represented. The one small failure time

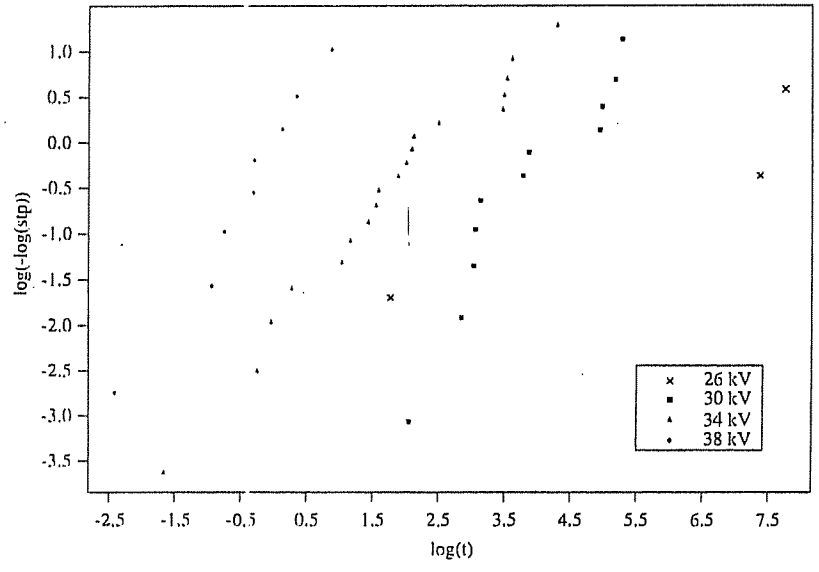

Figure 6.4. Weibull probability plots of electrical-insulation failure times at different voltage levels.

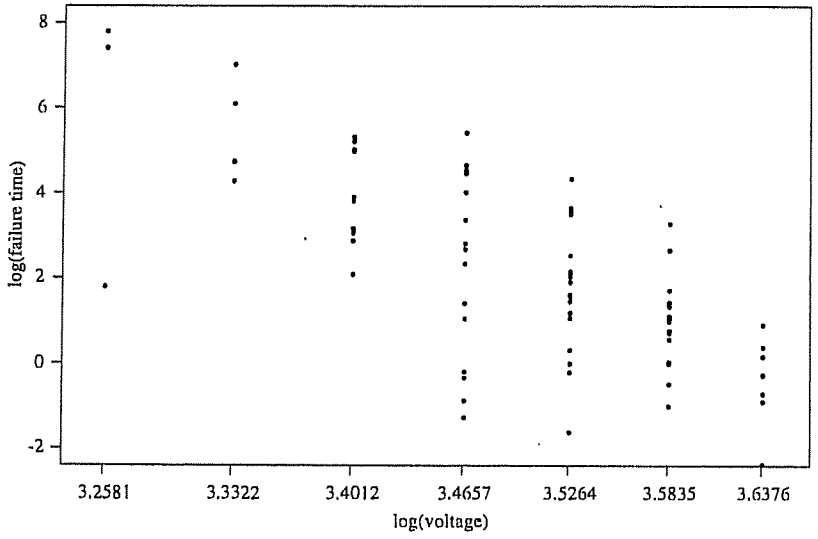

Figure 6.5. Scatter plot for electrical-insulation failure data.

at 26 kV should be noted as a possible outlier. Aside from this point, the plots are still not parallel, but are roughly so; they are also roughly linear. Bearing in mind that each plot is based on a small sample of failure times, we might consider Figure 6.4 broadly consistent with a Weibull AFT model. Formal hypothesis tests are indeed unable to contradict this model: Example 5.4.1 provides a test of the constancy of shape parameters, and Example 6.3.1 examines the power law and Weibull assumptions.

Figure 6.4 shows that with small data sets it can be difficult to infer from plots whether the data are consistent with a specific type of regression model (in this case, an AFT model, perhaps of Weibull form), even when there is only a single covariate. When the data are stratified into small groups according to covariate values, there is considerable variability in nonparametric estimates such as $\hat{S}_j(t)$, and it is asking too much to expect a clear picture to emerge. It may, however, be quite clear that there is a relationship between a covariate and certain characteristics of failure time. For example, Figure 6.4 clearly suggests that, although the failure time distributions at different voltage levels might overlap, the median failure time $t_{.50}$ or log failure time $y_{.50}$ decreases as voltage increases. Furthermore, a plot of $y_i (= \log t_i)$ versus $\log v_i$ $(i = 1, \ldots, 76)$, shown in Figure 6.5, suggests a roughly linear relationship between $y_{.50}$ and log voltage. Box plots of the log failure times at each voltage level would display similar information.

Example 6.2.3. (Times to Pulmonary Exacerbation). Censoring was not a serious factor in the preceding two examples. We now consider data on times to a

pulmonary exacerbation for persons with cystic fibrosis, which were introduced in Example 1.1.8 and subsequently examined in Examples 3.2.4 and 3.2.5. The data are from a clinical trial in which patients were randomly assigned to receive daily doses of an experimental treatment, rhDNase, or a placebo. Patients were followed for approximately 169 days, and the response considered here is the time T_i (in days) until the occurrence of a pulmonary exacerbation, or infection. Only 104 out of 321 rhDNase subjects and 139 out of 324 Placebo subjects had exacerbations during the study; for the remaining individuals T_i was right-censored at approximately 169 days, aside from a few persons who withdrew early.

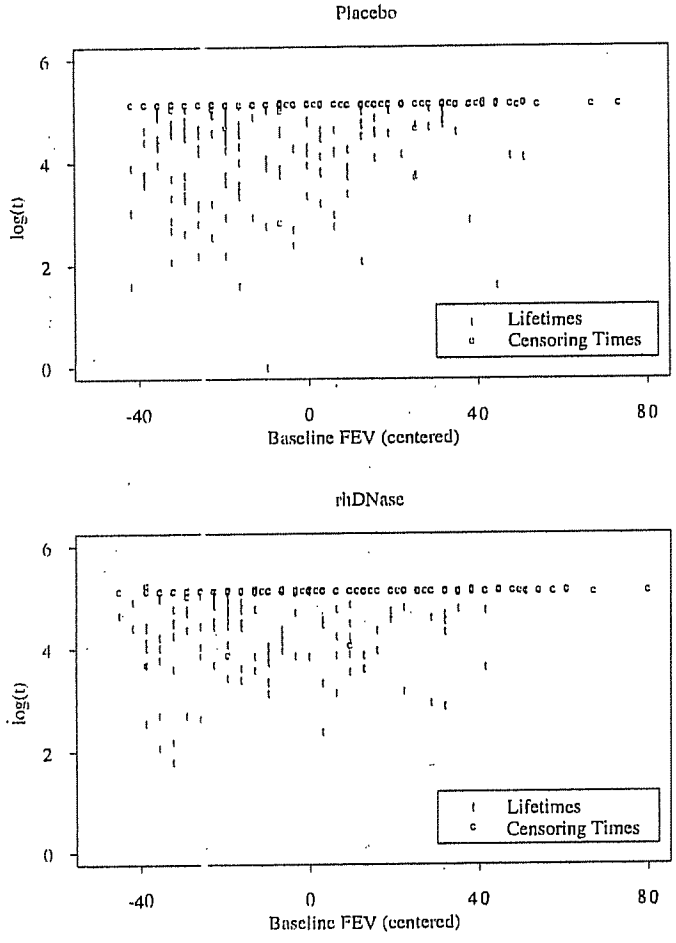

Figure 6.6. Scatter plots for time to pulmonary exacerbation data.

Figure 3.4 showed the Kaplan–Meier estimates of survival (no exacerbation) for each treatment group. The estimates are well separated and suggest that either an AFT or PH model to represent the treatment effect might be reasonable. A continuous covariate, forced expiratory volume (fev), was also measured on each subject at the start of the study; this is a measure of lung function, and one might expect higher fev to be associated with larger survival times. Figure 6.6 shows a plot of $y_i = \log t_i$ versus centered fev_i ($fevc_i$) for each treatment group, with lifetimes and censoring times represented by different symbols. This is an honest plot in the sense that it shows all of the available data. It suggests a positive association between T_i and fev_i, but the high degree of censoring makes it very difficult to characterize. The plot could be enhanced slightly by indicating numbers of censoring times at different fev values, since the overlaying of symbols makes this impossible to discern from the graph, but this does not help much in assessing the relationship between T_i and fev_i.

Plots that provide a bit more insight can be based on grouping subjects according to treatment and fev. For example, we could split the subjects in the two treatment groups into three subgroups by classifying their fev values as low, medium, or high. Figure 6.7 shows a transformed plot of the six Kaplan–Meier estimates of survival, $\hat{S}_j(t)$, based on subgroups of approximately equal sizes. Following the discussion preceding (6.2.1), we first plotted $\log[-\log \hat{S}_j(t)]$ versus $\log t$. A subsequent plot of $F_0^{-1}(1 - \hat{S}_j(t))$ versus $\log t$, where $F_0(z)$ is the standard normal cumulative distribution function (c.d.f.) and $F_0^{-1}(p)$ is the associated quantile function, gave more linear plots and is shown in Figure 6.7. This plot is somewhat "busy," but indicates

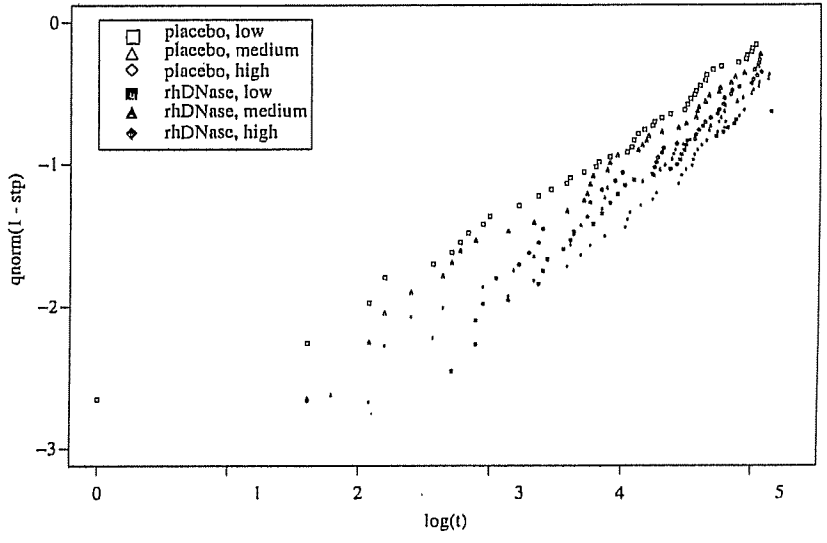

Figure 6.7. Normal probability plots for 6 treatment–fev groups for pulmonary exacerbation data.

the existence of a relationship between T_i and (trt_i, fev_i) which is reasonably consistent with a log-location-scale model (6.1.4) in which Z is normally distributed.

An alternative to Figure 6.7 that is less busy but sacrifices some information, is to show box plots or similar figures representing sample quantiles of lifetimes for the six subject groups. The sample quantiles are defined in terms of the Kaplan–Meier estimates, $\hat{S}_j(t)$, in view of the censoring; the pth quantile for a given group is the value t_p satisfying $\hat{S}_j(t_p) = 1 - p$. For some values of p there may be no such t_p, depending on the degree of censoring. In this example none of the $\hat{S}_j(t)$ $(j = 1, \ldots, 6)$ reaches .5, so only quantiles with $p < .5$ are available. Figure 6.8 shows a plot of estimated quantiles $\tilde{y}_p$ for $Y = \log T$. Sample quantiles for $p = .1$ and .3 are displayed. The estimates $\tilde{y}_{.10}$ and $\tilde{y}_{.30}$ in the plot are roughly an equal distance apart, consistent with a location-scale model for Y. A treatment effect is suggested and, at least for the Placebo group, an effect due to fev.

Confidence limits could be added to the points in the figure. An alternative graphical approach to Figure 6.8 would be to impose nonparametric quantile regression curves on scatter plots, as in Figure 6.6. This shows the actual data and avoids the grouping inherent in Figure 6.8, but requires some form of smoothing; see Gentleman and Crowley (1991) and Bowman and Wright (2000). Note also that Figure 6.8 summarizes only some of the information in Figure 6.6; it does not, for example, show that the fraction of persons experiencing an exacerbation within the 169-day follow-up period tends to decrease as fev increases.

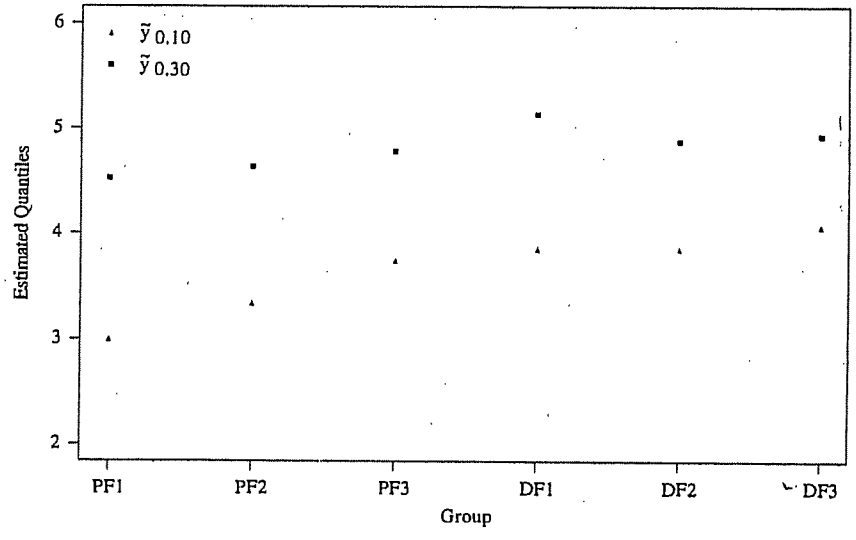

Figure 6.8. Estimated .10 and .30 quantiles for time to first pulmonary exacerbation, for 6 treatment–fev groups.

As always, it is important to remember that these plotting procedures have varying and substantial degrees of inherent variability. Formal model fitting and analysis of these data are considered in subsequent examples.

6.2.2 Assessment of Fitted Models

Once a regression model has been fitted, it is important to assess the assumptions underlying the model in the face of the observed data, and to check on the sensitivity of conclusions to changes in the model or the data. Although model fitting and related inference procedures are not discussed until Section 6.3, let us consider some general aspects of model assessment.

Plots described in the preceding section help to indicate whether certain models are plausible, and to identify extreme or influential data points $(t_i, \delta_i, \mathbf{x}_i)$ or $(y_i, \delta_i, \mathbf{x}_i)$. Once models have been fitted, several approaches to checking them can be considered. Model expansion, which involves adding parameters that represent specific types of departures from the current model, is very important; the need for the extra parameters can be assessed via hypothesis tests. Some examples of model expansion are: (1) adding covariates representing interactions or nonlinear terms to check a linear model; (2) allowing b in a location-scale model (6.1.4) to depend on $\mathbf{x}$, as a check on the constancy of b; (3) building time–covariate interactions into (6.1.6) as a check on the PH assumptions; (4) expanding the family of models $S_0(z)$ for Z in (6.1.4), as a check on the assumed error distribution.

Model expansion techniques are considered in subsequent sections of this chapter. Other model checking procedures based on hypothesis testing are considered in Section 10.4. The remainder of this section deals with graphical methods of model assessment, based on the examination of residuals and influence statistics.

6.2.2.1 Residual Analysis

Residual analysis is widely treated in books on regression (e.g., McCullagh and Nelder 1993; Weisberg 1985), though only uncensored data are usually considered. Residuals can be defined in various ways, but the key idea is that if a model for the distribution of Y given $\mathbf{x}$, specified in terms of a parameter $\boldsymbol{\theta}$, is fitted to independent data $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$, then residuals $\hat{e}_i = g(y_i, \mathbf{x}_i, \hat{\boldsymbol{\theta}})$ should have specific properties if the model is correct. In most settings we seek residuals $\hat{e}_1, \ldots, \hat{e}_n$ that are approximately independent and identically distributed when the model is correct. The following discussion incorporates censoring into the definition of residuals.

Let us first consider AFT or equivalent location-scale models (6.1.3) or (6.1.4), where $u(\mathbf{x})$ depends on a parameter vector $\boldsymbol{\beta}$. Let $(y_i, \delta_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ represent a censored random sample from the distributions of Y_i given $\mathbf{x}_i$; the $\mathbf{x}_i$ can be either random or fixed by the study design, but as usual in regression analysis we condition on their observed values. For convenience, write u_i for $u(\mathbf{x}_i; \boldsymbol{\beta})$ in (6.1.3) and let $\hat{\boldsymbol{\beta}}, \hat{b}$ be the maximum likelihood estimate (m.l.e.) obtained by fitting the data $(y_i, \delta_i, \mathbf{x}_i)$ to (6.1.3). The standardized variables $Z_i = (Y_i - u_i)/b$ are independent and identically distributed (i.i.d.) with survivor function $S_0(z)$ under (6.1.3), and the standard way

to define residuals is as

$$\hat{z}_i = \frac{y_i - \hat{u}_i}{\hat{b}}, \qquad i = 1, \ldots, n, \tag{6.2.3}$$

where $\hat{u}_i = u(\mathbf{x}_i; \hat{\boldsymbol{\beta}})$. If there is no censoring, the $\hat{z}_i$ should, at least for large n, appear like a random sample from $S_0(z)$, and therefore also be independent of the $\mathbf{x}_i$. Because they depend on $\hat{\boldsymbol{\beta}}$ and $\hat{b}$, the $\hat{z}_i$ are not of course exactly independent or identically distributed with distribution $S_0(z)$. Adjustments to the definition of $\hat{z}_i$ are sometimes considered, with the objective of making them look more like observations from $S_0(z)$. This is mainly of concern when n is small and precise model assessment is difficult.

Plots of the $\hat{z}_i$ against covariates or other factors such as the fitted values $\hat{u}_i$ can be used to check on the constancy of b in (6.1.3) or to look for systematic departures from the assumed specification $u(\mathbf{x}; \boldsymbol{\beta})$. Probability or other plots of the $\hat{z}_i$ can be used to assess the baseline distribution $S_0(z)$. Departures from the location-scale form itself are harder to detect; this is best approached through model expansion, or the graphical methods in Section 6.2.1.

The presence of censoring means that y_i is either a log-lifetime or log censoring time. In this case one can argue that if the model is correct, the $\hat{z}_i$ in (6.2.3) should appear roughly like a censored random sample from $S_0(z)$ and employ probability plots as discussed in Section 3.3. If censoring is light, then plots of $\hat{z}_i$ versus covariates are still useful, but censored and uncensored residuals should be designated with different symbols. However, as censoring becomes heavier, the usefulness of such plots is severely compromised. The essential problem is that when a censoring mechanism is operating, the distribution of y_i is not given by (6.1.3), but depends on the censoring process. For example, if log-lifetime Y_i has distribution (6.1.3) conditional on $\mathbf{x}_i$, and $\log C_i$ is a fixed log censoring time, then $Y_i^* = \min(Y_i, \log C_i)$ has distribution function

$$F^*(y_i) = \begin{cases} F_0\left[\dfrac{y_i - u_i}{b}\right] & y_i \le \log C_i \\ 1 & y_i > \log C_i \end{cases},$$

where $F_0(z) = 1 - S_0(z)$. That is, the possibly censored observation y_i is a realization of Y_i^*, which does not have distribution (6.1.3) and, indeed, has a mass of probability at $\log C_i$. Correspondingly, $z_i = (y_i - u_i)/b$ is a realization of $Z_i^* = (Y_i^* - u_i)/b$ and has a mass of probability at $(\log C_i - u_i)/b$.

Two approaches can be considered for dealing with plots of residuals versus covariates when censoring is substantial. One is to use adjusted residuals, for which $\hat{z}_i$ corresponding to censored observations are adjusted upwards. These are usually defined for location-scale models as

$$\hat{z}_i^{adj} = \delta_i \hat{z}_i + (1 - \delta_i) E(Z_i | Z_i \ge \hat{z}_i), \tag{6.2.4}$$

where $E(Z_i | Z_i \ge \hat{z}_i)$ denotes the expected value based on the distribution $S_0(z)$ for Z_i. This makes intuitive sense; we know that $\hat{z}_i$ for a censored observation is a lower bound on the true residual. It can also be noted that if $Z_i = (Y_i - u_i)/b$ and $R_i = (\log C_i - u_i)/b$, then

$$Z_i^{adj} = \delta_i Z_i + (1 - \delta_i) E(Z_i | Z_i \ge R_i)$$

has expectation

$$\begin{aligned} E(Z_i^{adj}) &= \int_{-\infty}^{R_i} z f_0(z)\, dz + S_0(R_i) \int_{R_i}^{\infty} \frac{z f_0(z)}{S_0(R_i)}\, dz \\ &= \int_{-\infty}^{\infty} z f_0(z)\, dz \\ &= E(Z_i). \end{aligned}$$

Therefore the adjusted residuals (6.2.4) should behave roughly like the Z_i in expectation, and "smooths," or nonparametric regression mean curves passed through plots of $\hat{z}_i^{adj}$ versus covariates, should be roughly constant and equal to $E(Z_i)$ if the model is correct. It is important to note, however, that the distribution of the adjusted residuals depends on the censoring times and may be very unlike the distribution of Z_i. In particular, it may have probability mass on certain values or along certain curves when censoring times are roughly equal, as illustrated in the continuation of Example 6.2.3 that follows. An inherent problem is also that the adjusted residuals (6.2.4) are computed using the model that we are trying to check, and if censoring is heavy, then the ability of the plots to detect model inadequacies is very limited.

A second approach that is useful when most censoring times are large is to pass nonparametric quantile regression curves $\tilde{y}_p(x_j)$ through plots of $\hat{z}_i$ versus specific covariates x_{ij}, as discussed in Example 6.2.3. If few of the smaller $\hat{z}_i$ are from censored observations, then it will at least be possible to obtain curves for smaller values of p. These should be roughly constant if the model (6.1.4) is correct, and approximately equal to the pth quantiles of Z in (6.1.4).

Residuals can be defined for arbitrary lifetime regression models, not necessarily of location-scale form, by considering quantities

$$e_i = g_i(T_i, \mathbf{x}_i, \boldsymbol{\theta}) \tag{6.2.5}$$

that are i.i.d. and whose distribution is known, given $\mathbf{x}_1, \ldots, \mathbf{x}_n$. In (6.2.5), T_i is the lifetime and $\boldsymbol{\theta}$ is the vector of parameters that specifies the model in question. If $\hat{\boldsymbol{\theta}}$ is the m.l.e. of $\boldsymbol{\theta}$ determined from complete data $(t_i, \mathbf{x}_i)$, $i = 1, \ldots, n$, then residuals $\hat{e}_i$ are defined by

$$\hat{e}_i = g_i(t_i, \mathbf{x}_i, \hat{\boldsymbol{\theta}}) \tag{6.2.6}$$

and for large n behave approximately like a random sample of size n from the distribution of the e_i. When the lifetimes T_i are subject to censoring then, as described

above, we can use the $\hat{e}_i$ as defined or choose to employ adjusted residuals for observations that are censored.

Quantities e_i are readily available with continuous models; for example, the probability integral transform values $e_i = F(T_i|\mathbf{x}_i; \boldsymbol{\theta})$ or $e_i = S(T_i|\mathbf{x}_i; \boldsymbol{\theta})$ have uniform distributions on (0, 1). An equivalent quantity that is especially convenient with censored data is the cumulative hazard transform $e_i = H(T_i|\mathbf{x}_i; \boldsymbol{\theta})$, which gives residuals

$$\hat{e}_i = H(t_i|\mathbf{x}_i; \hat{\boldsymbol{\theta}}) \qquad i = 1, \ldots, n. \tag{6.2.7}$$

Since $H(T_i|\mathbf{x}_i; \boldsymbol{\theta}) = -\log S(T_i|\mathbf{x}_i; \boldsymbol{\theta})$, these e_i are independent standard exponential random variables, so the $\hat{e}_i$ should behave approximately like a censored sample of standard exponential variables if the model is appropriate. Adjustments to these residuals are straightforward since if $e_i \sim \text{Exp}(1)$, then $E(e_i|e_i \geq \hat{e}_i) = \hat{e}_i + 1$. Thus it is customary to define adjusted exponential residuals as

$$\hat{e}_i^{adj} = H(t_i|\mathbf{x}_i; \hat{\boldsymbol{\theta}}) + 1 - \delta_i. \tag{6.2.8}$$

The residuals $\hat{e}_i$ or $\hat{e}_i^{adj}$ can be used in the ways previously described for the residuals $\hat{z}_i$ in location-scale models, and are subject to the same problems when censoring is heavy. Since the exponential distribution and the distribution of $\hat{e}_i$ or $\hat{e}_i^{adj}$ values is defined on $(0, \infty)$ and very skewed, it is often preferable to use equivalent extreme value residuals $\hat{z}_i = \log(\hat{e}_i)$ or $\hat{z}_i^{adj} = \log(\hat{e}_i^{adj})$, which should look roughly like a random sample from the standard extreme value distribution.

Example 6.2.4. For an AFT model (6.1.4) the distribution of $Y_i = \log T_i$, given $\mathbf{x}_i$ is of location-scale form (6.1.3). The cumulative hazard function for T_i is $H(t|\mathbf{x}_i) = -\log S(t|\mathbf{x}_i) = -\log S_0[(y - u_i)/b]$, where $y = \log t$, and so the exponential residuals (6.2.7) are equal to

$$\begin{aligned} \hat{e}_i &= -\log S_0[(y_i - \hat{u}_i)/\hat{b}] \\ &= -\log S_0(\hat{z}_i), \end{aligned} \tag{6.2.9}$$

where $\hat{z}_i$ is the residual (6.2.3). The adjusted residuals given by (6.2.4) and (6.2.8) are different; the latter gives

$$\hat{e}_i^{adj} = -\log S_0(\hat{z}_i) + 1 - \delta_i,$$

and the corresponding extreme value residuals are

$$\hat{z}_i^{adj} = \log[-\log S_0(\hat{z}_i) + 1 - \delta_i]. \tag{6.2.10}$$

An advantage of (6.2.10) is the closed-form expression; the residuals (6.2.4), on the other hand, require numerical integration for some distributions of Z.

6.2.2.2 *Influence Analysis*

In some cases a small number of points may have a strong influence on the fitted model. The topic of influence analysis is well developed for linear models with uncensored data (e.g., Cook and Weisberg 1982; Atkinson 1985), but more difficult with censored lifetime data. The main diagnostics for influence are based on a comparison of $\hat{\boldsymbol{\theta}}$ from the full data and the m.l.e. $\hat{\boldsymbol{\theta}}_{(-g)}$, with a specified group g consisting of one or more observations dropped. When g consists of just the ith observation, it is customary to assess the global influence of the observation through $\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(-i)}$ or the associated likelihood-drop (LD) statistic

$$LD_i = 2\ell(\hat{\boldsymbol{\theta}}) - 2\ell(\hat{\boldsymbol{\theta}}_{(-i)}). \tag{6.2.11}$$

This can be calibrated by reference to $\chi^2_{(p)}$ quantiles, where p is the dimension of $\boldsymbol{\theta}$. In particular, we can see whether $\hat{\boldsymbol{\theta}}_{(-i)}$ lies inside the q confidence region for $\boldsymbol{\theta}$ given by $\{\boldsymbol{\theta}: 2\ell(\hat{\boldsymbol{\theta}}) - 2\ell(\boldsymbol{\theta}) \leq \chi^2_{(p),q}\}$. The influence on individual parameters $\psi = g(\boldsymbol{\theta})$ is often measured by

$$\Delta\psi_i = \frac{\hat{\psi} - \hat{\psi}_{(-i)}}{se(\hat{\psi})}. \tag{6.2.12}$$

It is generally infeasible to compute (6.2.11) or (6.2.12) for all observations in a large data set, so approximations have been developed. Mostly these are obtained by first- or second-order Taylor series expansion of the score function $\mathbf{U}(\boldsymbol{\theta}) = \partial\ell/\partial\boldsymbol{\theta}$ or log-likelihood $\ell(\boldsymbol{\theta})$ about $\hat{\boldsymbol{\theta}}$. For example, letting $\mathbf{U}_{(-i)}(\boldsymbol{\theta})$ denote $\mathbf{U}(\boldsymbol{\theta}) - \mathbf{U}_i(\boldsymbol{\theta})$, the score function with observation i dropped, we have

$$\mathbf{U}_{(-i)}(\boldsymbol{\theta}) \doteq \mathbf{U}_{(-i)}(\hat{\boldsymbol{\theta}}) - I_{(-i)}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).$$

Since $\mathbf{U}_{(-i)}(\hat{\boldsymbol{\theta}}_{(-i)}) = \mathbf{0}$ and $\mathbf{U}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, this gives

$$\hat{\boldsymbol{\theta}}_{(-i)} - \hat{\boldsymbol{\theta}} \doteq -I_{(-i)}(\hat{\boldsymbol{\theta}})^{-1}\mathbf{U}_i(\hat{\boldsymbol{\theta}}). \tag{6.2.13}$$

A drawback of this approximation is the need to evaluate and invert the information matrix $I_{(-i)}(\hat{\boldsymbol{\theta}})$ with observation i dropped, and instead of (6.2.13) the alternative first-order approximation

$$\hat{\boldsymbol{\theta}}_{(-i)} - \hat{\boldsymbol{\theta}} \doteq -I(\hat{\boldsymbol{\theta}})^{-1}\mathbf{U}_i(\hat{\boldsymbol{\theta}}) \tag{6.2.14}$$

is typically used. This approximation is usually accurate enough to indicate which observations have the largest impact when deleted. A corresponding approximation to LD_i of (6.2.11) can be obtained from the quadratic approximation to $\ell(\boldsymbol{\theta})$ around $\hat{\boldsymbol{\theta}}$:

$$LD_i \doteq (\hat{\boldsymbol{\theta}}_{(-i)} - \hat{\boldsymbol{\theta}})' I(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_{(-i)} - \hat{\boldsymbol{\theta}}). \tag{6.2.15}$$

A practical guideline is that observations with especially large or small uncensored lifetimes, or with large right-censoring times, are potentially influential; their residuals are also often large. Their actual influence depends as well on whether their covariate value $\mathbf{x}_i$ is extreme, and on the effects of other observations. For AFT models for which $u(\mathbf{x}_i)$ in (6.1.3) is of the form $\boldsymbol{\beta}'\mathbf{x}_i$, the leverage values

$$h_{ii} = \mathbf{x}_i'(X'X)^{-1}\mathbf{x}_i, \tag{6.2.16}$$

where X is the $n \times p$ matrix whose ith row is the vector $\mathbf{x}_i'$, is a good measure of the extremeness of $\mathbf{x}_i$. This is also true for other models in which the covariates affect lifetime through the linear form $\boldsymbol{\beta}'\mathbf{x}$. An alternative approach to using approximations such as (6.2.14) or (6.2.15) is therefore to identify observations for which residuals and leverage values (6.2.16) are fairly large. Exact values of measures such as (6.2.11) or (6.2.12) can then be obtained by refitting models with certain individual observations dropped. Caveats for all methods are that influential observations may affect $\hat{\boldsymbol{\theta}}$ in such a way that their residuals do not appear extreme, and that a small number of observations that are highly influential as a group may not show up when observations are dropped one at a time.

In some settings certain observations might be considered as outliers, or values $(y_i, \delta_i, \mathbf{x}_i)$ that do not conform with the bulk of the data. Such values can arise because of errors made in recording data, but they can also occur because a small portion of the data is governed by a different process then the rest, or because the distributions of $Y|\mathbf{x}$ or of $\mathbf{x}$ are very long-tailed. The extent to which observations might be considered as outliers is reflected in their residuals and leverage values (6.2.16); better measures are given by deletion residuals, in which, for example, (6.2.6) uses the estimate $\hat{\boldsymbol{\theta}}_{(-i)}$ in place of $\hat{\boldsymbol{\theta}}$, and deletion leverage values (6.2.16), in which we replace $X'X$ with $X'_{(-i)}X_{(-i)}$.

6.2.2.3 *Examples*

Example 6.2.1. (continued). Examination of the leukemia remission time data of Example 6.2.1 suggests that AFT models could be satisfactory. Here we consider some diagnostic checks on a Weibull–extreme value model, in which (6.1.3) is used for $Y = \log T$ with $S_0(z) = \exp(-e^z)$ and $u(x; \boldsymbol{\beta}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, where $x_1 = AG$ and $x_2 = \log(wbc)$. In other words, in the form (6.1.4), the model is

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + bZ,$$

where Z has a standard extreme value distribution. Model fitting and inference for this model are discussed in Section 6.3.2; we merely note here that the m.l.e.'s, with standard errors in brackets, are $\hat{\beta}_0 = 3.841(.534)$, $\hat{\beta}_1 = 1.177(.427)$, $\hat{\beta}_2 = -.366(.150)$ and $\hat{b} = 1.119(.164)$.

Figure 6.9 shows in its left panel a plot of (unadjusted) residuals $\hat{z}_i$ given by (6.2.3) against values x_{2i} of $\log(wbc)$. No trend creating doubt about the model is observed. Note when assessing the plot that the standard extreme value distribution is skewed to the left (see Figure 1.5) and that $E(Z) = -.577$. The right panel of Figure 6.9 shows a probability plot of the residuals $\hat{z}_i$, designed to check on the regression specification combined with the assumption of the extreme value distribution for Z. This is obtained by treating the $\hat{z}_i$ as a censored sample of 33 log-lifetimes, and using the probability plot based on the values (3.3.1) in Section 3.3. To do this, we compute the Kaplan–Meier estimate $\hat{S}(z)$ based on the $\hat{z}_i$ and plot the points

$$(\hat{z}_i, \log(-\log \hat{w}_i)),$$

where $\hat{w}_i = .5\hat{S}(\hat{z}_i) + .5\hat{S}(\hat{z}_i+)$, for the residuals that correspond to uncensored observations. The plot is roughly linear and provides no evidence against the model.

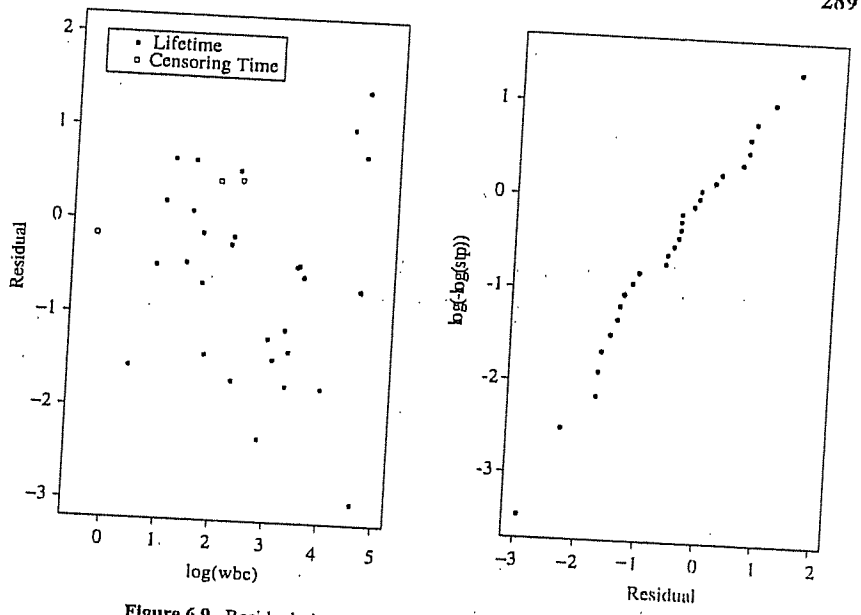

Figure 6.9. Residual plots for Weibull–EV model for leukemia data.

We note from the plot of the data in Figure 6.3 that there are four observations that appear quite influential; three points have high WBC's and long remission times, and one has a low WBC and a small remission time. The four points are those with $(t_i, wbc_i) = (65, 100.0)$, $(7, 1.5)$, $(30, 79.0)$, and $(43, 100.0)$. Individually their omissions have quite substantial effects; for example, values for (6.2.11) are 11.62, 7.79, 10.97, and 15.04, respectively. The fourth point is especially influential. Noting that $\chi^2_{(4),.995} = 14.86$, we see that its deletion moves the m.l.e. of $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, b)$ outside the joint .995 confidence region for $\boldsymbol{\theta}$ based on the full data. This observation also gives values of (6.2.12) with $\psi = \beta_1$ and β_2 of .51 and .83, respectively. If all four observations were omitted, then the effect of WBC would be increased significantly. There is no reason to isolate these observations in the present setting, but it is useful to note their influence.

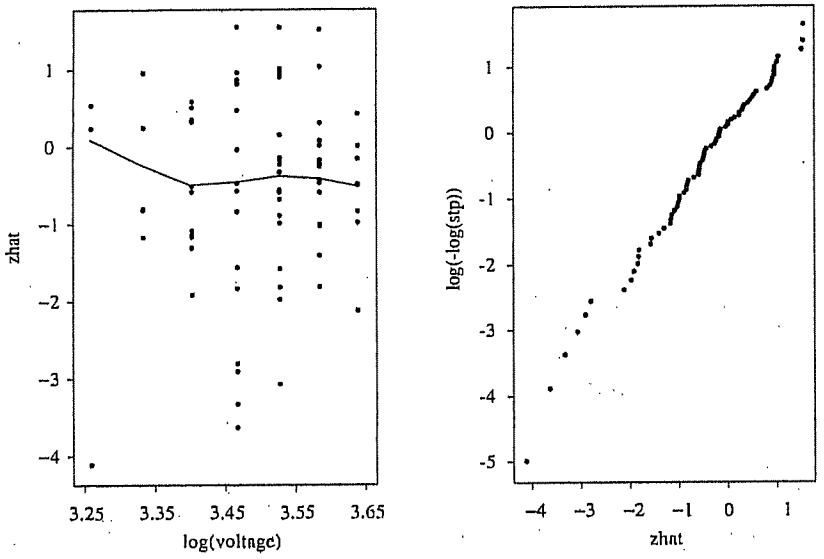

Figure 6.10. Residual plots for Weibull–EV model for electrical-insulation data.

Example 6.2.2. (continued). Plots of the electrical insulating fluid failure time data in Example 6.2.2 showed an approximately linear relationship between log failure times Y and log voltage, $x = \log v$, and as well that an AFT or log-location-scale model (6.1.3) might be reasonable. Figure 6.10 shows plots of residuals $\hat{z}_i$ given by (6.2.3) for the model

$$Y = \beta_0 + \beta_1 x + bZ,$$

where Z has a standard extreme value distribution. The m.l.e.'s, found later in Example 6.3.2, are (with standard errors in brackets) $\hat{\beta}_0 = 64.8(5.62)$, $\hat{\beta}_1 = -17.7(1.61)$, and $\hat{b} = 1.29(.11)$.

The left-hand panel in Figure 6.10 is a plot of $\hat{z}_i$ versus x_i, with a mean curve provided by a local linear smoother. This gives more or less the same picture as the plot of y_i versus x_i in Figure 6.5, and there is no reason to doubt the model. Note that variation in the number of items tested at the different voltages creates variation in the dispersion of residuals, but that the mean curve, or curves passing through median points or other quantiles at each voltage, are roughly horizontal. One small failure time at $v = 26$ stands out as unusual and should be noted. It also has the lowest value of x seen in the data, so is potentially influential. Dropping it and reestimating the parameters β_0, β_1, and b gives values for (6.2.12) of .30, −.30, and −.32 for $\hat{\psi} = \hat{\beta}_0$, $\hat{\beta}_1$, $\hat{b}$, respectively, so it is only moderately influential.

The right hand panel of Figure 6.10 is a Weibull probability plot of the $\hat{z}_i$, produced in the way described in the preceding example; the plot is satisfactorily linear

and provides no evidence against the model. The six smallest residuals stand out slightly, but this is quite consistent with the degree of variation expected in such plots under the model.

Example 6.2.3. (continued). The data in Example 6.2.3 on the times to a first pulmonary exacerbation for subjects in a randomized clinical trial are heavily censored, but the discussion in Example 6.2.3 suggests that log-location-scale models with normal errors are reasonable. Such a model is later fitted to the data in Example 6.3.4; it is of the form

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \sigma Z,$$

where Y is log failure time, $x_1 = I$ (treatment = rhDNase), $x_2 = fevc$, a centered form of fev, and Z has a standard normal distribution. The m.l.e.'s and standard errors for the parameters are $\hat{\beta}_0 = 5.40(.10)$, $\hat{\beta}_1 = .430(.137)$, $\hat{\beta}_2 = .022(.003)$, and $\hat{\sigma} = 1.45(.074)$; further details of the fitted model are given in Example 6.3.4.

Figure 6.11 shows a normal probability plot of the residuals $\hat{z}_i$ given by (6.2.3). This plot is based on the Kaplan–Meier estimate, $\hat{S}(z)$, from these residuals and consists of the points

$$(\hat{z}_i, \Phi^{-1}(1 - \hat{w}_i)),$$

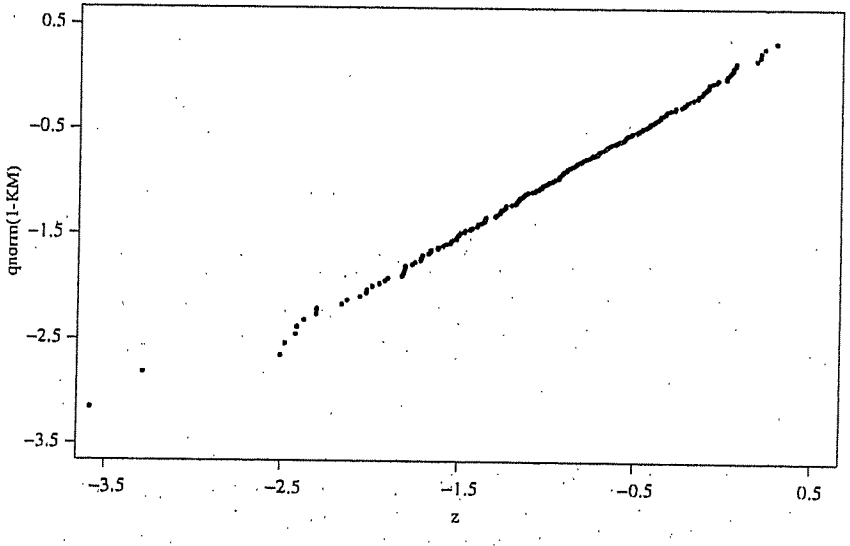

Figure 6.11. Normal probability plot of residuals for pulmonary exacerbation data.

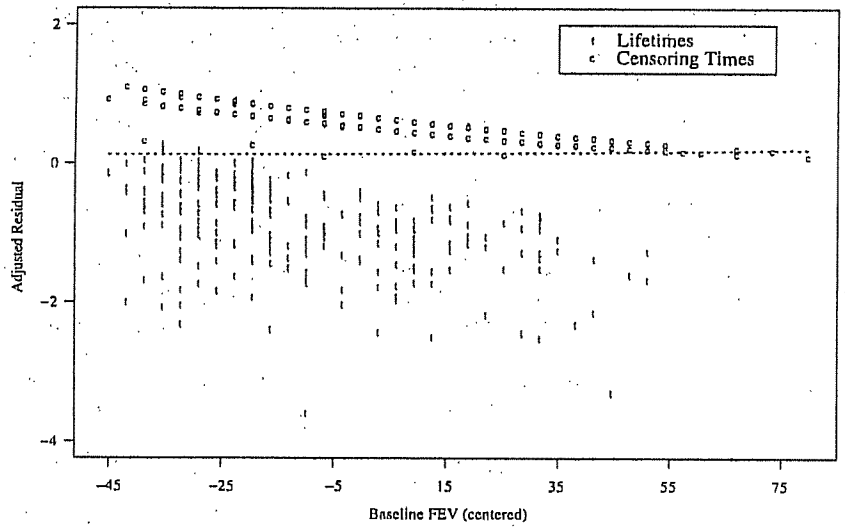

Figure 6.12. Plot of adjusted normal residuals versus fev for pulmonary exacerbation data.

where $\hat{w}_i = .5\hat{S}(\hat{z}_i) + .5\hat{S}(\hat{z}_i+)$ for the residuals that correspond to uncensored observations. The plot is very close to linear and does not suggest any problem with the model.

Figure 6.12 is a plot of adjusted (adj) residuals $\hat{z}_i^{adj}$, given by (6.2.4) and the expression (6.3.21) from Section 6.3, against the fev covariate x_{2i}, with different symbols denoting censored and uncensored failure times. A pronounced pattern from the censored residuals is seen. This is because most of the censoring times t_i are close to 169 days, so $\hat{z}_i = (\log 169 - \hat{u}_i)/\hat{\sigma}$ for these individuals. The estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\sigma}$ given earlier then give approximately $\hat{z}_i = -.19 - .015x_{2i}$ if $x_{1i} = 0$, and $\hat{z}_i = -.48 - .015x_{2i}$ if $x_{1i} = 1$. Thus when we plot either $\hat{z}_i$ or $\hat{z}_i^{adj}$ versus x_{2i} for the censored observations, we get the approximate parallel line patterns shown in Figure 6.12; the use of adjusted residuals merely introduces a little curvature. An estimated mean curve obtained from a scatterplot smoother has been added to the plot. It is close to the zero line, and the plot as a whole gives no reason to doubt the model.

6.3 INFERENCE FOR LOG-LOCATION-SCALE (ACCELERATED FAILURE TIME) MODELS

6.3.1 Likelihood Methods

Assume that a censored random sample consisting of data $(y_i, \delta_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ is available, where $y_i = \log t_i$ is a log-lifetime or log censoring time according

to whether $\delta_i = 1$ or 0, respectively. This section describes maximum likelihood estimation and related inference procedures for a general location-scale model for Y of the form (6.1.3) or (6.1.4), but with $u(\mathbf{x}_i; \boldsymbol{\beta})$ given by the linear specification

$$u(\mathbf{x}_i; \boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{x}_i, \tag{6.3.1}$$

where $\boldsymbol{\beta}$ and $\mathbf{x}_i$ are $p \times 1$ vectors.

Most regression models (6.3.1) have an intercept term, so that $\boldsymbol{\beta}'\mathbf{x}_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1}x_{i,p-1}$. It is generally better for accurate computation to center (i.e., choose the origin for) quantitative covariates so that their means are close to zero. The choice of origin for a covariate only affects the definition of the intercept β_0, and does not affect $\beta_1, \ldots, \beta_{p-1}$. The intercept is a more relevant parameter when covariates are centered, and correlations between $\hat{\beta}_0$ and the estimates of other regression coefficients are reduced. For examples throughout the book we will sometimes center covariates, but if their means are fairly small, the original uncentered covariates are often retained for simplicity.

The log-likelihood function $\ell(\boldsymbol{\beta}, b)$ is of exactly the same form as (5.1.4), with $z_i = (y_i - u_i)/b$, $u_i = u(\mathbf{x}_i; \boldsymbol{\beta})$, $f_0(z) = -S_0'(z)$ the probability density function (p.d.f.) for $Z_i = (Y_i - u_i)/b$, and $r = \sum \delta_i$ the number of uncensored lifetimes:

$$\ell(\boldsymbol{\beta}, b) = -r \log b + \sum_{i=1}^{n} [\delta_i \log f_0(z_i) + (1 - \delta_i) \log S_0(z_i)]. \tag{6.3.2}$$

Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$ and X be the $n \times p$ matrix with (i, j) entry x_{ij}. Then $\partial z_i/\partial \beta_j = -x_{ij}b^{-1}$, $\partial z_i/\partial b = -z_i b^{-1}$, and the first derivatives of $\ell(\boldsymbol{\beta}, b)$ are simple generalizations of (5.1.5) and (5.1.6):

$$\frac{\partial \ell}{\partial \beta_j} = -\frac{1}{b}\sum_{i=1}^{n}\left[\delta_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1 - \delta_i)\frac{\partial \log S_0(z_i)}{\partial z_i}\right] x_{ij} \tag{6.3.3}$$

$$\frac{\partial \ell}{\partial b} = -\frac{r}{b} - \frac{1}{b}\sum_{i=1}^{n}\left[\delta_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1 - \delta_i)\frac{\partial \log S_0(z_i)}{\partial z_i}\right] z_i. \tag{6.3.4}$$

The second derivatives of $\ell(\boldsymbol{\beta}, b)$ are

$$\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} = \frac{1}{b^2}\sum_{i=1}^{n}\left[\delta_i \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1 - \delta_i)\frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right] x_{ij}x_{ik} \tag{6.3.5}$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial b^2} = {} & \frac{r}{b^2} + \frac{2}{b^2}\sum_{i=1}^{n}\left[\delta_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1 - \delta_i)\frac{\partial \log S_0(z_i)}{\partial z_i}\right] z_i \\ & + \frac{1}{b^2}\sum_{i=1}^{n}\left[\delta_i \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1 - \delta_i)\frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right] z_i^2\end{aligned} \tag{6.3.6}$$

$$\frac{\partial^2 \ell}{\partial \beta_j \partial b} = \frac{1}{b^2}\sum_{i=1}^{n}\left[\delta_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1-\delta_i)\frac{\partial \log S_0(z_i)}{\partial z_i}\right] x_{ij}$$
$$+ \frac{1}{b^2}\sum_{i=1}^{n}\left[\delta_i \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1-\delta_i)\frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right] z_i x_{ij}. \quad (6.3.7)$$

The m.l.e.'s $\hat{\boldsymbol{\beta}}$ and $\hat{b}$ are found by solving the equations $\partial \ell/\partial \boldsymbol{\beta} = \mathbf{0}$, $\partial \ell/\partial b = 0$ or by direct maximization of $\ell(\boldsymbol{\beta}, b)$. Software is available to handle specific models, as discussed in the Computational Notes at the end of the chapter. The observed information matrix is, in partitioned form,

$$I(\boldsymbol{\beta}, b) = \begin{pmatrix} -\partial^2 \ell/\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}' & -\partial^2 \ell/\partial \boldsymbol{\beta}\partial b \\ -\partial^2 \ell/\partial b \partial \boldsymbol{\beta}' & -\partial^2 \ell/\partial b^2 \end{pmatrix}. \quad (6.3.8)$$

The usual large-sample normal approximation to the joint distribution of $\hat{\boldsymbol{\beta}}$ and $\hat{b}$ is to treat them as $(p+1)$-variate normal, with mean vector $(\boldsymbol{\beta}', b)$ and covariance matrix $I(\hat{\boldsymbol{\beta}}, \hat{b})^{-1}$. Since $\hat{\boldsymbol{\beta}}, \hat{b}$ satisfy the equations $\partial \ell/\partial \boldsymbol{\beta} = \mathbf{0}$, $\partial \ell/\partial b = 0$ given by (6.3.3) and (6.3.4), the first sums in each of (6.3.6) and (6.3.7) simplify at $(\hat{\boldsymbol{\beta}}, \hat{b})$.

Tests and interval estimates for parameters can be obtained either by using likelihood ratio tests or the approximate normality of the m.l.e.'s, as described in Appendix C. Two important inference problems will be considered explicitly. The first concerns the regression coefficients, $\boldsymbol{\beta}$. Hypotheses about $\boldsymbol{\beta}$ can frequently be put in the form $H: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0$, with $\boldsymbol{\beta}$ partitioned as $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')$, where $\boldsymbol{\beta}_1$ is $k \times 1$ $(k < p)$ and $\boldsymbol{\beta}_1^0$ is a specified vector. To test H we can use the likelihood ratio statistic

$$\Lambda = 2\ell(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{b}) - 2\ell(\boldsymbol{\beta}_1^0, \tilde{\boldsymbol{\beta}}_2, \tilde{b}), \quad (6.3.9)$$

where $\tilde{\boldsymbol{\beta}}_2$ and $\tilde{b}$ are the m.l.e.'s of $\boldsymbol{\beta}_2$ and b under H, and $\hat{b}$ and $(\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2') = \hat{\boldsymbol{\beta}}'$ are the m.l.e.'s under the full model. Large values of Λ provide evidence against H, and approximate p-values can be calculated by using the fact that for large samples Λ is approximately distributed as $\chi^2_{(k)}$ under H.

An alternative procedure for testing $H: \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^0$ is to use

$$\Lambda_1 = (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0)' V_{11}^{-1} (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^0) \quad (6.3.10)$$

as the test statistic. Here $V = I(\hat{\boldsymbol{\beta}}, \hat{b})^{-1}$ is partitioned as

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{12}' & V_{22} \end{pmatrix},$$

so that V_{11} is the $k \times k$ asymptotic covariance matrix for $\hat{\boldsymbol{\beta}}_1$. For large samples Λ_1 is approximately $\chi^2_{(k)}$ under H. The statistics (6.3.9) and (6.3.10) are asymptotically

equivalent, but for small samples it is usually preferable to use (6.3.9). It requires the m.l.e.'s under both the full model and the model specified by H, but given the speed of optimization software this is not a serious drawback. For individual regression coefficients β_j, estimates and standard errors are often used to test hypotheses $H: \beta_j = 0$ via $Z_j = (\hat{\beta}_j - 0)/se(\hat{\beta}_j)$, treating Z_j as approximately $N(0, 1)$ if H is true. Standard software generates values Z_j for certain models. In the case of small samples, we can compute likelihood ratio statistics as an alternative to the Z_j, if desired.

A second important problem is the estimation of quantiles or survival probabilities. Consider the pth quantile of Y for a given $\mathbf{x}$, which is

$$y_p(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x} + b w_p, \quad (6.3.11)$$

where $w_p = F_0^{-1}(p)$ is the pth quantile of the standard variable Z in (6.1.4). Tests or confidence intervals can be based on the approximate standard normal pivotal quantity

$$Z_p = \frac{\hat{y}_p(\mathbf{x}) - y_p(\mathbf{x})}{se(\hat{y}_p(\mathbf{x}))}, \quad (6.3.12)$$

where $\hat{y}_p(\mathbf{x}) = \hat{\boldsymbol{\beta}}'\mathbf{x} + \hat{b} w_p$ and, by the asymptotic variance formula (B2),

$$se(\hat{y}_p(\mathbf{x})) = \left[(\mathbf{x}', w_p) V (\mathbf{x}', w_p)'\right]^{1/2}, \quad (6.3.13)$$

with $V = I(\hat{\boldsymbol{\beta}}, \hat{b})^{-1}$. Alternatively, confidence intervals can be obtained via the likelihood ratio statistic by considering the hypothesis $H: y_p(\mathbf{x}) = y_{p0}$. The statistic for testing H is

$$\Lambda(y_{p0}) = 2\ell(\hat{\boldsymbol{\beta}}, \hat{b}) - 2\ell(\tilde{\boldsymbol{\beta}}, \tilde{b}), \quad (6.3.14)$$

where $\tilde{\boldsymbol{\beta}}, \tilde{b}$ maximize $\ell(\boldsymbol{\beta}, b)$ under H; they can be found by maximizing $\ell(\boldsymbol{\beta}, (y_{p0} - \boldsymbol{\beta}'\mathbf{x})/w_p)$ to obtain $\tilde{\boldsymbol{\beta}}$, and thus $\tilde{b} = (y_{p0} - \tilde{\boldsymbol{\beta}}'\mathbf{x})/w_p$.

Confidence intervals or tests for the survival probability $S(y_0|\mathbf{x})$, for specified y_0 and $\mathbf{x}$, can be based on an approximate $N(0, 1)$ pivotal quantity analogous to (5.1.14):

$$Z_0 = \frac{\hat{\psi}(\mathbf{x}) - \psi(\mathbf{x})}{se(\hat{\psi}(\mathbf{x}))}, \quad (6.3.15)$$

where $\psi(\mathbf{x}) = S_0^{-1}(S(y_0(\mathbf{x})) = (y_0 - \boldsymbol{\beta}'\mathbf{x})/b$. The asymptotic variance formula (B2) gives

$$se(\hat{\psi}(\mathbf{x})) = \frac{1}{\hat{b}}\left[(\mathbf{x}', \hat{z}_0) V (\mathbf{x}', \hat{z}_0)'\right]^{1/2}, \quad (6.3.16)$$

where $\hat{z}_0 = (y_0 - \hat{\beta}'\mathbf{x})/\hat{b}$. An alternative is to use the likelihood ratio statistic $\Lambda(s_0)$ for testing the hypothesis $H: S(y_0|\mathbf{x}) = s_0$. This leads to a statistic of the form (6.3.14), and involves the same type of constrained maximization of $\ell(\boldsymbol{\beta}, b)$.

Guidelines concernings the accuracy and use of the corresponding procedures for location-scale models with no covariates were given in Section 5.1.1, and apply here also. For small samples, improvements based on signed likelihood ratio square-root statistics are possible, as indicated in Appendix C.

6.3.1.1 *Exact Methods for Uncensored Data*

The exact test and confidence interval procedures introduced for location-scale parameter models in Section 5.1.2 can be extended to the regression model (6.1.4) for the case of uncensored data. Theorem E4 in Appendix E shows that if $\hat{\boldsymbol{\beta}}$, $\hat{b}$ are the m.l.e.'s in a model (6.1.3) or (6.1.4) with a specified distribution $S_0(z)$ for Z, then

$$\mathbf{Z}_1 = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\hat{b}, \qquad Z_2 = \hat{b}/b \tag{6.3.17}$$

are exact pivotal quantities. The distributions of these pivotals or others that can be written in terms of them are usually intractable, but are easily found to any desired degree of accuracy by simulation. To do this for a given sample size n and specified covariate vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$, we merely need to generate independent $Y_1, \ldots, Y_n$ from the models (6.1.3) with $\boldsymbol{\beta} = \mathbf{0}$, $b = 1$ and the given $\mathbf{x}_i$, obtain $\hat{\boldsymbol{\beta}}$ and $\hat{b}$ by fitting the full model (this uses the $\mathbf{x}_i$), and then obtain values for $\mathbf{Z}_1$ and Z_2. By repeating this procedure a large number of times, we can accurately estimate the distribution of $(\mathbf{Z}_1, Z_2)$. This procedure has not been used a great deal, but is feasible and useful, especially in small samples with few covariates.

Theorem E4 also indicates that the quantities $a_i = (y_i - \hat{\boldsymbol{\beta}}'\mathbf{x}_i)/\hat{b}$ are ancillary statistics; these are just the residuals (6.2.3). An alternative approach to inference about parameters is to consider the conditional distributions of $\mathbf{Z}_1$ and Z_2, given $\mathbf{a} = (a_1, \ldots, a_n)$. The form of this distribution is given in Theorem E5, but its use requires numerical integration in most cases. Problem 6.8 considers a simple setting involving an exponential regression model, where this approach is feasible.

When the distribution of Z in (6.1.4) is standard normal, the m.l.e. $\hat{\boldsymbol{\beta}}$ is the least-square estimate $(X'X)^{-1}X'\mathbf{y}$, where $\mathbf{y} = (y_1, \ldots, y_n)'$ and $\mathbf{Z}_1$ and Z_2 have a multivariate t and a transformed chi-squared distribution, respectively. Normal models are considered in Section 6.3.3.

6.3.2 Weibull and Extreme Value Regression Models

In this section we consider Weibull and extreme value regression models in more detail. We will deal with the location-scale form (6.1.2) for the distribution of log-lifetime Y, given $\mathbf{x}$, where $u(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$. Correspondingly, $\alpha(\mathbf{x})$ in the Weibull model (6.1.1) is $\exp(\boldsymbol{\beta}'\mathbf{x})$. The model for Y is therefore of the form

$$Y = \boldsymbol{\beta}'\mathbf{x} + bZ, \tag{6.3.18}$$

where Z has a standard extreme value distribution with p.d.f. and survivor function

$$f_0(z) = \exp(z - e^z), \qquad S_0(z) = \exp(-e^z) \qquad -\infty < z < \infty,$$

respectively. The general expressions in Section 6.3.1 are easily utilized to get the log-likelihood function and its derivatives. Numerous software packages provide m.l.e.'s and related inference procedures based on censored data from this model, as indicated in the Computational Notes at the end of the chapter.

From (6.3.5)–(6.3.8), the observed information matrix evaluated at $\hat{\boldsymbol{\beta}}$, $\hat{b}$ has the partitioned form

$$I(\hat{\boldsymbol{\beta}}, \hat{b}) = \frac{1}{\hat{b}^2}\begin{pmatrix} \sum_{i=1}^{n} e^{\hat{z}_i}\mathbf{x}_i\mathbf{x}_i' & \sum_{i=1}^{n} \hat{z}_i e^{\hat{z}_i}\mathbf{x}_i \\ \sum_{i=1}^{n} \hat{z}_i e^{\hat{z}_i}\mathbf{x}_i' & r + \sum_{i=1}^{n} \hat{z}_i^2 e^{\hat{z}_i} \end{pmatrix}. \tag{6.3.19}$$

Approximate confidence intervals or tests can be obtained by treating $(\hat{\boldsymbol{\beta}}', \hat{b})$ as multivariate normal with covariance matrix $I(\hat{\boldsymbol{\beta}}, \hat{b})^{-1}$ or via likelihood ratio procedures, following the general treatment of Section 6.3.1. For the latter, we use the extreme value form of the log-likelihood (6.3.2), which is

$$\ell(\boldsymbol{\beta}, b) = -r\log b + \sum_{i=1}^{n}(\delta_i z_i - e^{z_i}), \tag{6.3.20}$$

with $z_i = (y_i - \boldsymbol{\beta}'\mathbf{x}_i)/b$.

The following examples illustrate inference procedures based on extreme value and Weibull regression models.

Example 6.3.1. (Leukemia Survival Times). Data on the survival times for 33 leukemia patients were discussed in Example 6.2.1, with a binary blood cell characteristic AG and white blood cell count (wbc) at diagnosis as potential covariates. It was suggested that a Weibull AFT model with covariates $x_1 = AG$ and $x_2 = \log(wbc)$ seemed reasonable. We therefore consider the model (6.3.18) for log survival time Y, with $\mathbf{x} = (1, x_1, x_2)'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$, giving $\boldsymbol{\beta}'\mathbf{x} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$. A summary of results from maximum likelihood estimation is shown in Table 6.2. The Z values in the table show the test statistics $(\hat{\beta}_j - 0)/se(\hat{\beta}_j)$ for testing the hypotheses $H: \beta_j = 0$; for comparison, we also show the signed square-root likelihood ratio statistic, D. For example, the statistic D for testing $H: \beta_1 = 0$ is defined as $D = \text{sign}(\hat{\beta}_1)\Lambda_1(0)^{1/2}$, where

$$\Lambda_1(0) = 2\ell(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{b}) - 2\ell(\tilde{\beta}_0, 0, \tilde{\beta}_2, \tilde{b})$$

Table 6.2. Fitted Weibull–Extreme Value Model (Leukemia Data)

Parameter	Estimate	*se*	*Z*	*D*
β_0 (intercept)	3.841	.534	—	—
$\beta_1(AG)$	1.177	.427	2.74	2.63
$\beta_2(\log wbc)$	−.366	.150	−2.44	−2.46
b (scale)	1.119	.164	—	—

is the likelihood ratio statistic for testing H. Standard software returns the values of the log-likelihood evaluated at the m.l.e., so $\Lambda_1(0)$ is easily obtained by fitting the model with $\beta_1 = 0$, that is, with covariate x_1 dropped.

P-values concerning $H: \beta_j = 0$ can be calculated using the standard normal approximation for either Z or D. The effects of the covariates are clearly significant here, and the statistics Z and D are in close agreement. A positive AG test and low WBC are associated with longer survival. It can also be noted that there is no evidence against the hypothesis $H: b = 1$. This indicates that exponential lifetime distributions, with constant hazard functions for individuals, are plausible.

The full asymptotic covariance matrix $V = I(\hat{\boldsymbol{\beta}}, \hat{b})^{-1}$ is

$$V = \begin{pmatrix} .2856 & -.1302 & -.06710 & .00314 \\ -.1302 & .1824 & .01576 & .00596 \\ -.06710 & .01576 & .02238 & -.00528 \\ .00314 & .00596 & -.00528 & .02689 \end{pmatrix}.$$

This can be used to obtain confidence intervals for quantiles $y_p(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x} + \log(-\log(1-p))b$ or survival probabilities $S(y_0) = \exp[-\exp((y_0 - \boldsymbol{\beta}'\mathbf{x})/b)]$, as described in Section 6.3.1. For example, consider the probability of survival beyond time t_0 for an individual with covariate values $x_1 = AG$, $x_2 = \log(wbc)$, that is

$$S(\log t_0 | x_1, x_2) = \exp[-\exp((y_0 - \beta_0 - \beta_1 x_1 - \beta_2 x_2)/b)],$$

where $y_0 = \log t_0$. This is conveniently handled by considering the parameter

$$\begin{aligned} \psi &= \log[-\log S(\log t_0 | x_1, x_2)] \\ &= (y_0 - \beta_0 - \beta_1 x_1 - \beta_2 x_2)/b \end{aligned}$$

along with the approximate standard normal pivotal quantity (6.3.15). Consider the values $t_0 = 52$ weeks, $AG = 1$, $x_2 = \log(10)$ for illustration. Then we find $\hat{\psi} = -2.00$ and, using (6.3.16), $se(\hat{\psi}) = .271$. The approximate standard normal pivotal $Z = (\hat{\psi} - \psi)/se(\hat{\psi})$ then gives the approximate .95 confidence interval $\hat{\psi} \pm 1.96 se(\hat{\psi})$, or $-.731 \le \psi \le .330$. This transforms to the interval (.249, .618) for the probability of survival beyond 52 weeks, given $AG = 1$ and $wbc^{-1} = 10$. The interval is very wide, as a consequence of the small data set and the amount of variability in survival times for persons with similar covariate values.

Example 6.3.2. (Insulating Fluid Failure Times). Failure time data on electrical insulating fluids subjected to high-voltage stresses were given in Example 1.1.5 and examined in Example 6.2.2, where it was suggested that a Weibull power law model was reasonable. This is an AFT model for which the location-scale model (6.3.18) for log failure time Y has $\boldsymbol{\beta}'\mathbf{x} = \beta_0 + \beta_1 x$, where $x = \log v$ (v is the voltage in kV).

The parameter estimates (standard errors in brackets) are $\hat{\beta}_0 = 64.85(5.62)$, $\hat{\beta}_1 = -17.73(1.61)$, $\hat{b} = 1.288(.113)$, and the asymptotic covariance matrix $V = I(\hat{\boldsymbol{\beta}}, \hat{b})^{-1}$ for $(\hat{\beta}_0, \hat{\beta}_1, \hat{b})$ is

$$V = \begin{pmatrix} 31.5817 & -9.0266 & -.008889 \\ -9.0266 & 2.5819 & .000924 \\ -.008889 & .000924 & .012849 \end{pmatrix}.$$

The main objective in this setting is to use the model to estimate quantiles of the time-to-failure distribution at specified voltage levels. This is considered below, but first we consider some hypothesis tests of the model, based on model expansion.

A key assumption in (6.3.18) is that the scale parameter b does not depend on the voltage, and a second is that the location $u(x)$ is given by $\beta_0 + \beta_1 x$, for $x = \log(v)$. These assumptions seem reasonable from the plots in Example 6.2.2, but can be tested formally. Since there are several observations at each voltage level, we consider three models:

$$M_1: Y \sim EV(u(x), b(x)), \qquad \text{(14 parameters)},$$

where $u(x)$ and $b(x)$ are unrestricted;

$$M_2: Y \sim EV(u(x), b), \qquad \text{(8 parameters)},$$

where $u(x)$ is unrestricted; and

$$M_3: Y \sim EV(\beta_0 + \beta_1 x, b), \qquad \text{(3 parameters)},$$

which is the power law regression model. The assumption that $b(x)$ is constant can be examined by testing the null hypothesis of model M_2 versus model M_1. This was done in Example 5.4.1, where it was found that the maximum log-likelihoods under M_1 and M_2 were $\hat{\ell}(M_1) = -295.26$ and $\hat{\ell}(M_2) = -299.65$. This gave the likelihood ratio statistic value $\Lambda = 2\hat{\ell}(M_1) - 2\hat{\ell}(M_2) = 8.78$ and an approximate p-value from the $\chi^2_{(6)}$ distribution of .187; there is no strong evidence against the assumption of equal scale parameters.

The regression specification $u(x) = \beta_0 + \beta_1 x$ can be examined by testing the null hypothesis of model M_3 versus model M_2. The maximum log-likelihood under M_3 is $\hat{\ell}(M_3) = -300.82$, giving the likelihood ratio statistic $\Lambda = 2\hat{\ell}(M_3) - 2\hat{\ell}(M_2) = 2.34$. The approximate p-value from the $\chi^2_{(5)}$ distribution is .80, and is consistent with model M_3.

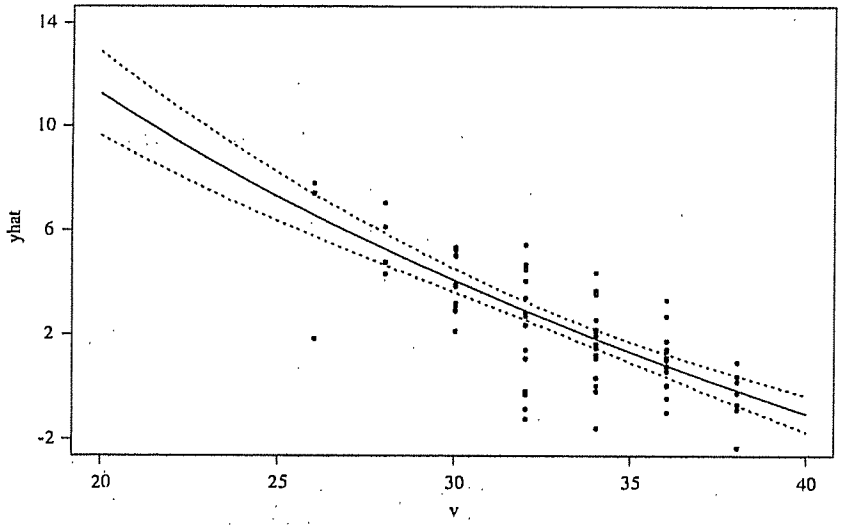

Figure 6.13. Pointwise .95 confidence intervals and m.l.e.'s (yhat, in log minutes) for median log failure time for electrical-insulation at different voltages.

An assessment of the assumption that Z in (6.3.18) has an extreme value distribution is made in Example 6.4.1.

Let us now consider the estimation of quantiles of the log-lifetime and lifetime distributions, using the approximate $N(0, 1)$ pivotal quantity (6.3.12), taking for illustration the median $y_{.50}(x)$. Figure 6.13 shows m.l.e.'s $\hat{y}_{.50}(x)$ and approximate .95 confidence intervals given by $\hat{y}_{.50}(x) \pm 1.96 se(\hat{y}_{.50}(x))$, for different values of x. For descriptive purposes, we show the intervals in terms of the voltage level, v. The log failure times are also shown in the plot. The model-based confidence intervals for $y_{.50}(x)$ are narrow, relative to intervals based on nonparametric methods applied to each voltage level separately (Section 3.2.3).

The standard error for $\hat{y}_{.50}(x)$ is computed using (6.3.13) with $\mathbf{x}' = (1, x)$ and $w_p = w_{.50} = \log(-\log .5) = -.3665$. The voltage level $v = 20$ ($x = 2.9957$) is of particular interest, since 20 kV is a standard operating voltage. For this value we have $\hat{y}_{.50}(\log 20) = 11.262$, with standard error .8225, and .95 confidence interval $9.650 \leq y_{.50}(\log 20) \leq 12.87$. This confidence interval is very wide, as we would expect from extrapolation well beyond the observed data. The interval is also subject to uncertainty regarding the validity of our model outside the range of the data. The interval for $t_{.50}(\log 20)$ is (15,520, 390,077) minutes, or approximately (10.8, 270.9) days.

Example 6.3.3. (Lung Cancer Survival Data). Example 1.1.9 described lung cancer survival data for patients assigned to one of two chemotherapy treatments (Standard and Test). The data, given in Table 1.5, include observations on 40

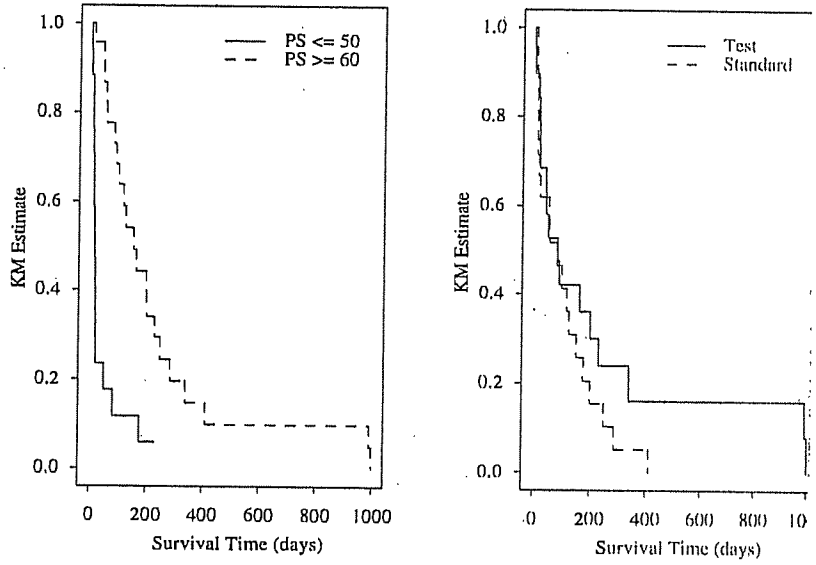

Figure 6.14. Lung cancer survival stratified by PS and by treatment.

patients, and are part of a larger study that is discussed in Example 6.4.3. In addit to treatment (trt), several factors thought to be relevant to an individual's prognc are given: performance status (PS) at diagnosis (a measure of general medical ci dition on a scale of 10 to 90, with lower numbers indicating poorer condition); of the patient at diagnosis (age); the number of months from diagnosis of cance entry into the study (diag); the cell-type of the tumor, classified as being in one four categories, squamous, small, adeno and large. Survival times are measured days from the date of entry to the study.

Preliminary analysis suggests that tumor cell-type and PS may be important. F ure 6.14 shows, for example, Kaplan–Meier plots of the survival times stratifie two ways: (1) PS 10–50 versus PS 60–90, and (2) Standard treatment versus T treatment. Treatment is approximately independent of the other covariates, and the right panel of Figure 6.14 suggests that treatment may not be a significant fac The left-hand panel of the plot suggests that performance status is important, tho association with other factors may explain some of the differences in Kaplan–M estimates. Note in particular that PS and cell-type display association; see Table

As a next step we fit Weibull AFT models, which the exploratory analysis sugg may be suitable. Table 6.4 shows estimates and standard errors for the model (6.3 for log survival time, with

$$u(\mathbf{x}) = \beta_0 + \beta_1 PS + \beta_2 \text{age} + \beta_3 \text{diag} + \beta_4 I(\text{cell} - \text{type} = \text{squamous})$$
$$+ \beta_5 I(\text{cell} - \text{type} = \text{small}) + \beta_6 I(\text{cell} - \text{type} = \text{adeno}) + \beta_7 I(\text{trt} = \text{T}$$

Table 6.3. Numbers of Individuals by PS and Cell-Type

		Cell-Type			
		Squamous	Small	Adeno	Large
PS	10–50	5	6	3	3
	60–90	9	5	2	7

Table 6.4. Fitted Weibull–Extreme Value Models (Lung Cancer Data)

	Parameter	Estimate	*se*	*Z*
Full model	β_0 (int)	.818	1.219	—
	β_1 (PS)	.0542	.0096	5.64
	β_2 (age)	.0094	.0176	.53
	β_3 (diag)	.0041	.0104	.39
	β_4 (squamous)	.377	.400	.95
	β_5 (small)	−.125	.426	−.30
	β_6 (adeno)	−.877	.514	−1.71
	β_7 (trt)	.270	.348	.78
	b (scale)	.874	.115	—
Reduced model	β_0 (int)	1.205	.556	—
	β_1 (PS)	.0604	.0095	6.38
	b (scale)	.981	.123	—

In this model β_4, β_5, and β_6 measure differences between each of the cell-types squamous, small, adeno, and the baseline cell-type large. It seems that only PS is important, but it is prudent to check on interactions; this is readily done by adding defined covariates to the model. For example, a treatment by cell-type interaction can be examined by adding covariates

$$x_8 = I(\text{trt} = \text{Test}) * I(\text{cell} - \text{type} = \text{squamous})$$

$$x_9 = I(\text{trt} = \text{Test}) * I(\text{cell} - \text{type} = \text{small})$$

$$x_{10} = I(\text{trt} = \text{Test}) * I(\text{cell} - \text{type} = \text{adeno}).$$

These checks do not reveal any significant interactions, and so as a next step we fit reduced models where first cell-type differences and then all covariates except PS are dropped from $u(\mathbf{x})$. The likelihood ratio statistics for testing these two submodels against the full model for $u(\mathbf{x})$ above are, respectively,

$$\Lambda_1 = 2\ell(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_7) - 2\ell(\tilde{\beta}_0, \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3, 0, 0, 0, \tilde{\beta}_7)$$

and

$$\Lambda_2 = 2\ell(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_7) - 2\ell(\tilde{\beta}_0, \tilde{\beta}_1, 0, \ldots, 0).$$

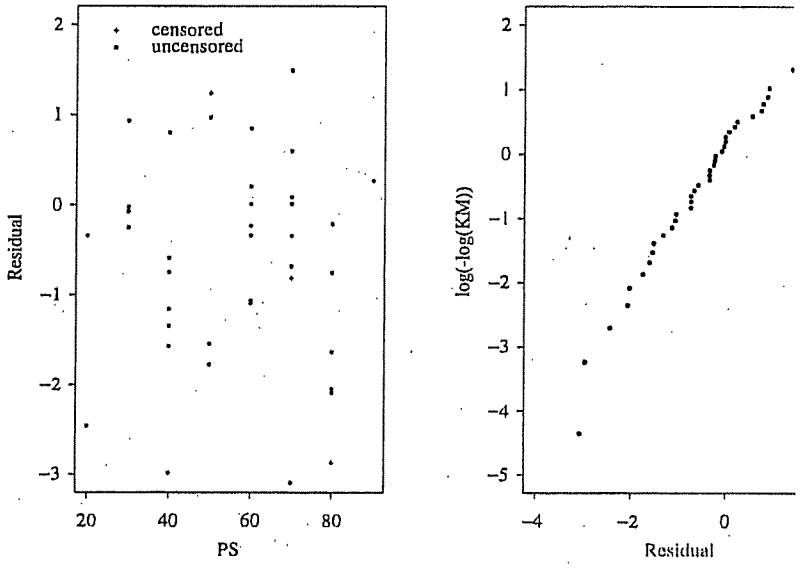

Figure 6.15. Residual plots for Weibull–EV model for lung cancer data.

Observed values are found to be $\Lambda_1 = 4.89$ and $\Lambda_2 = 5.87$. Comparing these quantiles for $\chi^2_{(3)}$ and $\chi^2_{(6)}$, respectively, we find no evidence against the model only PS included.

Table 6.4 also shows the reduced model with only PS included. Diagnostic ch on both the full model and the reduced model, as described in Section 6.2, do show any major problems. Figure 6.15 shows plots for the reduced model of extr value residuals $\hat{z}_i$ (6.2.3) versus PS_i and an extreme value probability plot base the $\hat{z}_i$, constructed as for Figure 6.9 in Example 6.2.1. Note that our model treat as a quantitative covariate, though it is actually an ordinal variate. Entering it linear term in $u(\mathbf{x})$ provides a reasonable fit to the data in this setting.

We conclude that PS is the only factor strongly related to survival time. Its e is roughly linear on the log time scale, with a PS increase of 10 giving an estim increase in median log survival time of approximately .6 when time is in days.

6.3.3 Normal-Log-Normal and Logistic-Log-Logistic Regression Models

The log-normal and log-logistic AFT models have location-scale forms (6.1.2), Z having a standard normal distribution and a standard logistic distribution, re tively. As for the extreme value distribution in the preceding section, we con models where $u(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$, so that

$$Y = \boldsymbol{\beta}'\mathbf{x} + bZ,$$

with (i) $Z \sim N(0, 1)$ or (ii) $Z \sim$ standard logistic, Logist(0,1). The p.d.f. and survivor functions for Z in the two cases are

$$\text{(i).}\quad f_0(z) = \frac{1}{\sqrt{2\pi}}\exp(-z^2/2), \qquad S_0(z) = \int_{-\infty}^{z} f_0(u)\,du, \qquad -\infty < z < \infty$$

$$\text{(ii).}\quad f_0(z) = \frac{e^z}{(1+e^z)^2}, \qquad S_0(z) = \frac{1}{(1+e^z)}, \qquad -\infty < z < \infty.$$

Many software packages provide m.l.e.'s and related inference procedures based on censored data from these distributions. Standard output includes the asymptotic covariance matrix $I(\hat{\beta}, \hat{b})^{-1}$ given by (6.3.8), and the value of the maximized log-likelihood function $\ell(\hat{\beta}, \hat{b})$. This makes it easy to test hypotheses concerning nested models and to get approximate confidence intervals for parameters, quantiles, or survival probabilities by using the methods described in Section 6.3.1. Methods based on normal approximations for m.l.e.'s are sufficiently accurate for most practical purposes if there are about 30 or more uncensored lifetimes. Likelihood ratio methods for obtaining confidence intervals can be used in circumstances where there are concerns about the accuracy of the normal approximations.

We note that the conditional expectations needed for the calculation of adjusted residuals (6.2.4) can be obtained in closed form for both the normal and logistic distributions. For standard normal Z it can be shown that

$$E(Z|Z \geq z) = \phi(z)/(1 - \Phi(z)) \tag{6.3.21}$$

where $\phi(z)$ and $\Phi(z)$ are the $N(0, 1)$ p.d.f. and c.d.f.. For Z a standard logistic random variable,

$$E(Z|Z \geq z) = \frac{1}{1+e^z}\left\{\frac{z}{1+e^z} - \log\left(\frac{e^z}{1+e^z}\right)\right\}. \tag{6.3.22}$$

In the case of uncensored data there are, of course, exact methods for the normal model that are described in books on regression analysis, and thus there is no need for the approximate likelihood methods. In particular, $\hat{\beta} = (X'X)^{-1}X'\mathbf{y}$, where X is the $n \times p$ matrix with (i, j) entry x_{ij} and $\mathbf{y} = (y_1, \ldots, y_n)'$, and $\hat{b}^2 = (\mathbf{y} - X\hat{\beta})'(\mathbf{y} - X\hat{\beta})/n$ give the m.l.e.'s. Confidence intervals and tests based on Student's-t and χ^2 pivotal quantities follow from the fact that $\hat{\beta} \sim N_p(\beta, b^2(X'X)^{-1})$, $n\hat{b}^2/b^2 \sim \chi^2_{(n-p)}$, and $\hat{\beta}$ and $\hat{b}$ are independent. Exact procedures for the logistic model can also be obtained via the pivotals (6.3.17) and simulation, as discussed in Section 6.3.1.

Let us also note how confidence intervals for quantiles can be calculated for the normal model when there is no censoring. This is based on a straightforward extension of the procedure described in Section 5.3.1, in which the pivotal quantity (5.3.3) is merely replaced with

$$Z_p = \frac{\tilde{y}_p(\mathbf{x}) - y_p(\mathbf{x})}{s}, \tag{6.3.23}$$

where $s^2 = n\hat{b}^2/(n-p)$ is the standard unbiased estimate of b^2. Since $(n-p)s^2/b^2 \sim \chi^2_{(n-p)}$, $\hat{\beta}'\mathbf{x} \sim N[\beta'\mathbf{x}, b^2A(\mathbf{x})^2]$, where $A(\mathbf{x})^2 = \mathbf{x}'(X'X)^{-1}\mathbf{x}$, and $\hat{\beta}$ and s^2 are independent, it follows that

$$Pr(Z_p \leq z) = Pr\left[\frac{(\hat{\beta}'\mathbf{x} - \beta'\mathbf{x})/bA(\mathbf{x}) - w_p/A(\mathbf{x})}{s/b} \leq \frac{z - w_p}{A(\mathbf{x})}\right]$$

$$= Pr\left[t'_{(n-p)}\left(\frac{-w_p}{A(\mathbf{x})}\right) \leq \frac{z - w_p}{A(\mathbf{x})}\right], \tag{6.3.24}$$

where $t'_{(v)}(\lambda)$ represents a noncentral t random variable, defined in Section 5.3.1. Confidence intervals for $y_p(\mathbf{x})$ can be found from probability statements for Z_p, exactly as described in Section 5.3.1 for the case of no covariates, with $A(\mathbf{x})^{-1}$ replacing $\sqrt{n}$ and $n - p$ replacing $n - 1$ in (5.3.4).

Example 6.3.4. (Times to Pulmonary Exacerbation). Example 6.2.3 examined clinical trial data on times to a pulmonary exacerbation for persons with cystic fibrosis, introduced in Example 1.1.8. There are two covariates, treatment and fev, and it was suggested by graphical examination that a log-normal accelerated failure time model would provide a good description of the data. This was verified in the continuation of Example 6.2.3 by diagnostic checks on a fitted log-normal model. We describe here the analysis of the data under a log-normal model, and under a log-logistic model, which is similar.

In both cases the model considered for $Y = \log T$ has location-scale form (6.1.4) with $u(\mathbf{x}_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$, where $x_{i1} = I(\text{treatment} = \text{rhDNase})$ and $x_{i2} = \text{fevc}_i = \text{fev}_i - \overline{\text{fev}}$, the mean fev across all subjects in the study. Before deciding on this model, the possibility of using different functions of fev, such as log(fev), was examined, and the linear specification was found satisfactory. The presence of a treatment–fev interaction was also checked, by fitting models in which β_2 depends on treatment, but proved insignificant.

Table 6.5 shows m.l.e.'s and standard errors for the parameters in each of the normal and logistic models. Note that the regression coefficients have the same interpretations in the two models, but that the scale parameters b do not. As would be expected, the estimates and standard errors under the two models are in close agreement, and indicate significant effects for both treatment and fev. Diagnostic checks carried out in Example 6.2.3 supported the log-normal model, and similar checks on

Table 6.5. Log-Normal and Log-Logistic Fits for Pulmonary Exacerbation Times

Parameter	Log-Normal		Log-Logistic	
	Estimate	*se*	Estimate	*se*
β_0 (int)	5.403	.105	5.353	.096
β_1 (trt)	.430	.137	.401	.130
β_2 (fevc)	.0217	.0029	.0207	.0028
b (scale)	1.446	.074	.796	.045

the log-logistic model support it as well. Because censoring is heavy and no failure times exceed 169 days, there is no information about the upper tail of the distribution of T or Y given $\mathbf{x}$, and the data do not discriminate between the models. The maximum log-likelihood values $\ell(\hat{\boldsymbol{\beta}}, \hat{b})$ under the two models for Y are -1625.04 (normal) and -1626.32 (logistic), indicating slightly more support for the normal, but no significant difference.

6.3.4 Some Comments on Least Squares, Robustness, and Efficiency

When the data are uncensored, the estimates of $\boldsymbol{\beta}$ and b in a location-scale regression model (6.1.4) with Z normally distributed are least-square estimates, and possess certain robustness properties. Linear location-scale models and estimates based on them also possess a weaker type of robustness. We will discuss these issues briefly.

Let us change our notation slightly from previous sections and write models in the form

$$Y_i = \beta_0 + \boldsymbol{\beta}'\mathbf{x}_i + bZ_i \qquad i = 1, \ldots, n, \tag{6.3.25}$$

where $b > 0$, $\mathbf{x}_i$ and $\boldsymbol{\beta}$ are $(p-1) \times 1$ vectors, and $Z_1, \ldots, Z_n$ are i.i.d. with some specified distribution on $(-\infty, \infty)$, which we assume has mean 0 and finite variance. Note that writing the model so $E(Z_i) = 0$ merely affects the intercept parameter. For example, the extreme value model (6.3.18) with $Z \sim EV(0, 1)$ has $E(Z) = -\gamma$, where $\gamma = .5772$ is Euler's constant. Thus $Z_i' = Z_i + \gamma$ has mean 0, and rewriting (6.3.18) with Z' replacing Z simply changes the intercept from β_0 to $\beta_0' = \beta_0 - b\gamma$. Assume also without loss of generality that the covariates are centered so that

$$\sum_{i=1}^{n} x_{ij} = 0 \qquad j = 1, \ldots, p-1. \tag{6.3.26}$$

The m.l.e.'s of $\boldsymbol{\beta}$ and β_0 when $Z_i \sim N(0, 1)$ are also least-square estimates:

$$\tilde{\beta}_0 = \bar{y} \qquad \tilde{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}, \tag{6.3.27}$$

where the $n \times (p-1)$ matrix $X = (x_{ij})$ is assumed to be of full rank, and $\mathbf{y} = (y_1, \ldots, y_n)'$. It is well known that $\tilde{\beta}_0$ and $\tilde{\boldsymbol{\beta}}$ are unbiased estimators of β_0 and $\boldsymbol{\beta}$, and the covariance matrix of $\tilde{\boldsymbol{\beta}}^* = (\tilde{\beta}_0, \tilde{\boldsymbol{\beta}}')'$ is

$$\operatorname{Var}(\tilde{\boldsymbol{\beta}}^*) = b^2 \operatorname{Var}(Z) \begin{pmatrix} n^{-1} & 0 \\ 0 & (X'X)^{-1} \end{pmatrix}. \tag{6.3.28}$$

These results hold true no matter what the distribution of Z, though the efficiency of the least-square estimates may not be high in some situations; this depends on the distribution of Z. The efficiency of least-square estimation is considered in Problem 6.13, where the points below are developed in more detail.

Suppose that the model (6.3.25) holds, but Z_i has some other distribution than $N(0, 1)$, still with $E(Z_i) = 0$. The Fisher information matrix for the parameters

b, β_0, and $\boldsymbol{\beta}$ when there is no censoring is easily seen by taking expectations of (6.3.5)–(6.3.7) to be of the form

$$\mathcal{I}(b, \beta_0, \boldsymbol{\beta}) = \begin{bmatrix} \mathcal{I}_1 & 0 \\ 0 & b^{-2}A_Z(X'X) \end{bmatrix}, \tag{6.3.29}$$

where $\mathcal{I}_1$ is 2×2, $X'X$ is $(p-1) \times (p-1)$, and

$$A_Z = E\left\{\frac{-\partial^2 \log f_0(z)}{\partial z^2}\right\}. \tag{6.3.30}$$

This shows that the m.l.e.'s $\hat{\boldsymbol{\beta}}$ and $(\hat{b}, \hat{\beta}_0)$ are asymptotically independent, and that the asymptotic covariance matrix for $\hat{\boldsymbol{\beta}}$ is

$$\operatorname{Asvar}(\hat{\boldsymbol{\beta}}) = \frac{b^2}{A_Z}(X'X)^{-1}. \tag{6.3.31}$$

This may be compared with the variance of the least-square estimator $\tilde{\boldsymbol{\beta}}$, which from (6.3.28) is $b^2 \operatorname{Var}(Z)(X'X)^{-1}$. The asymptotic efficiency of the least-square estimator relative to the m.l.e. under the true distribution $f_0(z)$ for Z is therefore given by

$$\frac{\operatorname{Asvar}(\hat{\beta}_j)}{\operatorname{Var}(\tilde{\beta}_j)} = \frac{1}{A_Z \operatorname{Var}(Z)} \qquad j = 1, \ldots, p-1. \tag{6.3.32}$$

Problem 6.13 outlines calculations showing that when Z has a logistic distribution, the asymptotic relative efficiency of least-squares is .91, and that when Z has an extreme value distribution, it is .61.

If a specific distribution for Z is assumed in the model (6.3.25), then we typically use m.l.e.'s for that distribution. Maximum likelihood under a normal distribution for Z, or least-squares, possesses the desirable robustness property that inferences about $\boldsymbol{\beta}$ (or β_0) are still valid when the distribution of Z is nonnormal. A weaker robustness property also holds for m.l.e.'s under other distributions for Z. First, note that when location-scale models are expressed in the form (6.3.25) with $E(Z_i) = 0$, there are interpretations of β_0 and $\boldsymbol{\beta}$ that are independent of the actual distribution of Z. In particular, $E(Y_i|\mathbf{x}_i) = \beta_0 + \boldsymbol{\beta}'\mathbf{x}_i$ so, for example, β_j represents the increase in the mean of Y when the jth covariate x_j increases by one unit. It similarly represents the increase in any quantile $y_p(\mathbf{x})$ under the same condition. Note that the interpretation of b does depend on the distribution of Z through the fact that $\operatorname{Var}(Y_i|\mathbf{x}_i) = b^2 \operatorname{Var}(Z)$, since we have not standardized Z to have variance 1. Silvapulle (1985) and Gould and Lawless (1988) show that the m.l.e. $\hat{\boldsymbol{\beta}}$ obtained using a specific distribution for Z in (6.3.25) is still consistent for the true regression parameter, provided that the true distribution is of the form (6.3.25). In other words, misspecification of the distribution of Z does not lead to inconsistency in the esti-

mation of regression coefficients, provided the location-scale assumption is correct. On the negative side, however, is the fact that estimates of β_0, b and the mean and quantiles of the distribution of Y given $\mathbf{x}$ are affected by misspecification, as is the estimation of standard errors for all parameters.

Because of the robustness properties of least squares for uncensored samples, attempts to extend least-square methods to censored data have been made. Other robust estimation methods for location-scale models can also be developed; this is discussed in Section 8.2. The robustness properties of least-squares and, more generally, m.l. estimation for (6.3.25) do not extend to censored data, though they hold approximately when censoring is light.

All models are of course only approximations to reality, and it is worth repeating the recommended approach for dealing with concerns about robustness or model misspecification: first, carry out diagnostic checks that can provide support for basing inferences and conclusions on any specific family of models and, second, perform sensitivity analyses in which the effects of varying the model can be examined. The latter activity is aided by model expansion, and is discussed in Section 6.4 for location-scale regression models. Earlier examples not involving covariates were provided in Examples 5.5.1 and 5.5.2.

6.3.5 Experimental Design

In studies where values of explanatory factors can be controlled, experimental design can be used to increase efficiency and reduce cost. The considerations for general linear location-scale models (6.3.25) are very much the same as for the classic normal linear model. When there is no censoring we can indeed use results directly from classic theory, in view of the similarity between the information matrix (6.3.29) for general models and that for normal (least-square) estimates in (6.3.28). Guidelines for multifactor experiments as in Example 6.1.2, or for the choice of covariate values in problems such as Example 6.2.2, can then be found in standard experimental design references. Problem 6.15 provides an illustration.

The presence of time constraints that lead to censoring complicates matters. One problem is that when covariates affect lifetimes substantially the degree of censoring in fixed-time studies may vary considerably across different levels of the covariates. Another is that information about parameters has to be considered in the context of a specific distribution, since we no longer have the direct link with least-squares. Moreover, trade-offs in the choice of covariate values and the duration of experimentation can be made, in addition to the usual trade-offs between duration and sample size. A good approach is to explore designs or study plans by simulating data sets or by numerical evaluation of information matrices, as described for the case of no covariates in Section 5.6.1.

The information matrix $I(\boldsymbol{\beta}, b)$ is given by (6.3.8), and a comparison of expressions (6.3.5)–(6.3.7) with expressions (5.1.7)–(5.1.9) shows the close similarity to the information matrix in the no-covariate case. In fact, the Fisher information matrix based on (6.3.8) can be obtained directly from the analogous expressions for $\mathcal{I}_{0,11}$, $\mathcal{I}_{0,12}$, and $\mathcal{I}_{0,22}$ in (5.6.2)–(5.6.4) for the no-covariate case. Taking expected

values of (6.3.5)–(6.3.7) and noting that the first terms in square brackets in each of (6.3.6) and (6.3.7) simplify (since $E(\partial\ell/\partial\beta_j) = 0$ and $E(\partial\ell/\partial b) = 0$), we find

$$\mathcal{I}(\boldsymbol{\beta}, b) = E[I(\boldsymbol{\beta}, b)] \tag{6.3.33}$$

is a $(p+1) \times (p+1)$ matrix with entries

$$\mathcal{I}(\boldsymbol{\beta}, b)_{jk} = \frac{1}{b^2}\sum_{i=1}^{n} \mathcal{I}_{0,11i} x_{ij} x_{ik} \qquad j, k = 1, \ldots, p \tag{6.3.34}$$

$$\mathcal{I}(\boldsymbol{\beta}, b)_{j,p+1} = \mathcal{I}(\boldsymbol{\beta}, b)_{p+1,j} = \frac{1}{b^2}\sum_{i=1}^{n} \mathcal{I}_{0,12i} x_{ij} \qquad j = 1, \ldots, p \tag{6.3.35}$$

$$\mathcal{I}(\boldsymbol{\beta}, b)_{p+1,p+1} = \frac{1}{b^2}\sum_{i=1}^{n} \mathcal{I}_{0,22i}. \tag{6.3.36}$$

In these expressions we have denoted terms $I_{0,k\ell i}$ $(k, \ell = 1, 2)$, since they depend on $(\log C_i - \boldsymbol{\beta}'\mathbf{x}_i)/b$, and thus they vary for $i = 1, \ldots, n$. Tables, charts, and general evaluations of the information matrix terms (5.6.2)–(5.6.4) discussed in Section 5.6.1 can also be applied to models with covariates.

We provide a simple example, following which some results that have been obtained for certain specific models are discussed.

Example 6.3.5. Consider the extreme value regression model (6.3.18). Since $\partial^2 \log f_0(z)/\partial z^2 = -\exp(z) = \partial^2 \log S_0(z)/\partial z^2$, expressions (5.6.2)–(5.6.4) and (6.3.34)–(6.3.36) give, in partitioned form,

$$\mathcal{I}(\boldsymbol{\beta}, b) = \frac{1}{b^2}\begin{pmatrix} \sum_{i=1}^{n} E(e^{Z_i})\mathbf{x}_i\mathbf{x}_i' & \sum_{i=1}^{n} E(Z_i e^{Z_i})\mathbf{x}_i \\ \sum_{i=1}^{n} E(Z_i e^{Z_i})\mathbf{x}_i' & \sum_{i=1}^{n} E(\delta_i + Z_i^2 e^{Z_i}) \end{pmatrix}. \tag{6.3.37}$$

The expectations $E(e^{Z_i})$, $E(Z_i e^{Z_i})$, and $E(\delta_i + Z_i^2 e^{Z_i})$ are considered in Problem 5.4. In particular, note that under a Type 1 censoring scheme where individual i has potential censoring time C_i,

$$\begin{aligned} Z_i &= [\min(Y_i, \log C_i) - \boldsymbol{\beta}'\mathbf{x}_i]/b \\ &= \min(Z_i^*, R_i), \end{aligned}$$

where $Z_i^* = (Y_i - \boldsymbol{\beta}'\mathbf{x}_i)/b$ is $EV(0, 1)$ and $R_i = (\log C_i - \boldsymbol{\beta}'\mathbf{x}_i)/b$. In addition, $E(\delta_i) = Pr(\delta_i = 1) = 1 - \exp(-e^{R_i})$.

In many applications, such as comparative experiments, precise estimation of regression coefficients is of interest. Estimation of survival probabilities or of dis-

tribution quantiles $y_p(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x} + w_p b$ may also be important. In either case it is necessary to specify working values for $\boldsymbol{\beta}$ and b for planning purposes. Then, a specific plan consisting of values for n, censoring times $C_1, \ldots, C_n$, and covariate values $\mathbf{x}_1, \ldots, \mathbf{x}_n$ can be assessed through an examination of asymptotic variances for parameter estimates of interest, obtained from the asymptotic covariance matrix $V = \mathcal{I}(\boldsymbol{\beta}, b)^{-1}$. We can obtain V by direct numerical calculation of (6.3.37) in the case of Type 1 censoring or, more generally, we can estimate it via simulation. Simulation also allows an examination of the observed information matrix $I(\hat{\boldsymbol{\beta}}, \hat{b})$ and how sampling variation affects variance estimates from $I(\hat{\boldsymbol{\beta}}, \hat{b})^{-1}$, and provides empirical estimates of the distributions of estimates or test statistics.

Let us consider models with a single covariate, so that $\boldsymbol{\beta}'\mathbf{x}_i = \beta_0 + \beta_1 x_i$. The observed information matrix (6.3.19) in this case, reduces to

$$I(\hat{\beta}_0, \hat{\beta}_1, \hat{b}) = \frac{1}{\hat{b}^2} \begin{pmatrix} \sum_{i=1}^n \hat{z}_i & \sum_{i=1}^n \hat{z}_i x_i & \sum_{i=1}^n \hat{z}_i e^{\hat{z}_i} \\ \sum_{i=1}^n \hat{z}_i x_i & \sum_{i=1}^n \hat{z}_i x_i^2 & \sum_{i=1}^n \hat{z}_i e^{\hat{z}_i} x_i \\ \sum_{i=1}^n \hat{z}_i e^{\hat{z}_i} & \sum_{i=1}^n \hat{z}_i e^{\hat{z}_i} x_i & \sum_{i=1}^n (\delta_i + \hat{z}_i^2 e^{\hat{z}_i}) \end{pmatrix}, \qquad (6.3.38)$$

and the Fisher information matrix $\mathcal{I}(\beta_0, \beta_1, b)$ is of the same form, with $\hat{z}_i$ replaced by $E(Z_i)$, $\hat{z}_i \exp(\hat{z}_i)$ replaced by $E(Z_i \exp(Z_i))$, and so on. Even if the censoring times C_i are equal, Z_i depends on x_i and so direct calculation of I or $\mathcal{I}$ depends critically on these values. Algorithms for evaluation of terms in $\mathcal{I}(\beta_0, \beta_1, b)$ have been published by Escobar and Meeker (1994), as discussed in Section 5.6.1. The observed information matrix $I(\hat{\beta}_0, \hat{\beta}_1, \hat{b})$ is easily calculated for any simulated sample once the m.l.e.'s have been found.

The evaluation of experimental plans that involve censoring is rather laborious, either by simulation or by numerical evaluation of Fisher information matrices. A good practical approach in many settings is first to consider plans with no censoring (all $C_i = \infty$). In this case, the Fisher information matrix has the form (6.3.29), and the covariance matrix for $\hat{\boldsymbol{\beta}}$ is proportional to that for the least-square estimator, which is $b^2 \text{var}(Z)(X'X)^{-1}$. Traditional design considerations can then be used to suggest sample size and choice of covariate values $\mathbf{x}_1, \ldots, \mathbf{x}_n$. The effect of censoring is to reduce information and increase asymptotic variances of parameter estimates. The amount of variance inflation is variable across parameters, and is difficult to judge because a given censoring time produces different censoring probabilities for different covariate values. However, an examination of variation inflation for the no-covariate setting of Section 5.6.1 often provides reasonable guidance and, at least, a starting point for simulation or further numerical evaluation.

The special problem of accelerated life testing has received considerable attention. Here the objective is to estimate quantiles $y_p(\mathbf{x})$ for settings involving one or two

covariates that represent factors such as voltage, temperature, or humidity. Meeker and Escobar (1998, Ch. 20) and Nelson (1990) provide detailed treatments of this area.

6.4 EXTENSIONS OF LOG-LOCATION-SCALE MODELS

6.4.1 Families of Error Distributions

Section 5.5 discussed extensions to location-scale parameter models in which a parametric family of standard distributions for $Z = (Y - u)/b$ was considered. The same extensions can be applied to settings with covariates; we consider models

$$Y_i = \boldsymbol{\beta}'\mathbf{x}_i + bZ_i \qquad i = 1, \ldots, n \qquad (6.4.1)$$

for log-lifetime Y_i given covariate vector $\mathbf{x}_i$, where Z_i has a distribution with survivor function $S_0(z; k)$, as in (5.5.1). Fitting such models allows a check on the sensitivity of inferences to variations in the error distribution in (6.4.1), provides increased modeling flexibility, and allows an assessment of common error distributions such as the extreme value and normal. Estimation is best carried out as described in Section 5.5.1, by maximizing the log-likelihood function $\ell(\boldsymbol{\beta}, b, k)$ for $\boldsymbol{\beta}$ and b with k held fixed at different values. This gives estimates $\tilde{\boldsymbol{\beta}}(k)$ and $\tilde{b}(k)$ and the profile log-likelihood function

$$\ell_p(k) = \ell(\tilde{\boldsymbol{\beta}}(k), \tilde{b}(k), k).$$

Maximization of $\ell_p(k)$ gives the m.l.e. $\hat{k}$ and the m.l.e.'s $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{k})$, $\hat{b} = \tilde{b}(\hat{k})$ for $\boldsymbol{\beta}$ and b. Plausible values for k can be assessed using $\ell_p(k)$ or the equivalent likelihood ratio statistic

$$\Lambda(k) = 2\ell(\hat{\boldsymbol{\beta}}, \hat{b}, \hat{k}) - 2\ell(\tilde{\boldsymbol{\beta}}(k), \tilde{b}(k), k).$$

For interior values k_0 in the parameter space, the distribution of $\Lambda(k_0)$ is typically $\chi^2_{(1)}$ in large samples if $k = k_0$; this can be used to test $H: k = k_0$ or to obtain confidence intervals for k.

As discussed in Section 5.5.1, it is often reasonable to provide inferences about the distribution of Y with k treated as fixed, and to assess informally the effect of varying k on such inferences. When there is no censoring the estimate $\hat{k}$ is asymptotically independent of the m.l.e. for the regression coefficients. In particular, if we write the model (6.4.1) in the alternative form

$$Y_i = \beta_0 + \boldsymbol{\beta}'\mathbf{x}_i + bZ_i, \qquad i = 1, \ldots, n \qquad (6.4.2)$$

and center covariates as in (6.3.26), then a simple extension of the development in Section 6.3.4 shows that $(\hat{\beta}_0, \hat{b}, \hat{k})$ are asymptotically independent of $\hat{\boldsymbol{\beta}}$. This proves that $\hat{k}$ and $\hat{b}$ are asymptotically independent of $\hat{\boldsymbol{\beta}}$ regardless of whether the covariates are centered or not, because centering affects only the intercept and not any of the other parameters or their estimates. When lifetimes are censored the asymptotic

independence no longer holds, but the correlation between $\hat{k}$ and $\hat{\boldsymbol{\beta}}$ is typically small. This means that inferences about $\boldsymbol{\beta}$ will usually change little as k is varied across a range of plausible values. The estimates $\hat{\beta}_0$, $\hat{b}$ are often rather highly correlated with $\hat{k}$, but estimates of quantiles or survival probabilities are much less so, especially away from the tails of the distribution. These are usually of much more interest than β_0 or b on their own. These points are illustrated in the examples that are presented later.

Two useful extended families are the generalized log-Burr and generalized log-gamma, discussed in Sections 5.5.2 and 5.5.3 for the case of no covariates. For the log-Burr regression family the p.d.f. $f_0(z_i; k)$ of Z_i in (6.4.1) is given by (5.5.6), and the log-likelihood from a censored random sample of lifetimes by (5.5.7):

$$\ell(\boldsymbol{\beta}, b, k) = -r \log b + \sum_{i=1}^{n} \delta_i \log f_0(z_i; k) + \sum_{i=1}^{n} (1 - \delta_i) \log S_0(z_i; k), \quad (6.4.3)$$

where $z_i = (y_i - \boldsymbol{\beta}'\mathbf{x}_i)/b$. The extreme value regression model of Section 6.3.2 is given by the limit $k \to \infty$ and the logistic regression model by $k = 1$, and standard software can be used for those cases. For other fixed values of k the general methods of Section 6.3.1 apply, and are readily implemented with general optimization software. A plot of the profile log-likelihood $\ell_p(k)$ or likelihood ratio statistic $\Lambda(k) = 2\ell_p(\hat{k}) - 2\ell_p(k)$ can be used to assess plausible values of k. The behavior of $\ell_p(k)$ is the same as discussed for the no-covariate case in Section 5.5.2.

The log-gamma regression model has likelihood functions (6.4.3), but with $f_0(z; k)$ and $S_0(z; k)$ given by (5.5.10) and (5.5.8), respectively. The special case $k = 1$ gives the extreme value regression model, and the limit as $k \to \infty$ gives the normal linear model. Standard software handles these cases, and general optimization software will deal with other values of k. As discussed in Section 5.5.3, it is better to use the parameter $\lambda = k^{-1/2}$ for estimation purposes, and the special case, $\lambda = 0$, can be made into an interior point in the parameter space by a simple extension of the model to allow $\lambda < 0$ as well as $\lambda \geq 0$. In this case, the p.d.f. of Z_i takes the form

$$f_0(z; \lambda) = \begin{cases} \dfrac{|\lambda|}{\Gamma(\lambda^{-2})} (\lambda^{-2})^{\lambda^{-2}} \exp[\lambda^{-2}(\lambda z - e^{\lambda z}).] & \lambda \neq 0 \\ \dfrac{1}{(2\pi)^{1/2}} \exp(-z^2/2) & \lambda = 0 \end{cases} \quad (6.4.4)$$

for $-\infty < z < \infty$.

The examples that follow illustrate the application of the generalized log-Burr and log-gamma families. It is possible to consider a log F model with two shape parameters that includes these as special cases (Kalbfleisch and Prentice 1980, Sec. 3.9; Gould and Lawless 1988), but experience suggests that the log-Burr and log-gamma models provide sufficient flexibility for most practical purposes.

Example 6.4.1. Let us reconsider the data on failure times of electrical insulating fluid subjected to high-voltage stresses, introduced in Example 1.1.5. A Weibull accelerated failure time model, or extreme value location-scale model with $u(\mathbf{x})$ in (6.1.4) given by $\beta_0 + \beta_1 x$ and $x = \log(\text{voltage})$, was fitted in Example 6.3.2. Checks on the model there and in Example 6.2.2 did not reveal any significant evidence against it. Nevertheless, let us push the data a bit further and examine a generalized log-gamma regression model given by (6.4.1), with the errors having the density function $f_0(z; \lambda)$ of (6.4.4). The extreme value–Weibull model is represented by the special case $\lambda = 1$, so this provides a check on that model, and allows a sensitivity analysis of conclusions based on it.

Values of the profile log-likelihood function $\ell_p(k) = \ell(\tilde{\boldsymbol{\beta}}(k), \tilde{b}(k), k)$, where $k = \lambda^{-2}$, are easily obtained by maximizing (6.4.3) with respect to $\boldsymbol{\beta}$ and b. The m.l.e. $\hat{k}$ is found by maximizing $\ell_p(k)$, or $\ell(\boldsymbol{\beta}, b, k)$, and the associated likelihood ratio statistic takes values $\Lambda(k) = 2\ell(\hat{\boldsymbol{\beta}}, \hat{b}, \hat{k}) - 2\ell_p(k)$. We find that $\hat{k} = 1.818$, $\hat{\beta}_0 = 64.328$, $\hat{\beta}_1 = -17.631$, $\hat{b} = 1.348$, and that the extreme value model ($k = 1$) has $\Lambda(k) = .72$; assessing this relative to a $\chi^2_{(1)}$ distribution indicates that there is no evidence whatsoever against the model. The normal location-scale model (6.4.1), on the other hand, corresponds to $\lambda = 0$ ($k = \infty$), and we find that $\Lambda(\infty) = 6.306$. This gives a p-value of .012 with the $\chi^2_{(1)}$ approximation and provides evidence against the model.

Confidence intervals for quantiles $y_p(x)$ were given under the extreme value model in Example 6.3.2. Inferences about central quantiles, $y_p(x)$, are insensitive to varying k away from the extreme value model ($k = 1$), provided we consider k values that are well supported by the data. Quantiles for p close to 0 or 1 and for x-values beyond the experimental data are more sensitive to variations in k, as illustrated for some data with no covariates in Example 5.5.2. We reiterate the remark made there that confidence limits based on a specific parametric model are often optimistically narrow, given uncertainty about the model.

Crowder et al. (1991, p. 82) and Farewell and Prentice (1977) noted that separate log-gamma models fitted for units with $v \leq 32$ kV and those with $v > 32$ kV gave rather different estimates of k. Fitting a model with separate parameters β_0, β_1, b, and k for the two voltage categories gives $\hat{k} = .55$ for the lower voltages and $\hat{k} = 9.81$ for the higher ones. However, a likelihood ratio test of this eight-parameter model versus the four-parameter model discussed earlier gives the likelihood ratio statistic $\Lambda = 5.0$, which yields a p-value from $\chi^2_{(4)}$ of .29, and thus no evidence of a difference in parameters for the two voltage categories. A five-parameter model in which only k is different for low and high voltages gives a likelihood ratio statistic $\Lambda = 3.57$, which gives a p-value from $\chi^2_{(1)}$ of .06. This provides mild but inconclusive evidence regarding a possibly different shape for the failure time distribution at low and high voltages. Separate extreme value probability plots of the residuals from the four-parameter extreme value model ($k = 1$) fitted in Example 6.3.2 for units with $v \leq 32$ and $v > 32$, respectively, similarly do not show pronounced evidence against the extreme value model, nor does a plot of residuals against voltage levels.

Example 6.4.2. Crowder (2000) discussed some data of Watson and Smith (1985) on the breaking strengths T of single carbon fibers of different lengths; the

data are given in Appendix G. Generalized log-Burr models have often been found to fit breaking strength data rather well, and we consider them here.

For the log-Burr distribution with location and scale parameters $u(\mathbf{x})$ and b and shape parameter k, the survivor function for log-lifetime Y given $\mathbf{x}$ is from (5.5.5),

$$S(y|\mathbf{x}) = \left[1 + \frac{1}{k}\exp\left(\frac{y-u(\mathbf{x})}{b}\right)\right]^{-k} \qquad -\infty < y < \infty. \tag{6.4.5}$$

In the current example "lifetime" is breaking strength T, $Y = \log T$, and the only covariate is fiber length. Probability plots described in Section 6.2.1 can help to see whether models of the form (6.4.5) are plausible for given fiber lengths. The data involve approximately 60 fiber specimens for each of four lengths, $\ell = 1, 10, 20,$ and 50. Let $\hat{S}_\ell(t)$ represent the Kaplan–Meier estimate of the survivor function for T for fiber length ℓ ($\ell = 1, 10, 20, 50$). For a model of the form (6.4.5) with a specified value of k, we have

$$\log\left[k\left(S(y|\mathbf{x})^{-1/k} - 1\right)\right] = \frac{y-u(\mathbf{x})}{b}, \tag{6.4.6}$$

so if the model is plausible, then plots of $\log[k(\hat{S}_\ell(t)^{-1/k} - 1)]$ versus $\log t$ should be roughly linear and parallel.

Figure 6.16 shows such a plot using the value $k = 2$. This value was found to give quite linear plots, as shown. The logistic model given by $k = 1$ also gave reasonably linear plots, but the extreme value model ($k = \infty$), for which the left side of (6.4.6)

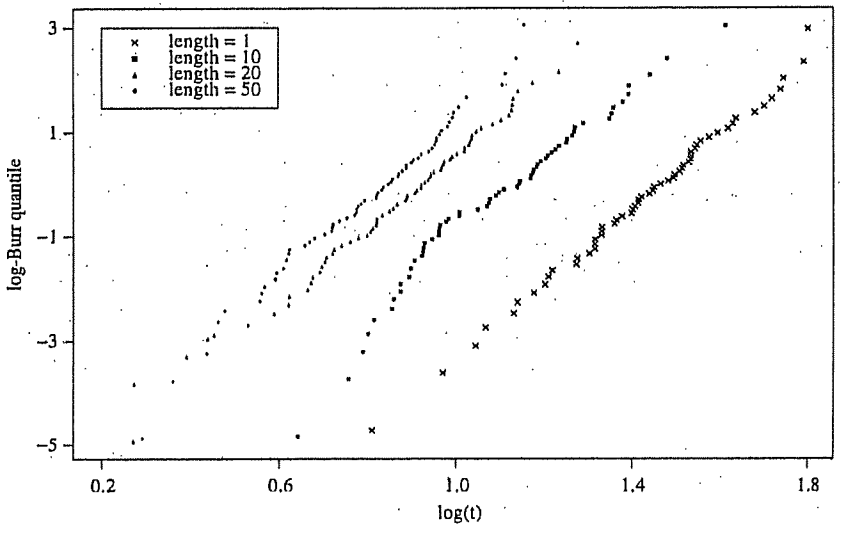

Figure 6.16. Log-Burr probability plots ($k = 2$) for fiber breaking strength data.

becomes the limit $\log(-\log S(y|\mathbf{x}))$ does not and is considerably less plausible. A slight anomaly for the breaking strengths with $\ell = 10$ is noted, with the left tail of the distribution being a little shorter than the log-Burr model suggests.

The probability plots for the four lengths $\ell = 1, 10, 20, 50$ are reasonably close to parallel, which suggests that a common value for the scale parameter b should suffice, as in (6.4.5). Breaking strengths tend to be lower when the fiber length is longer, and it is of interest to see whether a simple relationship exists.

There is considerable theory concerning length effects for the breaking strength of strands or fibers. One simple model relies on weakest link arguments and leads to log-Burr models for Y (e.g., Crowder et al. 1991, Sec. 7.3; Watson and Smith 1985) where the survivor function for fibers of length ℓ is of the form

$$S(y|\ell) = \left[1 + \frac{\ell}{k}\exp\left(\frac{y-u}{b}\right)\right]^{-k}. \tag{6.4.7}$$

This can be written in the form (6.4.5) with $x = \log \ell$ and

$$u(x) = \beta_0 + \beta_1 x, \tag{6.4.8}$$

where $\beta_1 = -b$. As a check on this model we fitted three models to the data:

M_0: Model (6.4.5) with (6.4.8) and β_1 not constrained to equal b; this has four-parameters (β_0, β_1, b, k).

M_1: Model (6.4.5) with $u(x) = u_\ell$ for $\ell = 1, 10, 20, 50$; this has six parameters $(u_1, u_{10}, u_{20}, u_{50}, b, k)$.

M_2: Log-Burr models with common k but distinct location and scale parameters; this has nine parameters $(u_1, u_{10}, u_{20}, u_{50}, b_1, b_{10}, b_{20}, b_{50}, k)$.

Parameter estimates and maximized log-likelihood values $\log_{\hat{L}}$ under the three models are as follows:

M_0: $\hat{\beta}_0 = 1.47, \hat{\beta}_1 = -.167, \hat{b} = .139, \hat{k} = 2.046, \log_{\hat{L}} = -219.59$

M_1: $\hat{u}_1 = 1.47, \hat{u}_{10} = 1.14, \hat{u}_{20} = .926, \hat{u}_{50} = .839, \hat{b} = .140, \hat{k} = 2.275, \log_{\hat{L}} = -215.29$

M_2: $\hat{u}_1 = 1.47, \hat{u}_{10} = 1.138, \hat{u}_{20} = .927, \hat{u}_{50} = .843, \hat{b}_1 = .141, \hat{b}_{10} = .147, \hat{b}_{20} = .143, \hat{b}_{50} = .133, \hat{k} = 2.365, \log_{\hat{L}} = -215.03.$

The likelihood ratio statistic for testing M_1 versus M_2 is $\Lambda = .52$ on three degrees of freedom, with an associated χ^2 p-value of .915. The analogous test for M_0 versus M_1 gives $\Lambda = 8.62$ on two degrees of freedom, and a χ^2 p-value of .013. A third test of M_2 against a 12-parameter model in which the k's were allowed to vary with length gave a likelihood ratio statistic of 2.24 on four degrees of freedom. Thus, there is considerable support for a log-Burr model (6.4.5) for log breaking strength, in which $u(x)$ varies with fiber length, ℓ. However, the model (6.4.8) with $x =$

log ℓ is contradicted to some degree by the data, though not exceptionally strongly. It can also be noted that in model M_0 the estimates $\hat{\beta}_1$ and $-\hat{b}$ are similar, as the weakest link model suggests, though the observed difference is significant. These results agree broadly with empirical work that suggests that the weakest link model is approximately correct in some circumstances, but that with sufficient data, departures from it are frequently discernible. In this instance a plot of values $\hat{u}_\ell$ from model M_2 versus log ℓ suggests mild departures from linearity, but data at more than four values for ℓ are needed to provide a clearer picture.

Example 6.4.3. Examples 1.1.9 and 6.3.3 considered data on the survival of 40 lung cancer patients. These were part of a larger study involving 137 patients, which we discuss here. The covariates include those in Example 6.3.3, labeled as before: performance status (PS), age in years at diagnosis (age), number of months from diagnosis to entry into the study (diag), tumor cell-type (squamous, small, adeno) with large as baseline, and I(treatment = Standard) (trt). The 40 patients in Example 1.1.9 and Example 6.3.3 are those who received prior therapy; the remaining 97 patients did not. This is represented here by the additional covariate I(prior therapy), denoted below by "therapy." Survival times are measured in days from the date of entry to the study. The full data are available in electronic form (see Appendix G).

Preliminary analysis as in Example 6.3.3 suggests that the covariates age and months since diagnosis are not important, and this is confirmed by fitting separate log-Burr models (6.4.5) to the patients who received and did not receive prior therapy. These variables were then dropped and models with only treatment, performance status, cell-type, and prior therapy were examined. Models fitted with interaction terms suggested that only interactions involving prior therapy needed consideration. This leads to Table 6.6, which shows fits of three log-Burr models (6.4.5): a model with main effects for PS, trt, cell type, and therapy for all 137 patients, plus separate models with main effects for PS, trt, and cell type for patients with and without therapy. A likelihood ratio test of the model with only a main effect for prior therapy

Table 6.6. Log-Burr Models Fitted to Lung Cancer Data

	Full Data (n = 137)		Therapy (n = 40)		No Therapy (n = 97)	
Parameter	Estimate	se	Estimate	se	Estimate	se
β (int)	2.74	.41	1.49	.63	3.29	.46
β (PS)	.034	.005	.053	.0095	.027	.005
β (squamous)	.133	.277	.366	.387	.193	.316
β (small)	−.649	.248	−.050	.408	−.739	.275
β (adeno)	−.760	.265	−.868	.511	−.671	.293
β (trt)	−.090	.178	.255	.306	−.276	.201
β (therapy)	−.082	.197	—	—	—	—
b	.665	.092	.886	.116	.625	.096
k	1.63	.79	4.179	—	1.72	.94
	$\hat{\ell} = -192.34$		$\hat{\ell} = -56.57$		$\hat{\ell} = -129.95$	

versus the model stratified on prior therapy gives an observed value $\Lambda = 11.64$ and a p-value from $\chi^2_{(7)}$ of .1. There is no conclusive evidence of a prior therapy interaction effect, though it is interesting to note that the error distributions suggested for the two prior therapy groups is different: essentially extreme value ($k = \infty$) for the prior therapy group and closer to logistic for the no prior therapy group. For the group receiving prior therapy the log-likelihood is flat in the region of $\hat{k} = 4179$, and for larger values of k, so no standard error is given in Table 6.6.

The picture regarding covariate effects is quite clear. Performance status has a strong effect in both therapy groups (slightly more so for those with prior therapy), treatment has no significant effect in either case, and there is a suggestion of a difference for the small and adeno cell-types versus the squamous and large types, the evidence being primarily due to the no prior therapy group.

The results given here agree with those of Farewell and Prentice (1977), who fitted log-gamma models to these data. Diagnostics do not cast doubt on either the log-Burr or log-gamma models.

6.4.2 Variable Scale Parameters

Standard location-scale models (6.1.3) and associated AFT models (6.1.5) have a scale parameter b that does not depend on the covariates. This assumption is sometimes unsuitable, and we may seek to retain the location-scale form but specify some form of dependency of b on $\mathbf{x}$. Since b has to be nonnegative, a common specification is $b(\mathbf{x}) = \exp(\gamma'\mathbf{x})$, where γ is a vector of parameters. In some cases we may wish to use different covariates in $b(\mathbf{x})$ than in the location model $u(\mathbf{x})$. For example, it is often adequate to let b depend on the levels of a small number of discrete factors. The term heteroscedasticity is sometimes used in regression analysis to describe settings where $\mathrm{Var}(Y|\mathbf{x})$ is not constant.

We can carry out graphical diagnostic checks on the constancy of b, as for ordinary linear regression models with uncensored data. For example, plots of log failure time y against single covariates as in Section 6.2.1 may directly suggest nonconstancy; the probability plots in Section 6.2.2, which are based on grouping individuals according to covariate values, may suggest nonconstancy through nonparallel plots; plots of residuals (6.2.3) versus covariates or fitted values, $\hat{u}_i$, may show nonuniform dispersion patterns.

Formal tests of the constancy of b can be carried out through model expansion. The electrical insulating fluid failure time data in Example 6.3.2 was considered through a set of location-scale models that allowed a check on whether $b(x) = b$. It should be noted that strong evidence of heteroscedasticity is often absent even when plots suggest it; a good deal of data are needed to reach clear conclusions. For the insulating fluid data, for example, the probability plots in Figure 6.4 suggest that an extreme value location-scale model may be reasonable, but the variation in slopes (or the nonparallelism of the plots) raises doubts about the constancy of b. However, the formal hypothesis test in Example 6.3.2 gave a p-value of about .19, thus providing no significant evidence of heteroscedasticity. Detecting heteroscedasticity in data sets with several covariates is even more difficult.

Regression models with survivor functions of the form

$$S(y|\mathbf{x}) = S_0\left[\frac{y - u(\mathbf{x};\boldsymbol{\beta})}{b(\mathbf{x};\boldsymbol{\gamma})}\right] \tag{6.4.9}$$

can be fitted by maximum likelihood using general optimization software. The log-likelihood function from a censored random sample is a direct generalization of (6.3.2),

$$\ell(\boldsymbol{\beta},\boldsymbol{\gamma}) = -\sum_{i=1}^{n}\delta_i \log b_i + \sum_{i=1}^{n}[\delta_i \log f_0(z_i) + (1-\delta_i)\log S_0(z_i)], \tag{6.4.10}$$

where $z_i = (y_i - u_i)/b_i$, $u_i = u(\mathbf{x}_i;\boldsymbol{\beta})$ and $b_i = b(\mathbf{x}_i;\boldsymbol{\gamma})$. For the model with specifications $u_i = \boldsymbol{\beta}'\mathbf{x}_i$ and $b_i = \exp(\boldsymbol{\gamma}'\mathbf{x}_i)$, the derivatives of $\ell(\boldsymbol{\beta},\boldsymbol{\gamma})$ take simple forms: first derivatives are, for $j = 1,\ldots,p$,

$$\frac{\partial \ell}{\partial \beta_j} = -\sum_{i=1}^{n}\left[\delta_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1-\delta_i)\frac{\partial \log S_0(z_i)}{\partial z_i}\right]\frac{x_{ij}}{b_i} \tag{6.4.11}$$

$$\frac{\partial \ell}{\partial \gamma_j} = -\sum_{i=1}^{n}\delta_i x_{ij} - \sum_{i=1}^{n}\left[\delta_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1-\delta_i)\frac{\partial \log S_0(z_i)}{\partial z_i}\right] z_i x_{ij}. \tag{6.4.12}$$

Second derivatives are, for j and $k = 1,\ldots,p$,

$$\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^{n}\left[\delta_i \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1-\delta_i)\frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right]\frac{x_{ij}x_{ik}}{b_i^2} \tag{6.4.13}$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \gamma_j \partial \gamma_k} = &\sum_{i=1}^{n}\left[\delta_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1-\delta_i)\frac{\partial \log S_0(z_i)}{\partial z_i}\right] z_i x_{ij}x_{ik} \\ &+ \sum_{i=1}^{n}\left[\delta_i \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1-\delta_i)\frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right] z_i^2 x_{ij}x_{ik}\end{aligned} \tag{6.4.14}$$

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \beta_j \partial \gamma_k} = &\sum_{i=1}^{n}\left[\delta_i \frac{\partial \log f_0(z_i)}{\partial z_i} + (1-\delta_i)\frac{\partial \log S_0(z_i)}{\partial z_i}\right]\frac{x_{ij}x_{ik}}{b_i} \\ &+ \sum_{i=1}^{n}\left[\delta_i \frac{\partial^2 \log f_0(z_i)}{\partial z_i^2} + (1-\delta_i)\frac{\partial^2 \log S_0(z_i)}{\partial z_i^2}\right]\frac{z_i x_{ij}x_{ik}}{b_i}.\end{aligned} \tag{6.4.15}$$

Example 6.4.4. Kimber (1990) and Crowder (2000) discuss data on the times to failure of steel specimens subjected to cyclic stress loading of various amplitudes. The data given by Crowder are reproduced in Appendix G; they are for 20 specimens at each of the 14 stress amplitudes 32.0, 32.5, 33.0, ..., 38.0, 38.5. Failure times t are in numbers of thousands of stress cycles. None of the 280 times are censored.

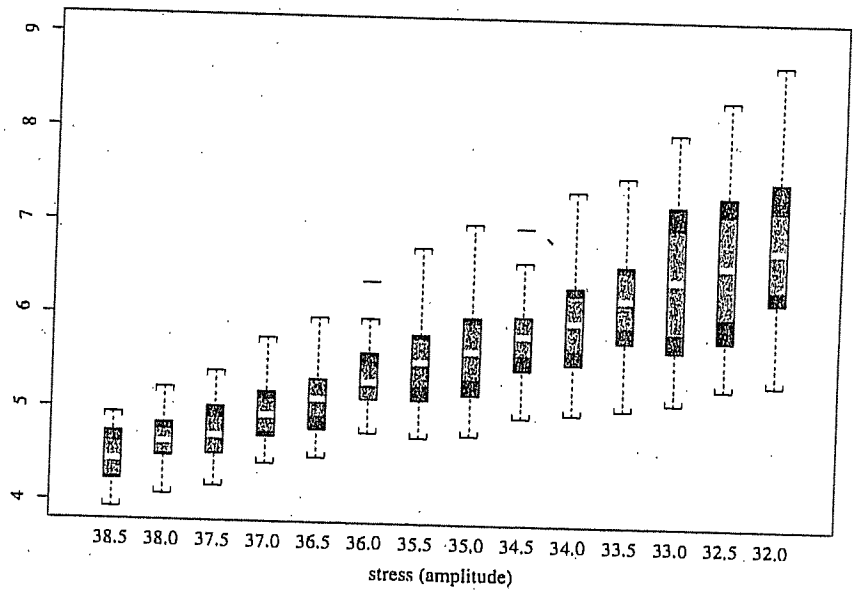

Figure 6.17. Box plots of steel-specimen log failure times at 14 stress levels.

Figure 6.17 shows box plots of the log failure times at each stress amplitude. This suggests that log failure times tend to be smaller at higher amplitudes, and also that the dispersion decreases as the amplitude increases. Further preliminary examination of the data suggests that a log-Burr distribution may provide a satisfactory description. Fitting separate log-Burr models as well as models in which k is the same across amplitude levels indicates no need for different k's, so we next consider models of the form (6.4.9) with $S_0(z)$ of the form (5.5.5).

The specifications

$$u(x) = \beta_0 + \beta_1 x, \qquad \log b(x) = \gamma_0 + \gamma_1 x, \tag{6.4.16}$$

with $x = \log$ (amplitude), are suggested by plots of the data and by numerical summaries of location and spread such as $y_{.5}(x)$ and $sd(x)$. Figure 6.18 shows plots of $\log \hat{u}_j$ and $\log \hat{b}_j$ versus x_j for the log-Burr model with constant k, but different parameters (u_j, b_j), $j = 1,\ldots,14$. The models with 29 parameters (different u and b but the same k at each value of x), 5 parameters ((6.4.16) and constant k), and 4 parameters ((6.4.16) with $\gamma_1 = 0$ and constant k) give maximized log-likelihoods $\hat{\ell} = -211.12$, -214.26, and -243.43, respectively. This provides strong support for the model (6.4.16) with both location and scale effects. The m.l.e.'s for this model are $\hat{\beta}_0 = 47.92$, $\hat{\beta}_1 = -11.92$, $\hat{\gamma}_0 = 22.13$, $\hat{\gamma}_1 = -6.61$, and $\hat{k} = .643$.

Finally, we remark that a log-Burr probability plot of residuals from the fitted five-parameter log-Burr model shows a strong linear pattern, but with a few residuals at

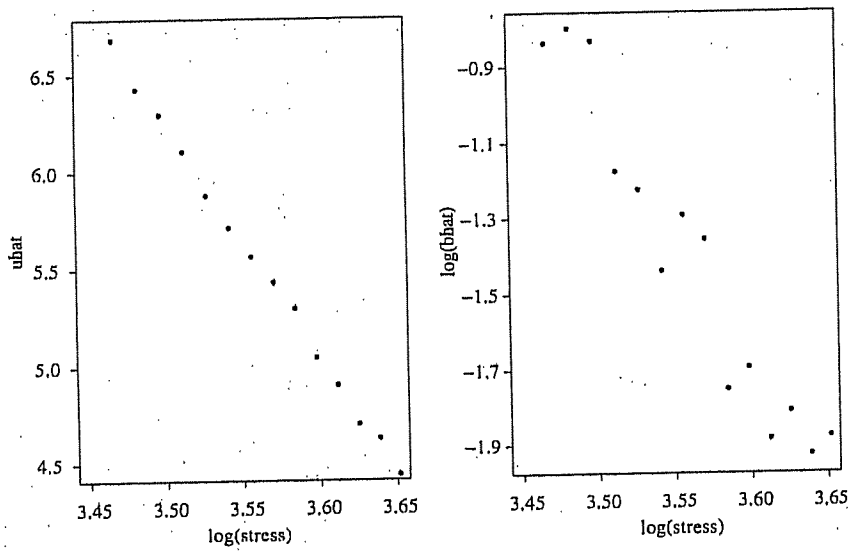

Figure 6.18. Location and scale estimates at 14 stress levels for steel-specimen failures.

either end departing from this. This is not unusual with large data sets, but one could refine the model, for example, by allowing some mild nonlinearity in $u(x)$ or by considering other distributions $S_0(z)$. Crowder (2000) considers a larger family that includes the log-Burr, but his results do not reject the log-Burr model used here.

6.4.3 Time-Varying Covariates

Time-varying covariates can be incorporated into accelerated failure time or time-transformation models. The key idea is that the covariate $\mathbf{x}(t)$ effectively alters the rate at which time t passes. This leads to models in which the survivor function for T, conditional on an external covariate path $X = \{x(t), t \geq 0\}$, is of the form

$$S(t|X) = S_0\left[\int_0^t g(\mathbf{x}(u))\,du\right], \tag{6.4.17}$$

where $g(\cdot)$ is a positive-valued function and $S_0(\cdot)$ is a survivor function on $(0, \infty)$. If $\mathbf{x}(u) = \mathbf{x}$ is constant over time, then (6.4.17) reduces to $S_0(g(\mathbf{x})t)$, which is the standard accelerated failure time model.

A common procedure is to specify $g(\mathbf{x}(u))$ parametrically as $\exp(\boldsymbol{\gamma}'\mathbf{x}(u))$, in which case (6.4.17) becomes

$$S(t|X) = S_0\left[\int_0^t e^{\boldsymbol{\gamma}'\mathbf{x}(u)}\,du\right]. \tag{6.4.18}$$

Fully parametric models assume further that $S_0(\cdot)$ has a specific form; a Weibull model, for example, would take $S_0(z) = \exp\left[-(z/\alpha)^\delta\right]$. Fully parametric models have been used in engineering applications such as variable stress accelerated life tests (e.g., Nelson 1990, Ch. 10). It is also possible to develop semiparametric procedures in which $S_0(\cdot)$ in (6.4.18) is left arbitrary. This topic is discussed briefly in Section 8.2.

Example 6.4.5. Step-Stress Accelerated Life Tests. In some accelerated life tests the covariate $x(t)$ represents a stress that is varied over time, so that

$$x(t) = x_\ell \qquad a_{\ell-1} \leq t < a_\ell$$

for $\ell = 1, \ldots, k$, where $a_0 = 0$ and $a_k = \infty$. Suppose that the model (6.4.18) applies, with $\boldsymbol{\gamma}'\mathbf{x}(u) = \gamma x(u)$. Then, given a stress history $X = \{x(t), t > 0\}$, the survivor function for failure time T is

$$\begin{aligned} S(t|X) &= S_0\left[\int_0^t e^{\gamma x(u)}\,du\right] \\ &= S_0\left[\sum_{\ell=1}^{k} e^{\gamma x_\ell}\Delta_\ell(t)\right], \end{aligned} \tag{6.4.19}$$

where, as defined in (1.3.26), $\Delta_\ell(t) = 0$, $t - a_{\ell-1}$, or $a_\ell - a_{\ell-1}$ according to whether $t < a_{\ell-1}$, $a_{\ell-1} \leq t < a_\ell$, or $t > a_\ell$, respectively. With a Weibull model for $S_0(z)$, (6.4.19) becomes

$$S(t|X) = \exp\left[-\frac{1}{\alpha}\sum_{\ell=1}^{k} e^{\gamma x_\ell}\Delta_\ell(t)\right]^\delta. \tag{6.4.20}$$

The variable $z(t) = \int_0^t g(\mathbf{x}(u))\,du$ in (6.4.17) can be thought of as a generalized time scale on which units have survivor function $S_0(z)$. Such time scales, which can depend on time-varying covariates, appear in various theories about the physics of failure. This concept is considered more generally by Duchesne and Lawless (2000).

6.5 SOME OTHER MODELS

Location-scale or AFT models provide considerable flexibility in fitting data, are easily interpreted, and in many settings are plausible with respect to knowledge of the underlying failure process. However, other formulations are often necessary. This section provides a brief discussion of some other models. First we consider specifications based on hazard functions, then models derived from mechanistic assumptions about failure, and we finish with some comments on models that are linked to certain transformations.

6.5.1 Hazard-Based Models

Section 6.1 indicated that the two main approaches to regression modeling of lifetimes are time-transformation and hazard-based modeling. In the latter, models are specified in terms of the way that covariates $\mathbf{x}$ affect the hazard function for T. Flexible but easily interpreted models are sought, and the most common is the PH or multiplicative model (6.1.6). Additive models with $h(t|\mathbf{x})$ of the form (6.1.9) are also useful. This section outlines some parametric hazard-based specifications for lifetime regression. Semiparametric multiplicative models are considered at much greater length in Chapter 7.

If there is a sufficient number of failures and not too many different covariate values in the data, then nonparametric estimates of the hazard function for different groups of individuals can be examined for insight into potential models. Estimation of hazard functions was discussed in Section 3.4. Insight can also be gained through the examination of Nelson–Aalen plots of cumulative hazard function estimates $\hat{H}(t|\mathbf{x})$, as described in Section 6.2.1. More generally, we rely on a combination of data exploration, model fitting, and model assessment. For the latter, the cumulative hazard residuals (6.2.7) and (6.2.8) are useful.

Software for fitting parametric models is not widely available, except for special cases such as the Weibull PH model, which is also in the accelerated failure time family. However, estimation and inference is generally easy to implement using optimization software. If a parametric model $h(t|\mathbf{x};\boldsymbol{\theta})$ is considered, then the log-likelihood function based on a censored random sample $(t_i;\delta_i)$ $i = 1,\ldots,n$ of lifetimes, given covariate values $\mathbf{x}_1,\ldots,\mathbf{x}_n$, can be written in the form (2.2.17), giving

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \{\delta_i \log h(t_i|\mathbf{x}_i;\boldsymbol{\theta}) - H(t_i|\mathbf{x}_i;\boldsymbol{\theta})\}, \tag{6.5.1}$$

where $H(t|\mathbf{x}) = \int_0^t h(u|\mathbf{x})\,du$ is the cumulative hazard function. This can be maximized directly, with or without the use of analytic derivatives for $\ell(\boldsymbol{\theta})$.

Multiplicative models are convenient and flexible. Consider models (6.1.6) with fixed covariates $\mathbf{x}$, in which

$$h(t|\mathbf{x};\boldsymbol{\theta}) = h_0(t;\boldsymbol{\alpha})\exp(\boldsymbol{\beta}'\mathbf{x}), \tag{6.5.2}$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha},\boldsymbol{\beta})$ and $h_0(t;\boldsymbol{\alpha})$ is a positive-valued baseline hazard function. The specification $r(\mathbf{x};\boldsymbol{\beta}) = \exp(\boldsymbol{\beta}'\mathbf{x})$ in (6.1.6) is convenient, but other specifications are also easily handled. The log-likelihood function (6.5.1) becomes

$$\ell(\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left\{\log h_0(t_i;\boldsymbol{\alpha}) + \boldsymbol{\beta}'\mathbf{x}_i\right\} - \sum_{i=1}^{n} H_0(t_i;\boldsymbol{\alpha})\exp(\boldsymbol{\beta}'\mathbf{x}_i), \tag{6.5.3}$$

where $H_0(t;\boldsymbol{\alpha}) = \int_0^t h_0(u)\,du$ is the baseline cumulative hazard function. First and second derivatives of $\ell(\boldsymbol{\alpha},\boldsymbol{\beta})$ are of simple form; the latter are, in particular,

$$\frac{\partial^2 \ell}{\partial\alpha_j\partial\alpha_k} = \sum_{i=1}^{n}\left\{\delta_i \frac{\partial^2 \log h_0(t_i;\boldsymbol{\alpha})}{\partial\alpha_j\partial\alpha_k} - \frac{\partial^2 H_0(t_i;\boldsymbol{\alpha})}{\partial\alpha_j\partial\alpha_k}\exp(\boldsymbol{\beta}'\mathbf{x}_i)\right\} \tag{6.5.4}$$

$$\frac{\partial^2 \ell}{\partial\alpha_j\partial\beta_k} = -\sum_{i=1}^{n} \frac{\partial H_0(t_i;\boldsymbol{\alpha})}{\partial\alpha_j} x_{ik}\exp(\boldsymbol{\beta}'\mathbf{x}_i) \tag{6.5.5}$$

$$\frac{\partial^2 \ell}{\partial\beta_j\partial\beta_k} = -\sum_{i=1}^{n} x_{ij}x_{ik}H_0(t_i;\boldsymbol{\alpha})\exp(\boldsymbol{\beta}'\mathbf{x}_i), \tag{6.5.6}$$

from which the information matrix $I(\boldsymbol{\alpha},\boldsymbol{\beta})$ and estimated asymptotic covariance matrix $I(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})^{-1}$ for $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})$ can be obtained.

Models for which $h_0(t)$ or $\log h_0(t)$ are linear in parameters $\alpha_1,\ldots,\alpha_r$ are often useful; this includes polynomial or regression spline models, as discussed in Section 4.2.3. A very simple model to handle is one with piecewise-constant $h_0(t)$; although the hazard function is not smooth, this can be quite flexible and useful. The following example outlines maximum likelihood estimation.

Example 6.5.1. Models with Piecewise-Constant Hazard Functions. Consider (6.5.2) with $h_0(t;\boldsymbol{\alpha})$ piecewise-constant: for a specified set of points $a_1,\ldots,a_{k-1}$ with $a_0 = 0$, $a_k = \infty$, we have

$$h_0(t;\boldsymbol{\alpha}) = \alpha_j \qquad a_{j-1} \le t < a_j \qquad (j = 1,\ldots,k). \tag{6.5.7}$$

The baseline cumulative hazard function is

$$H_0(t) = \sum_{j=1}^{k} \alpha_j \Delta_j(t), \tag{6.5.8}$$

where $\Delta_j(t) = 0$, $t - a_{j-1}$, or $a_j - a_{j-1}$ if $t < a_{j-1}$, $a_{j-1} \le t < a_j$, or $t \ge a_j$, respectively, as defined in (1.3.26). The log-likelihood (6.5.3) can then be written as

$$\begin{aligned}\ell(\boldsymbol{\alpha},\boldsymbol{\beta}) &= \sum_{i=1}^{n}\delta_i\left[\sum_{j=1}^{k} I(a_{j-1} \le t_i < a_j)\log\alpha_j + \boldsymbol{\beta}'\mathbf{x}_i\right] \\ &\quad - \sum_{i=1}^{n}\left[\sum_{j=1}^{k}\Delta_j(t_i)\alpha_j\right]e^{\boldsymbol{\beta}'\mathbf{x}_i} \\ &= \sum_{j=1}^{k} d_j\log\alpha_j + \sum_{i=1}^{n}\delta_i\boldsymbol{\beta}'\mathbf{x}_i - \sum_{j=1}^{k}\alpha_j\left[\sum_{i=1}^{n}\Delta_j(t_i)e^{\boldsymbol{\beta}'\mathbf{x}_i}\right],\end{aligned} \tag{6.5.9}$$

where $d_j = \sum_i \delta_i I(a_{j-1} \le t_i < a_j)$ is the number of lifetimes that lie in $[a_{j-1}, a_j)$. Maximization of (6.5.9) can be simplified by noting that for a given $\boldsymbol{\beta}$, the equation

$\partial \ell / \partial \boldsymbol{\alpha} = \mathbf{0}$ has the closed-form solution

$$\tilde{\alpha}_j(\boldsymbol{\beta}) = \frac{d_j}{\sum_{i=1}^{n} \Delta_j(t_i) e^{\boldsymbol{\beta}' \mathbf{x}_i}} \qquad j = 1, \ldots, k. \tag{6.5.10}$$

Insertion of these values into $\ell(\boldsymbol{\alpha}, \boldsymbol{\beta})$ gives the profile log-likelihood for $\boldsymbol{\beta}$,

$$\ell_p(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \boldsymbol{\beta}' \mathbf{x}_i - \sum_{j=1}^{k} d_j \log \left[\sum_{\ell=1}^{n} \Delta_j(t_\ell) e^{\boldsymbol{\beta}' \mathbf{x}_\ell} \right] + c, \tag{6.5.11}$$

where $c = \sum (d_j \log d_j - d_j)$ does not depend on $\boldsymbol{\beta}$, and can be omitted.

An easy way to obtain the m.l.e.'s is therefore to maximize (6.5.11) to get $\hat{\boldsymbol{\beta}}$ and then to determine $\hat{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}})$ from (6.5.10). The asymptotic covariance matrix $I(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})^{-1}$ is easily obtained from the information matrix $I(\boldsymbol{\alpha}, \boldsymbol{\beta})$, whose entries are given by (6.5.4)–(6.5.8) as

$$\frac{-\partial^2 \ell}{\partial \alpha_j \partial \alpha_m} = I(j = m) \frac{d_j}{\alpha_j^2}$$

$$\frac{-\partial^2 \ell}{\partial \beta_r \partial \beta_s} = \sum_{j=1}^{k} \alpha_j \left[\sum_{i=1}^{n} \Delta_j(t_i) x_{ir} x_{is} e^{\boldsymbol{\beta}' \mathbf{x}_i} \right]$$

$$\frac{-\partial^2 \ell}{\partial \alpha_j \partial \beta_r} = \sum_{i=1}^{n} \Delta_j(t_i) x_{ir} e^{\boldsymbol{\beta}' \mathbf{x}_i}.$$

Inference procedures for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and characteristics of the lifetime distribution for T given $\mathbf{x}$ are easily implemented using the standard approaches described in Appendix C. We will find in Chapter 7 that there is a close connection between these procedures and ones developed for the semiparametric PH model.

The assumption of PH is strong and it is important that it be checked. The multiplicative model (6.5.2) can be expanded to allow for nonproportionality. One way is through the use of time–covariate interactions, as, for example, in

$$h(t|x) = h_0(t; \boldsymbol{\alpha}) \exp[\beta x + \gamma x g(t)], \tag{6.5.12}$$

where $g(t)$ is a specified function. Such models are easily fitted, though depending on the form of $g(t)$ it may be necessary to use numerical integration in the computation of the log-likelihood (6.5.1). Another way to interpret (6.5.12) is as a model in which β is allowed to vary with t:

$$h(t|\mathbf{x}) = h_0(t; \boldsymbol{\alpha}) \exp[\boldsymbol{\beta}(t)' \mathbf{x}], \tag{6.5.13}$$

where $\boldsymbol{\beta}(t)$ is a parametrically specified vector. The model (6.5.12) has a single covariate, with $\boldsymbol{\beta}(t) = \beta + \gamma g(t)$.

Models like (6.5.12) are log linear for $h(t|\mathbf{x})$; they are of the form

$$\log h(t|\mathbf{x}) = g_0(t; \boldsymbol{\alpha}) + g_1(\mathbf{x}; \boldsymbol{\beta}) + g_2(\mathbf{x}, t; \gamma). \tag{6.5.14}$$

Additive models for which $h(t|\mathbf{x})$ has the form on the right-hand side of (6.5.14) are also useful; they are generally easier to work with in terms of likelihood computations, but on the other hand, may require constraints on parameters to make $h(t|\mathbf{x})$ positive. In addition to polynomial specifications for g_0, g_1, and g_2 in (6.5.14), piecewise functions such as regression splines can be used to achieve a great deal of flexibility. The Bibliographic Notes reference some work in this direction.

6.5.2 Mechanistic Models

Sometimes there is knowledge about the mechanics of the failure process that points to certain types of models. Some of the accelerated failure time models used when testing materials or products at high temperature or voltage stress levels are motivated by such considerations, for example. A discussion of physical failure processes is beyond the scope of this book, but we will note a pair of settings that motivate classes of models.

The first setting concerns situations where failure is largely determined by a process of deterioration within individuals or items. In that case covariates may affect rates of deterioration, and thus times to failure; they may also affect the strength or (conversely) the susceptibility of an individual to failure. The following example illustrates this type of situation.

Example 6.5.2. The inverse Gaussian distribution was introduced in Section 1.3.7, and can be motivated as the distribution for the time T until a Wiener diffusion process with positive drift coefficient γ and dispersion parameter σ crosses a threshold level d. The lifetime distribution $IG(\mu, \lambda)$ given by (1.3.23) has parameters defined as $\mu = d/\gamma$, $\lambda = d^2/\sigma^2$. Regression models of different types can be considered. One form takes the distribution of lifetime T, given a vector of covariates $\mathbf{x}$, to be $IG(\mu(\mathbf{x}), \lambda)$, with $\mu(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$. A more plausible model in many cases is one with $\mu(\mathbf{x}) = (\boldsymbol{\beta}'\mathbf{x})^{-1}$, as considered by Whitmore (1983). This can be motivated by allowing the drift rate γ in the corresponding Wiener process to depend linearly on covariates; the choice $\gamma = \boldsymbol{\beta}'\mathbf{x}$ leads to $\mu(\mathbf{x}) = (\boldsymbol{\beta}'\mathbf{x})^{-1}$, provided that d and σ do not depend on $\mathbf{x}$. A third model is one that assumes that the threshold level d associated with failure depends on $\mathbf{x}$; this gives a model $IG(\mu, \lambda)$ in which both μ and λ vary with $\mathbf{x}$.

As an example we consider the data on times to failure of 20 aluminum-reduction cells given in Example 4.2.2. There was actually a covariate x associated with the units: the failure times (repeated from Example 4.2.2) and corresponding covariate values are as follows.

t_i: .468, .725, .838, .853, .965, 1.139, 1.142, 1.304, 1.317, 1.427
1.554, 1.658, 1.764, 1.776, 1.990, 2.010, 2.224, 2.279*, 2.244*, 2.286*

x_i: 51, 64, 90, 83, 61, 78, 91, 88, 12, 42, 95, 92, 92, 76, 19, 34, 6, 7, 42, 0

Asterisks denote censoring times. Both types of models for $\mu(x)$ just described fit the data quite well. The model with $\mu(x) = (\beta_0 + \beta_1 x)^{-1}$ is readily fitted by maximizing the log-likelihood function $\ell(\beta_0, \beta_1, \lambda)$ given by (4.2.8), with μ replaced by $\mu_i = (\beta_0 + \beta_1 x_i)^{-1}$. This gives estimates and standard errors $\hat{\beta}_0 = .360(.132)$, $\hat{\beta}_1 = .00515(.00207)$, $\hat{\lambda} = 7.453(2.546)$, and asymptotic correlation estimates $\text{corr}(\hat{\beta}_0, \hat{\beta}_1) = -.844$, $\text{corr}(\hat{\beta}_0, \hat{\lambda}) = .216$, $\text{corr}(\hat{\beta}_1, \hat{\lambda}) = -.163$.

Model checks can be carried out by using uniform residuals $\hat{u}_i = F(t_i; \hat{\mu}_i, \hat{\lambda})$ where the c.d.f. $F(t; \mu, \lambda)$ is given by (1.3.24) and $\hat{\mu}_i = (\hat{\beta}_0 + \hat{\beta}_1 x_i)^{-1}$. A plot of $\hat{u}_i$ versus x_i (noting that three residuals correspond to censored lifetimes) and a uniform probability plot of the $\hat{u}_i$ both give no reason to doubt the model. Note that the uniform probability plot is carried out as described in Section 3.5.1. That is, we plot the 17 points $(u_j^*, 1 - \hat{S}_j^*)$, $j = 1, \ldots, 17$, where u_j^* is the jth smallest value among the uncensored residuals $\hat{u}_1, \ldots, \hat{u}_{20}$, and $\hat{S}_j^* = .5[\hat{S}(u_j^*) + \hat{S}(u_j^*+)]$, with $\hat{S}(u)$ the Kaplan–Meier estimate obtained from $\hat{u}_1, \ldots, \hat{u}_{20}$ and the censoring indicators.

The second setting we mention is where background knowledge suggests that individuals vary considerably in terms of their susceptibility to failure, to the extent that a mixture model is considered. An extreme but important case is where some individuals are immune from failure. This leads to a mixture model (6.1.12) for the distribution of T given $\mathbf{x}$. Here, $p(\mathbf{x})$ is the fraction of individuals with covariate vector $\mathbf{x}$ who are susceptible to failure, and $S_0(t|\mathbf{x})$ is the s.f. for such individuals. Both $p(\mathbf{x})$ and $S_0(t|\mathbf{x})$ can be specified as parametric models.

Example 6.5.3. (Example 4.4.1 Revisited). In Example 4.4.1 we fitted separate mixture models of the form (4.4.1) to the Control and Therapy groups for the time to recurrence of colon cancer in patients entered in a randomized clinical trial. It is of some interest to consider whether the two groups differ mainly in the probability of recurrence, or whether the times to recurrence are also distributed differently. The only covariate is the treatment indicator $x = I$ (Therapy group), and log-logistic models were found to fit the data well in Example 4.4.1, so let us now consider the model (6.1.12) with

$$S_0(t|x) = [1 + (t/\alpha(x))^{\beta(x)}]^{-1}. \tag{6.5.15}$$

The separate models fitted for the two treatment groups in Example 4.4.1 correspond to (6.5.15) with six independent parameters $p(x)$, $\alpha(x)$, $\beta(x)$, $x = 0, 1$; let us call this model M_2. It is clear from plots of the data in Figure 4.8 that $p(0)$ and $p(1)$ are different, but let us fit models M_1, with $\beta(0) = \beta(1)$, but $\alpha(0)$ and $\alpha(1)$ allowed to be different, and M_0, with $\beta(0) = \beta(1)$ and $\alpha(0) = \alpha(1)$.

Table 6.7. Estimates and Maximum Likelihood Values for Three Models

Model	$\ell_{\max}$	$p(0)$	$p(1)$	$\alpha(0)$	$\alpha(1)$	$\beta(0)$	$\beta(1)$
M_0	−2564.66	.610	.426	444.5	444.5	1.61	1.61
M_1	−2564.01	.605	.430	415.9	485.6	1.62	1.62
M_2	−2563.91	.608	.426	419.5	479.0	1.58	1.68

Table 6.7 shows m.l.e.'s and maximum log-likelihood values $\ell_{\max} = \ell(\hat{\mathbf{p}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ for the three models. It is clear from the log-likelihoods that neither of the models M_0 and M_1 is contradicted by the data, when compared with M_2. In particular, the likelihood ratio statistic for testing M_0 versus M_2 gives the observed value

$$\Lambda = 2(-2563.91) - 2(-2564.66) = 1.50$$

and a p-value on $\chi^2_{(2)}$ of .47.

A model in which persons in either treatment group who experience recurrence have the same time-to-recurrence distribution is therefore supported by the data. The proportion of persons who experience an eventual recurrence is, however, significantly lower for the therapy group. Plots of the survivor functions for the two groups under model M_0 are nearly indistinguishable from those in Figure 4.8, and approximate the Kaplan–Meier estimates of $S(t)$ very closely.

6.5.3 Transformations and Some Other Models

Transformation families can be used to define families of lifetime models. We will comment briefly on two such approaches.

First, the log-location-scale family of models can be expanded via transformations. In particular, we might consider a parametric family of monotonic transformations $g(t; \boldsymbol{\alpha})$ that maps $(0, \infty)$ onto $(-\infty, \infty)$, and assume that $g(T; \boldsymbol{\alpha})$ has for some $\boldsymbol{\alpha}$ a location-scale distribution. This gives a survivor function for T given $\mathbf{x}$ of the form

$$S(t|\mathbf{x}) = S_0\left[\frac{g(t; \boldsymbol{\alpha}) - u(\mathbf{x}; \boldsymbol{\beta})}{b}\right], \tag{6.5.16}$$

where $S_0(z)$ is a survivor function on $(-\infty, \infty)$. The log-location-scale models are given by the special case $g(t) = \log t$. In terms of the plotting techniques discussed in Section 6.2.1, this model has the property that if we plot $S(t|\mathbf{x}_j)$ or any monotonic function of it against $g(t; \boldsymbol{\alpha})$ for two covariate vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, the plots will be (horizontal) translations of one another. The extended power family of transformations $g(x; \alpha) = (x^\alpha - 1)/\alpha$ is sometimes useful; it includes $g(x; 0) = \log x$ as the limit of $g(x; \alpha)$ as $\alpha \to 0$.

Another family of models is obtained by considering transformations $\psi(s)$ that map $(0,1)$ onto $(-\infty, \infty)$, and assuming that for some ψ the survivor function $S(t|\mathbf{x})$

satisfies (6.1.11):

$$\psi[S(t|\mathbf{x})] = \psi[S_0(t)] + w(\mathbf{x}; \boldsymbol{\beta}), \tag{6.5.17}$$

where $S_0(t)$ is a baseline survivor function The PH family is given by the special case where $\psi(s) = \log(-\log s)$. The case where $\psi(s) = \log((1-s)/s)$ is called the proportional odds (PO) family (see Problem 6.3). A broader class of models is based on the parametric family of transformations $\psi(s) = \log((s^{-\alpha} - 1)/\alpha)$; the special cases $\alpha \to \infty$ and $\alpha = 1$ give the PH and PO families. A way to think about the models (6.5.17) is that we seek a function ψ such that plots of $\psi[S(t|\mathbf{x}_1)]$ and $\psi[S(t|\mathbf{x}_2)]$ versus t (or any monotone function of t) are vertical translations of one another.

Fully parametric versions of (6.5.16) are obtained by assuming parametrically specified or known functions $g(t)$ and $S_0(t)$. Similarly, fully parametric versions of (6.5.17) assume known or parametrically specified functions $\psi(s)$ and $S_0(t)$. Semiparametric models, which are considered in Chapters 7 and 8, leave some of these functions unspecified. For example, we might assume in (6.5.17) that $\psi(s) = \log(-\log s)$ and that $S_0(t)$ is unspecified; this is the semiparametric PH model considered in Chapter 7.

In spite of their seeming flexibility, models of the form (6.5.16) and (6.5.17) make the strong assumption that the survivor functions $S(t|\mathbf{x})$ are ordered in the sense that $S(t|\mathbf{x}_1)$ and $S(t|\mathbf{x}_2)$ do not cross. As for log-location-scale or PH models, this can be relaxed at the cost of more complexity, for example, by letting b in (6.5.16) depend on $\mathbf{x}$. When models with one or more shape or transformation parameters are considered, there is generally a range of models that fit the data adequately, and there is a natural inclination to choose models that yield simple interpretations. As a result, the AFT, PH, and PO models are widely used, and the broader families (6.5.16) or (6.5.17) much less so.

The carbon-fiber breaking strength data of Example 6.4.2 provide an illustration of the points just made. In this case, there are only four levels for the covariate x (log fiber length) and a relatively large number of observations at each level, so it is feasible to examine both the shapes of the distribution of T given x, and the effect of x on these distributions. The probability plots based on the log-Burr AFT (log-location-scale) model shown in Figure 6.16 indicate that a model with parameter k in (6.4.5) approximately equal to 2 provides a good fit to the data. The data for $\ell = 10$ ($x = \log 10$) display a mild departure from the log-Burr distribution, and the four plots in the figure indicate the possibility that k varies with x, but formal hypothesis tests for the dependence of k or the scale parameter b in (6.4.5) on x do not provide significant evidence against the basic model. Because of its simplicity and the fact that the AFT model with a log linear effect for fiber length ℓ is physically plausible, we are inclined to base our conclusions on it. However, other models that fit the data well can be found. An examination of Figure 6.16 suggests, for example, that models of the form (6.5.17) with a suitably chosen function $\psi(s)$ would be satisfactory; the choice $\psi(s) = \log((s^{-.5} - 1)/.5)$ corresponds to the vertical scale in Figure 6.16, and so gives a good fit. The figure also indicates, however, that no choice of $S_0(t)$ in (6.5.17) will provide a substantially better fit than the log-Burr AFT model. We are consequently reluctant to replace it with a model (6.5.17) in which the interpretation of the length effect is less natural.

BIBLIOGRAPHIC NOTES

Numerous books discuss linear and nonlinear regression models based on normal error distributions and on least-square procedures, for example, Weisberg (1985), Ryan (1997), and Seber and Wild (1989). Cox and Oakes (1984, Ch. 5) discuss various ways in which covariates might affect lifetimes. General discussions of methodology for accelerated failure time and location-scale models were given by Kalbfleisch and Prentice (1980), Lawless (1982), Cox and Oakes (1984), Nelson (1990), Crowder et al. (1991), and Meeker and Escobar (1998), the last three emphasizing accelerated life testing and reliability applications. Gentleman and Crowley (1991) and Bowman and Wright (2000) are useful additional references on graphical methods. The use of nonparametric regression techniques such as generalized additive modeling (Hastie and Tibshirani 1990) should also be mentioned as an exploratory tool.

Cox and Snell (1968) gave an early discussion of generalized residuals. Lawless (1986) considered probability plots and relationships between families of models such as accelerated failure time, proportional hazards, and proportional odds. There has been relatively little detailed study of the properties of residuals under censoring, though see Baltazar–Aban and Pena (1995). Influence analysis for censored lifetime data has been considered by Hall et al. (1982), Weissfeld and Schneider (1990), and Escobar and Meeker (1992). Atkinson (1985) and Cook and Weisberg (1982) are basic references for uncensored data.

Lawless (1982, Ch. 6) was the precursor to the treatment of log-location-scale models given in Section 6.3, especially for the extreme value and normal models. Bennett (1983a) considered log-logistic models, Farewell and Prentice (1977) the log-gamma models of Section 6.4.1, and Lancaster and Nickell (1980) the log-Burr models of the same section. Kalbfleisch and Prentice (1980, Ch. 3) and Ciampi et al. (1986) discussed a log–F family that includes all of the aforementioned models as special cases.

Cox and Hinkley (1968) developed most of the results in Section 6.3.4; Gould and Lawless (1988) and Silvapulle (1985) demonstrated robustness properties of m.l.e.'s $\hat{\boldsymbol{\beta}}$ within location-scale families and considered efficiency. Experimental designs for linear and additive models are described in many books (e.g., Box et al. 1978; Wu and Hamada 1999). Many of the key ideas of experimental design and analysis of variance are applicable to the analysis of log-lifetimes, though censoring complicates matters and destroys properties such as parameter orthogonality. Early work on accelerated life test designs was done by Nelson and Kielpinski (1976) and Nelson and Meeker (1978). Meeker and Escobar (1998, Ch. 20) and Nelson (1990) are invaluable references for accelerated life tests, with many references to the literature and practical guidance for designing studies. Chaloner

and Larntz (1992) discuss design from a Bayesian perspective; also see Chaloner and Verdinelli (1995). Hamada (1995) and Hamada and Wu (1991, 1995) consider fractional factorial life test experiments; Wu and Hamada (1999) give many references. Lawless and Singhal (1978) discuss all-subsets regression for location-scale models.

Location-scale models with nonconstant scale parameters have been considered many times in the literature (e.g., Nelson 1984; Smyth 1989; Anderson 1991). Meeter and Meeker (1994) discuss life test design for this case. So-called parameter design studies for reliability improvement (e.g., Taguchi 1986; Condra 1993; Wu and Hamada 1999) often involve nonconstant scale parameters. Nelson (1990, Ch. 10) discusses time-varying covariates in the context of accelerated life testing. Bagdonavicius and Nikulin (2002) give a thorough account of many models for dealing with time-varying covariates.

Other regression models, like those in Section 6.5, have been considered by many authors. Most hazard-based modeling tends to be semiparametric, but parametric methods are very flexible. Models with piecewise-constant hazard functions have been studied by Holford (1976), Laird and Olivier (1981), and Friedman (1982). Carstensen (1996) and Kim (1997) fitted such models to interval censored data, and Lindsey and Ryan (1998) considered piecewise-constant and piecewise-polynomial models in the same setting. Kooperberg et al. (1995) and Kooperberg and Clarkson (1997) consider regression spline hazard-based models for right-censored and interval-censored lifetimes, respectively. Additive hazards models for discrete or grouped data have been considered by Aranda-Ordaz (1983), Breslow and Day (1987), and others. Inverse Gaussian regression models are considered by Whitmore (1983) and others, and mixtures like those in Example 6.5.4 by Farewell (1982), Maller and Zhou (1996), and Sy and Taylor (2000). Other types of mechanistic models arising, for example, from physics of failure considerations, can be found in many fields of application. See, for example, LuValle (1993), LuValle et al. (1988), and Meeker and LuValle (1995). Transformation models such as (6.5.17) have also been considered quite widely; for parametric approaches, see, for example, Mackenzie (1996) and Younes and Lachin (1997).

COMPUTATIONAL NOTES

Methodology for the accelerated failure time or log-location-scale regression models discussed in Sections 6.1 to 6.3 is available in many commercial packages, including S-Plus (see function censorReg) and SAS (see procedure LIFEREG). The reliability data-analysis system SPLIDA (Meeker 2002) has many special features, including experimental design capabilities. Some packages also handle the log-Burr or log-gamma models of Section 6.4.1. Software for dealing with variable scale parameters is not available in the major packages, nor is methodology for time-varying covariates. The models in Section 6.5 are likewise not handled by major packages, although the models with piecewise-constant hazards in Section 6.5.1 can be implemented using log linear generalized linear model software.

PROBLEMS AND SUPPLEMENTS

6.1 Show that the only models in both the proportional hazards family (6.1.6) for T given $\mathbf{x}$ and the accelerated failure time family (6.1.5) for T given $\mathbf{x}$ are Weibull distributions of the form (6.1.1).

(Section 6.1; Lawless 1986)

6.2 Consider accelerated failure time models (6.1.5) in the case of a single indicator covariate, $x = 0$ or 1, and examine the ratio of the hazard functions $h(t|x)$ for $x = 0$ and 1 for the following two models:

(a) $S_0^*(t) = (1+t)^{-1}$ in (6.1.5) is of log-logistic form.

(b) $S_0^*(t) = \exp(-t)$ in (6.1.5) is of Weibull form, but δ depends on x.

In each case assume that $\alpha(x) = \exp(\beta_0 + \beta_1 x)$.

(Section 6.1)

6.3 *Proportional odds models.* The proportional odds family of regression models is of the form (6.1.11), with $\psi(p) = \log((1-p)/p)$. Note that for such models the odds ratio for the probability of survival past time t, for any two covariate values, $\mathbf{x}_1$ and $\mathbf{x}_2$, is independent of t:

$$\frac{[1 - S(t|\mathbf{x}_1)]/S(t|x_1)}{[1 - S(t|\mathbf{x}_2)]/S(t|x_2)} = \exp[w(\mathbf{x}_1; \boldsymbol{\beta}) - w(\mathbf{x}_2; \boldsymbol{\beta})].$$

(a) Show that the survivor function $S(t|\mathbf{x})$ for proportional odds (PO) models can be written in the form

$$S(t|\mathbf{x}) = [1 + \exp(A_0(t) + w(\mathbf{x}; \boldsymbol{\beta}))]^{-1}, \qquad (6.6.1)$$

where $A_0(t) = \log[(1 - S_0(t))/S_0(t)]$.

(b) Show that the log-logistic regression model for which

$$S(t|\mathbf{x}) = [1 + (t/\alpha(\mathbf{x}))^{\delta}]^{-1} \qquad (6.6.2)$$

is both a PO and an accelerated failure time (AFT) model. Show that any regression model that is in both the PO and AFT families must be of the form (6.6.2).

(Section 6.1; Bennett 1983b; Lawless 1986)

6.4 Derive the results (6.3.21) and (6.3.22) concerning the mean lifetime of left truncated normal and logistic random variables.

(Section 6.2.2)

6.5 *Concavity of the log-likelihood for location-scale models.* Consider the log-likelihood function (6.3.2) for a location-scale model with $u(\mathbf{x}; \boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{x}$ and right-censored data. Rewrite the log-likelihood as $\ell(\boldsymbol{\gamma}, \phi)$, where the parameters $\boldsymbol{\gamma}$ and ϕ are defined as

$$\boldsymbol{\gamma} = b^{-1}\boldsymbol{\beta}, \qquad \phi = b^{-1}.$$

Note that $z_i = [y_i - \boldsymbol{\beta}'\mathbf{x}_i]/b = \phi y_i - \boldsymbol{\gamma}'\mathbf{x}_i$ is linear in ϕ and $\boldsymbol{\gamma}$. It then follows from results on convex functions (Rockafellar 1970) that $\ell(\boldsymbol{\gamma}, \phi)$ is always concave with respect to $\boldsymbol{\gamma}$ and ϕ if $\log f_0(z)$ and $\log S_0(z)$ are concave with respect to z.

(a) Prove that if $\log f_0(z)$ is concave, that is, $d^2 \log f_0(z)/dz^2 < 0$ for all z, then $\log S_0(z)$ is also concave.

(b) Show for the standard extreme value, normal, and logistic p.d.f.'s that $\log f_0(z)$ is concave. This implies that $\ell(\boldsymbol{\gamma}, \phi)$ is concave for these models and that if the m.l.e. $(\hat{\boldsymbol{\gamma}}, \hat{\phi})$ exists it is unique.

Burridge (1981) shows that these results hold more generally for interval censored y's.

(Section 6.3; Burridge 1981)

6.6 *Nonexistence of the m.l.e. for a regression coefficient.* With location-scale models and censored data as in Problem 6.5, it is possible for some components of the m.l.e.'s $\hat{\boldsymbol{\beta}}$ or $\hat{\boldsymbol{\gamma}}$ to be at $+\infty$ or $-\infty$. That is, the log-likelihood is strictly increasing in some direction, so there does not exist a maximum at any finite $\hat{\boldsymbol{\beta}}$ or $\hat{\boldsymbol{\gamma}}$. To illustrate, suppose that Y is extreme value with $b = 1$ and $u(\mathbf{x}; \boldsymbol{\beta}) = \beta_0 + \beta_1 x$, where x takes only values 0 and 1. The log-likelihood (6.3.2) in terms of $t_i = \exp(y_i)$ and censoring indicators δ_i is then (with $r = \sum \delta_i$)

$$\ell(\beta_0, \beta_1) = -r\beta_0 - \beta_1 \sum_{i=1}^{n} \delta_i x_i - \sum_{i=1}^{n} t_i e^{-\beta_0 - \beta_1 x_i}.$$

Show that if $\sum \delta_i x_i = 0$ (and not all x_i are 0), then $\partial \ell / \partial \beta_1 > 0$ for all finite (β_0, β_1). Thus, the m.l.e. $\hat{\beta}_1$ does not exist.

Silvapulle and Burridge (1986) provide general conditions for the nonexistence of $\hat{\boldsymbol{\beta}}$ in location-scale models, as well as other models.

(Section 6.3)

6.7 For the location-scale regression model (6.1.3) with $u(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$, prove that the m.l.e.'s $\hat{\boldsymbol{\beta}}$ and $\hat{b}$ from a complete random sample $(y_i|\mathbf{x}_i)$, $i = 1, \ldots, n$ satisfy the equivariance conditions (E7) of Appendix E. Show that the least-square estimator of $\boldsymbol{\beta}$ and the corresponding residual mean square estimator of b are also equivariant.

(Section 6.3; Appendix E)

6.8 *Exact inference for exponential regression models.* Consider the exponential regression model where $S(t|x)$ depends on a scalar covariate and is of the form (6.1.1), with $\delta = 1$ and $\alpha(x) = \exp(\beta_0 + \beta_1 x)$. Note that this corresponds to an extreme value location-scale model for $Y = \log T$ given x, with scale parameter $b = 1$. Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the m.l.e.'s from an uncensored random sample $(y_i|\mathbf{x}_i)$, $i = 1, \ldots, n$, and recall from Problem 6.7 that $\hat{\beta}_0$ and $\hat{\beta}_1$ are

equivariant. Let $\mathbf{a} = (a_1, \ldots, a_n)$, with $a_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$, and define the pivotal quantities $Z_0 = \hat{\beta}_0 - \beta_0$, $Z_1 = \hat{\beta}_1 - \beta_1$.

Use (E9) of Appendix E to show that, given $\mathbf{a}$, the density function of (Z_0, Z_1), is of the form

$$g(z_0, z_1|\mathbf{a}) = k(\mathbf{a}, \mathbf{x}, n) \exp\left[nz_0 + \sum_{i=1}^{n} (a_i + z_1 x_i) - e^{z_0} \sum_{i=1}^{n} e^{a_i + z_1 x_i} \right],$$

where $-\infty < z_0 < \infty$, $-\infty < z_1 < \infty$, and $\mathbf{x} = (x_1, \ldots, x_n)$. Thus show that the p.d.f. of Z_1, given $\mathbf{a}$, is of the form

$$g_1(z_1|\mathbf{a}) = k_1(\mathbf{a}, \mathbf{x}, n) e^{nz_1\bar{x}} \Big/ \left(\sum_{i=1}^{n} e^{a_i + z_1 x_i} \right)^n.$$

(Section 6.3.1; Lawless 1976)

6.9 The data in Table 6.8 are from a more comprehensive set given by Krall et al. (1975). The problem is to relate survival times for multiple myeloma patients to a number of prognostic variables. The data given here show survival times, in months, for 65 patients and include measurements on each patient for the following five covariates:

x_1 Logarithm of a blood urea nitrogen measurement at diagnosis
x_2 Hemoglobin measurement at diagnosis
x_3 Age at diagnosis
x_4 Sex: 0, male; 1, female
x_5 Serum calcium measurement at diagnosis

Asterisks denote censoring times.

Examine the relationship of these variables to survival time by fitting Weibull regression models of the form (6.1.1) with $\alpha(\mathbf{x}) = \exp(\boldsymbol{\beta}'\mathbf{x})$. Assess the fit of the models on which you choose to base your conclusions.

(Section 6.3; Lawless and Singhal 1978)

6.10 McCool (1980) gives the failure times for hardened steel specimens in a rolling contact fatigue test shown in Table 6.9; 10 independent observations were taken at each of four values of contact stress.

(a) Engineering background suggests that at stress level s failure time should have approximately a Weibull distribution with a scale parameter α related to s by a power law relationship $\alpha = cs^p$ and with a shape parameter δ that is independent of s. Assess this model graphically and by a formal test of it against the alternative that failure times at the ith stress level have a Weibull distribution with parameters α_i and δ ($i = 1, 2, 3, 4$). Are you

Table 6.8. Survival Times and Covariates for Multiple Myeloma Patients

t	x_1	x_2	x_3	x_4	x_5	t	x_1	x_2	x_3	x_4	x_5
1	2.218	9.4	67	0	10	26	1.230	11.2	49	1	11
1	1.940	12.0	38	0	18	32	1.322	10.6	46	0	9
2	1.519	9.8	81	0	15	35	1.114	7.0	48	0	10
2	1.748	11.3	75	0	12	37	1.602	11.0	63	0	9
2	1.301	5.1	57	0	9	41	1.000	10.2	69	0	10
3	1.544	6.7	46	1	10	42	1.146	5.0	70	1	9
5	2.236	10.1	50	1	9	51	1.568	7.7	74	0	13
5	1.681	6.5	74	0	9	52	1.000	10.1	60	1	10
6	1.362	9.0	77	0	8	54	1.255	9.0	49	0	10
6	2.114	10.2	70	1	8	58	1.204	12.1	42	1	10
6	1.114	9.7	60	0	10	66	1.447	6.6	59	0	9
6	1.415	10.4	67	1	8	67	1.322	12.8	52	0	10
7	1.978	9.5	48	0	10	88	1.176	10.6	47	1	9
7	1.041	5.1	61	1	10	89	1.322	14.0	63	0	9
7	1.176	11.4	53	1	13	92	1.431	11.0	58	1	11
9	1.724	8.2	55	0	12	4*	1.945	10.2	59	0	10
11	1.114	14.0	61	0	10	4*	1.924	10.0	49	1	13
11	1.230	12.0	43	0	9	7*	1.114	12.4	48	1	10
11	1.301	13.2	65	0	10	7*	1.532	10.2	81	0	11
11	1.508	7.5	70	0	12	8*	1.079	9.9	57	1	8
11	1.079	9.6	51	1	9	12*	1.146	11.6	46	1	7
13	.778	5.5	60	1	10	11*	1.613	14.0	60	0	9
14	1.398	14.6	66	0	10	12*	1.398	8.8	66	1	9
15	1.602	10.6	70	0	11	13*	1.663	4.9	71	1	9
16	1.342	9.0	48	0	10	16*	1.146	13.0	55	0	9
16	1.322	8.8	62	1	10	19*	1.322	13.0	59	1	10
17	1.230	10.0	53	0	9	19*	1.322	10.8	69	1	10
17	1.591	11.2	68	0	10	28*	1.230	7.3	82	1	9
18	1.447	7.5	65	1	8	41*	1.756	12.8	72	0	9
19	1.079	14.4	51	0	15	53*	1.114	12.0	66	0	11
19	1.255	7.5	60	1	9	57*	1.255	12.5	66	0	11
24	1.301	14.6	56	1	9	77*	1.079	14.0	60	0	12
25	1.000	12.4	67	0	10						

Table 6.9. Failure Times for Steel Specimens at Four Stress Levels

Stress (psi $\div 10^6$)	Ordered Failure Times
.87	1.67, 2.20, 2.51, 3.00, 2.90, 4.70, 7.53, 14.70, 27.8, 37.4
.99	.80, 1.00, 1.37, 2.25, 2.95, 3.70, 6.07, 6.65, 7.05, 7.37
1.09	.012, .18, .20, .24, .26, .32, .32, .42, .44, .88
1.18	.073, .098, .117, .135, .175, .262, .270, .350, .386, .456

Table 6.10. Failure Times for Epoxy Insulation Specimens at Three Voltage Levels

Voltage (kV)	Failure Times (min)
52.5	4690, 740, 1010, 1190, 2450, 1390, 350, 6095, 3000, 1458, 6200*, 550, 1690, 745, 1225, 1480, 245, 600, 246, 1805
55.0	258, 114, 312, 772, 498, 162, 444, 1464, 132, 1740*, 1266, 300, 2440*, 520, 1240, 2600*, 222, 144, 745, 396
57.5	510, 1000*, 252, 408, 528, 690, 900*, 714, 348, 546, 174, 696, 294, 234, 288, 444, 390, 168, 558, 288

Note: Censored observations are indicated by asterisks.

satisfied that δ can be considered the same for all four stresses? Are you satisfied with the assumption of a Weibull model?

(b) Fit a log-logistic distribution (6.6.2) to the data, with $\alpha = cs^p$ and δ constant. Compare the fit and conclusions under this model with those in part (a).

(Section 6.3; McCool 1980)

6.11 Stone (1978) reports an experiment in which specimens of solid epoxy electrical-insulation were studied in an accelerated voltage life test. In all, 20 specimens were tested at each of three voltage levels: 52.5, 55.0, and 57.5 kV. Failure times, in minutes, for the insulation specimens are given in Table 6.10.

(a) Examine whether the data at each voltage level might have arisen from a Weibull distribution. It may be necessary to consider three parameter distributions, since with the failure process involved here there is an initiation period during which failure does not normally occur. In this case, examine the possibility that each of the three parameters might depend on voltage.

(b) Does a Weibull model in which there is a constant threshold and shape parameter and a scale parameter related to voltage by a power law relationship $\alpha = cv^p$ appear plausible?

(Section 6.3.2; Stone 1978)

6.12 Consider the extreme value regression model (6.1.2) with $u(\mathbf{x}) = \alpha + \beta x$. Let $\hat{\alpha}$, $\hat{\beta}$, and $\hat{b}$ be the m.l.e.'s from a random sample $y_1, \ldots, y_n$ corresponding to covariate values $x_1, \ldots, x_n$. Obtain the p.d.f. of $Z_1 = (\hat{\alpha} - \alpha)/\hat{b}$, $Z_2 = (\hat{\beta} - \beta)/\hat{b}$, and $Z_3 = \hat{b}/b$, given the ancillary statistics $a_i = (y_i - \hat{\alpha} - \hat{\beta}x_i)/\hat{b}$ from (E9) of Appendix E. Show that z_1 can be integrated out of this density, to give the p.d.f. of Z_2 and Z_3, given $\mathbf{a}$, as

$$k(\mathbf{a}, \mathbf{x}) z_3^{n-2} \exp\left(z_3 \sum_{i=1}^{n} a_i\right) \Big/ \left(\sum_{i=1}^{n} e^{a_i z_3 + x_i z_2 z_3}\right)^n \qquad -\infty < z_2 < \infty, \quad z_3 > 0.$$

Note that double numerical integration is required to evaluate probabilities for Z_2 or Z_3.

(Section 6.3)

6.13 *Efficiency of least-square estimation.* Consider the linear location-scale regression model in the form (6.3.25), and assume that the covariates $x_1, \ldots, x_{p-1}$ are centered as in (6.3.26). Let $\hat{\beta}_0, \hat{\boldsymbol{\beta}}, \hat{b}$ denote the m.l.e.'s based on an uncensored random sample from (6.3.25), and let $\tilde{\beta}_0, \tilde{\boldsymbol{\beta}}$ denote the least-square estimators of β_0 and $\boldsymbol{\beta}$, given by (6.3.27).

(a) Derive (6.3.29) and thus prove (6.3.32).

(b) Calculate A_Z in (6.3.29) for the cases (i) where Z has an extreme value distribution with mean 0, that is, with p.d.f.

$$\exp[(z-\gamma) - \exp(z-\gamma)], \qquad -\infty < z < \infty,$$

where $\gamma = .5772$ is Euler's constant, and (ii) where Z has a standard logistic distribution with p.d.f. $\exp(z)/[1+\exp(z)]^2$. Thus show that when Z in (6.3.25) has an extreme value distribution the efficiency (6.3.32) is .61, and that when Z has a logistic distribution it is .91.

(c) Outline how you could extend your investigation to estimation of the scale parameter b and intercept β_0; assume that with least-squares you will use the estimator of $\sigma^2 = \mathrm{Var}(Z_i)$ given by

$$\tilde{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - \tilde{\beta}_0 - \tilde{\boldsymbol{\beta}}'\mathbf{x}_i)^2.$$

(Section 6.3.4; Cox and Hinkley 1968)

6.14 The data in Table 6.11 show the lifetimes (in km) of front disk brake pads on a randomly selected set of 40 cars (same model) that were monitored by a dealer network. Three factors are shown for each car:

Model Year	1 or 2
Driving Conditions	A = predominantly city,
	B = predominantly highway,
	C = mixed
Geographic region	N = Northern,
	S = Southern

Assess whether any of the factors appear related to lifetime. Provide a predictive model for brake-pad lifetime. Give a confidence interval for the median lifetime on a car used for mixed city and highway driving, taking the model year and region into account only if you feel it necessary.

(Section 6.3)

6.15 Table 6.12 shows the design for a multifactor experiment to improve the lifetimes of fluorescent lamps (Taguchi 1987). There are five two-level factors A, B, C, D, E, whose effects on lifetime are to be investigated. The proposed design is a 2^{5-2} fractional factorial with defining relations $D = AC$, $E = BC$ (see Box et al. 1978). The experiment was replicated twice, so there are 16 runs.

Table 6.11. Lifetimes of Disk Brake Pads on 40 Cars

Pad Life (1,000 km)	Car Year	GR	DC	Pad Life (1,000 km)	Car Year	GR	DC
86.2	1	S	C	48.8	2	S	C
45.1	1	N	C	81.7*	2	S	C
52.1	1	N	C	61.5	2	N	C
54.2	1	S	C	53.6	2	N	C
59.0	1	S	C	50.7*	2	S	C
38.4	1	N	C	42.8*	2	S	C
41.0*	1	S	C	102.5*	2	S	C
56.4	1	S	C	42.7	2	N	B
81.3*	1	S	B	80.6	2	N	B
62.4	1	N	B	64.5	2	S	B
45.5	1	N	B	73.1*	2	S	B
36.7	1	S	B	28.4	2	S	B
42.2*	1	S	B	46.9	2	N	A
51.6	1	S	A	45.9	2	S	A
34.4	1	S	A	33.8	2	S	A
22.7	1	S	A	59.8*	2	S	A
22.6	1	N	A	31.7	2	S	A
40.0	1	S	A	33.9	2	S	A
38.8	1	S	A	50.6	2	S	A
50.2	1	S	A	56.7	2	N	A

Note: Censored observations are indicated by asterisks.

Table 6.12. Design and Lifetimes for Fluorescent Lamp Experiment

Factor						
A	B	C	D	E	Lifetime	
1	1	1	1	1	(14,16)	(20,∞)
1	1	2	2	2	(18,20)	(20,∞)
1	2	1	1	2	(8,10)	(10,12)
1	2	2	2	1	(18,20)	(20,∞)
2	1	1	2	1	(20, ∞)	(20, ∞)
2	1	2	1	2	(12, 14)	(20, ∞)
2	2	1	2	2	(16,18)	(20,∞)
2	2	2	1	1	(12,14)	(14,16)

The lamp lifetimes were interval censored; they are also shown in Table 6.12 and are examined further in Problem 6.18.

Discuss what factor effects or contrasts can be examined in this study.

(Section 6.3.5, Hamada 1995)

6.16 The S-Plus software system includes some data on the times (in months) to recurrences of bladder cancer in 85 patients; the data set is contained in the

data frame bladder. Some individuals actually experienced several recurrences over the period of the study, but consider here only the time to first recurrence. Covariates included in the data are

rx: Treatment group (1 = Placebo, 2 = Drug Thiotepa)
number: The number of tumors present at the initial diagnosis of cancer
size: The size of the largest initial tumor

(a) Investigate the relationship of these covariates to the time to the first recurrence, using the AFT models of Section 6.3. Assess the fit of the models used.

(b) Repeat the analysis using a proportional hazards model with piecewise-constant baseline hazard, as in Example 6.5.1. Use models with three to five pieces.

(Sections 6.1–6.5; Wei et al. 1989)

6.17 Consider the fluorescent lamp experiment of Problem 6.15. The purpose of the study was to determine a combination of factor levels (i.e., levels 1 or 2 for each of factors A–E) that would maximize lifetime.

Engineering background suggested that the main effects of factors A–E and the AB interaction might be important.

a. Use a combination of plots and the AFT models of Section 6.3 to assess the factor effects. Comment on the precision with which the effects are estimated.

b. Make recommendations concerning the choice of levels for factors A–E in order to maximize median lifetime.

(Section 6.3; Hamada 1995; Wu and Hamada 1999)

6.18 Consider the data set of Problem 6.9. Fit a log-Burr model (6.4.5) and determine plausible values for the shape parameter, k. Does the evidence about which covariates are important vary much with k?

(Section 6.4)

6.19 *A factorial carcinogenicity experiment.* The data in Table 6.13 are from a nine-month study on the effect of known carcinogens DES and DMBA in the induction of mammary tumors in female rats (Shellabarger et al. 1980). The experimental animals were allocated to four treatment groups: Control (no carcinogen), DES only, DMBA only, and DES and DMBA. After treatment, the times to tumor appearance for the animals were noted; they are given in Table 6.13. All of the response times in the control group, and some in the other treatment groups, are censored.

Carry out an analysis of these data with a view to comparing the effects of the carcinogens on the time to tumor distribution. It is of special interest to

Table 6.13. Times to Mammary Tumor Appearance or Censoring

Control	DES only		DMBA only		DES and DMBA	
112*	64*	218*	94	192*(3)	57*	129(3)
266*(19)	88	224*	101	214*	67*	136(2)
	107	231*(2)	113*	224	88	143
	129(4)	238*	120*	229	94	144
	163(2)	238	163	235	100	191(2)
	191*(2)	256*	169*	252*	107	192
	192*	263*	170	260*	113	211
	200*	266*(2)	184*	266*(13)	123(2)	218
	210*(2)	266	190	266(2)	125	266*(2)
	217*(3)					

Note: Censored observations are indicated by asterisks; multiplicities are in parentheses.

ascertain whether there is an interaction, or synergistic effect, between DES and DMBA.

(Sections 6.1–6.4; Machado and Bailey 1985)

6.20 *Collapsible time-scale models.* Let $\mathbf{x}(t)$ be a vector of external time-varying covariates associated with the lifetime variable T. A collapsible model (Oakes 1995) is one for which

$$Pr(T > x|X) = S_0[t, \mathbf{x}(t)], \tag{6.6.3}$$

where $X = \{\mathbf{x}(t), t \geq 0\}$. That is, the probability that lifetime exceeds t depends only on the covariate values at t, and not prior values. For the case of a single covariate $x(t)$, several authors have considered linear time scale models for which

$$S_0[t, x(t)] = G[\alpha_0 t + \alpha_1 x(t)], \tag{6.6.4}$$

where G is a survivor function that may involve additional parameters β.

(a) Show that a time transformation model (6.4.17) is not in general expressible in the form (6.6.3). Give an example in which it is. (*Note:* $\mathbf{x}(t)$ in (6.6.3) could be defined as a function of the $\mathbf{x}(t)$ process in (6.4.17).)

(b) Investigate maximum likelihood estimation for the model (6.6.4) in the case where either failure or censoring times are observed for n independent individuals. What information is needed to obtain the likelihood function?

(Sections 6.4.3, 6.5; Duchesne and Lawless 2000)

$h(t|x) = h_0(t)\, r(x)$

$S(t|x) = S_0(t)^{r(x)}$

CHAPTER 7

Semiparametric Multiplicative Hazards Regression Models

7.1 METHODS FOR CONTINUOUS MULTIPLICATIVE HAZARDS MODELS

Models in which covariates have a multiplicative effect on the hazard function play a prominent role in the analysis of lifetime data. Proportional hazards (PH) models were described in Section 6.1 and some parametric versions were considered in Section 6.5. The present chapter deals with semiparametric PH models in which $S_0(t)$ in (6.1.7) or $h_0(t)$ in (6.1.6) is left arbitrary. Extensions to (6.1.6) are also considered, in which the multiplicative form is retained but the PH property is not; we refer to such models generally as multiplicative or log-additive hazard models.

Let T be a continuous lifetime variable and $\mathbf{x}$ a $p \times 1$ vector of fixed covariates. We will consider the PH model (6.1.6) in the case where $r(\mathbf{x}; \boldsymbol{\beta}) = \exp(\boldsymbol{\beta}'\mathbf{x})$, so that the hazard function for T given $\mathbf{x}$ takes the form

$$h(t|\mathbf{x}) = h_0(t) \exp(\boldsymbol{\beta}'\mathbf{x}), \tag{7.1.1}$$

with $\boldsymbol{\beta}$ a $p \times 1$ vector of regression coefficients. The procedures in this chapter can be applied to models with other forms for $r(\mathbf{x}; \boldsymbol{\beta})$ with obvious modification, but the exponential form is convenient and flexible enough for many purposes. Later (7.1.1) will be extended to allow $\mathbf{x}$ to be time-varying, but we assume for now that it is fixed. Finally, it should be noted that with the semiparametric PH model (7.1.1) no intercept term is included in $\boldsymbol{\beta}'\mathbf{x}$, because it is subsumed in $h_0(t)$.

The methods in this chapter are distribution-free in the sense that their validity and certain properties do not depend on the true form of $h_0(t)$, provided the multiplicative form (7.1.1) is correct. However, the form (7.1.1) is a strong assumption and requires careful checking in applications. The remainder of this section deals with inference procedures for the model and with methods of model assessment; examples are given in Section 7.2. Subsequent sections consider grouped or discrete lifetimes, and incomplete data.

The model (7.1.1) has two components, $\boldsymbol{\beta}$ and $h_0(t)$. Equivalently one can consider the baseline cumulative hazard function $H_0(t) = \int_0^t h_0(u)\,du$ or survivor function $S_0(t) = \exp[-H_0(t)]$ in place of $h_0(t)$. The survivor function for T given $\mathbf{x}$ is then

$$S(t|\mathbf{x}) = [S_0(t)]^{\exp(\boldsymbol{\beta}'\mathbf{x})}. \tag{7.1.2}$$

Given a censored random sample of lifetimes (t_i, δ_i), $i = 1, \ldots, n$, and corresponding covariate vectors $\mathbf{x}_i$, we want to estimate $\boldsymbol{\beta}$ and $S_0(t)$. Cox (1972a) introduced an ingenious way of estimating $\boldsymbol{\beta}$ without having to consider $S_0(t)$ explicitly; this is now known as the partial likelihood method. Because of its great simplicity and usefulness, methodology related to this approach will be described first, and the estimation of $H_0(t)$ and $S_0(t)$ after that.

7.1.1 Estimation and Tests for $\boldsymbol{\beta}$

Suppose that a censored random sample (t_i, δ_i), $i = 1, \ldots, n$, yields k distinct observed lifetimes $t_{(1)} < \cdots < t_{(k)}$ and $n - k$ censoring times. Let $R_i = R(t_{(i)})$ denote the set of individuals who are alive and uncensored just prior to time $t_{(i)}$; this is referred to as the risk set at $t_{(i)}$ since it consists of those individuals who could be observed to die at $t_{(i)}$, given what has occurred up to that time. Cox (1972a) suggested the following likelihood function for estimating $\boldsymbol{\beta}$ in (7.1.1):

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k} \left(e^{\boldsymbol{\beta}'\mathbf{x}_{(i)}} \Big/ \sum_{\ell \in R_i} e^{\boldsymbol{\beta}'\mathbf{x}_\ell} \right) \tag{7.1.3}$$

where $\mathbf{x}_{(i)}$ is the covariate associated with the individual observed to die at $t_{(i)}$. The motivation for (7.1.3) was that given $R(t)$ and that a death occurs at t, the probability it is individual $i \in R(t)$ who dies is

$$\frac{h(t|\mathbf{x}_i)}{\sum_{\ell \in R(t)} h(t|\mathbf{x}_\ell)} = \frac{e^{\boldsymbol{\beta}'\mathbf{x}_i}}{\sum_{\ell \in R(t)} e^{\boldsymbol{\beta}'\mathbf{x}_\ell}},$$

directly from (7.1.1). However, as presented (7.1.3) is not a likelihood in the usual sense, since it does not arise from the probability of some observable outcome (see Appendix C). It turns out that $L(\boldsymbol{\beta})$ can be treated as an ordinary likelihood, though: maximization of (7.1.3) yields an estimator $\hat{\boldsymbol{\beta}}$ which is consistent and asymptotically normal under suitable conditions, and score, information, and likelihood ratio statistics based on $L(\boldsymbol{\beta})$ behave as though it is an ordinary likelihood.

Formal justification of $L(\boldsymbol{\beta})$ will be taken up in Section 7.1.3. For now its validity is assumed, and likelihood inference procedures based on it will be considered. In doing this it is convenient to rewrite (7.1.3) in a slightly different form. For individual i, define as in Section 2.2.2

$$Y_i(t) = I(t_i \geq t) \qquad i = 1, \ldots, n. \tag{7.1.4}$$

Then $Y_i(t) = 1$ if and only if $i \in R(t)$, and (7.1.3) can be rewritten as

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{e^{\boldsymbol{\beta}'\mathbf{x}_i}}{\sum_{\ell=1}^{n} Y_\ell(t_i) e^{\boldsymbol{\beta}'\mathbf{x}_\ell}} \right)^{\delta_i}. \tag{7.1.5}$$

Note that (7.1.5) is also defined when ties in lifetimes occur, that is, when two or more individuals (with $\delta_i = 1$) have the same observed lifetime. Additional discussion of ties is provided in Section 7.1.3, but for now we assume they are accommodated under (7.1.5).

The log-likelihood function from (7.1.5) is

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left[\boldsymbol{\beta}'\mathbf{x}_i - \log \left(\sum_{\ell=1}^{n} Y_\ell(t_i) e^{\boldsymbol{\beta}'\mathbf{x}_\ell} \right) \right]. \tag{7.1.6}$$

The score vector $\mathbf{U}(\boldsymbol{\beta}) = (\partial\ell/\partial\beta_1, \ldots, \partial\ell/\partial\beta_p)'$ and information matrix take simple forms. Define for any $t > 0$ the $p \times 1$ vector

$$\bar{\mathbf{x}}(t, \boldsymbol{\beta}) = \frac{\sum_{\ell=1}^{n} Y_\ell(t)\mathbf{x}_\ell e^{\boldsymbol{\beta}'\mathbf{x}_\ell}}{\sum_{\ell=1}^{n} Y_\ell(t) e^{\boldsymbol{\beta}'\mathbf{x}_\ell}}, \tag{7.1.7}$$

which is a weighted average of the covariate vectors of individuals at risk at time t. Then it is easily seen that

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i [\mathbf{x}_i - \bar{\mathbf{x}}(t_i, \boldsymbol{\beta})]. \tag{7.1.8}$$

In addition, the $p \times p$ information matrix $I(\boldsymbol{\beta}) = -\partial^2\ell/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'$ is easily seen to be

$$I(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \left\{ \frac{\sum_{\ell=1}^{n} Y_\ell(t_i) e^{\boldsymbol{\beta}'\mathbf{x}_\ell} [\mathbf{x}_\ell - \bar{\mathbf{x}}(t_i, \boldsymbol{\beta})][\mathbf{x}_\ell - \bar{\mathbf{x}}(t_i, \boldsymbol{\beta})]'}{\sum_{\ell=1}^{n} Y_\ell(t_i) e^{\boldsymbol{\beta}'\mathbf{x}_\ell}} \right\}. \tag{7.1.9}$$

The maximum likelihood equations $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$ are easily solved by Newton–Raphson iteration (see Appendix D) or other methods. Numerous software packages give the maximum likelihood estimate (m.l.e.) $\hat{\boldsymbol{\beta}}$ and standard errors, tests, or confidence intervals based on the standard asymptotic normal approximation

$$\hat{\boldsymbol{\beta}} \simeq N_p(\boldsymbol{\beta}, I(\hat{\boldsymbol{\beta}})^{-1}).$$

Inferences can also be based on likelihood ratio statistics such as $\Lambda(\boldsymbol{\beta}) = 2\ell(\hat{\boldsymbol{\beta}}) - 2\ell(\boldsymbol{\beta})$, or on the score statistic, just as described for ordinary likelihoods in Appendix C. Score procedures lead to especially simple tests for comparing distributions. We consider this topic next since it also provides a simple illustration of inference under the PH model. The validity and properties of $L(\boldsymbol{\beta})$ will then be considered in Section 7.1.3.

7.1.2 Comparison of Two or More Lifetime Distributions

Regression models can be used for comparing or testing the equality of distributions. The PH model gives distribution-free tests for the equality of distributions that are effective at detecting differences when the distributions in question have roughly proportional hazard functions. Consider to start two distributions, and a test of equality of their survivor functions,

$$H : S_1(t) = S_2(t).$$

This can be approached through a regression model that includes an indicator covariate x taking on the value 1 or 0 according to whether an individual is associated with the first or second distribution:

$$x = I \text{ (individual's lifetime comes from } S_1(t)).$$

One then tests H by testing for the absence of a covariate effect.

If the PH regression model (7.1.2) with $\boldsymbol{\beta}'\mathbf{x} = \beta x$ is adopted, then $S_2(t) = S_0(t)$ and $S_1(t) = S_0(t)^{\exp(\beta)}$, so that

$$S_1(t) = S_2(t)^{\exp(\beta)}. \tag{7.1.10}$$

A test of $\beta = 0$ is a test of the hypothesis H; the alternative hypotheses (7.1.10) with $\beta \neq 0$ are sometimes referred to as the Lehmann family of alternatives.

The hypothesis $H: \beta = 0$ can be tested by fitting the PH model with survivor function

$$S(t|x) = S_0(t)^{\exp(\beta x)} \tag{7.1.11}$$

and applying standard large-sample procedures based on the m.l.e. $\hat{\beta}$ and its standard error, or on the likelihood ratio statistic $\Lambda(\beta)$; these approaches are illustrated later in Example 7.1.1. However, the score function $U(\beta)$ leads to an especially simple test that does not require $\hat{\beta}$ to be obtained, as we now describe.

Assume that we have independent censored random samples of lifetimes from $S_1(t)$ and $S_2(t)$ of sizes N_1 and N_2. For the model (7.1.11) the score function (7.1.8) is

$$U(\beta) = \sum_{i=1}^{N} \left(d_{1i} - \frac{d_i n_{1i} e^{\beta}}{n_{1i} e^{\beta} + n_{2i}} \right), \tag{7.1.12}$$

where $N = N_1 + N_2$, and we introduce the notation

$$d_i = \delta_i = I(t_i \text{ is a lifetime})$$

$$d_{1i} = \delta_i x_i = I(t_i \text{ is a lifetime from } S_1(t))$$

$$n_{1i} = \sum_{\ell=1}^{N} Y_\ell(t_i) x_\ell = \text{ number at risk from } S_1(t) \text{ at time } t_i$$

$$n_{2i} = \sum_{\ell=1}^{N} Y_\ell(t_i)(1 - x_\ell) = \text{ number at risk from } S_2(t) \text{ at time } t_i.$$

The information (7.1.9) is correspondingly

$$I(\beta) = \sum_{i=1}^{N} \frac{d_i n_{1i} n_{2i} e^{\beta}}{(n_{1i} e^{\beta} + n_{2i})^2}. \tag{7.1.13}$$

Confidence intervals or tests for β can be obtained from the approximate pivotal quantity

$$Z(\beta) = U(\beta)/I(\beta)^{1/2}. \tag{7.1.14}$$

As described in Appendix C, this has an asymptotic standard normal distribution when β is the true value of the parameter.

The approximate pivotal $Z(0)$ provides a very simple test of the hypothesis $H: \beta = 0$, and hence of $S_1(t) = S_2(t)$. We have from (7.1.12) and (7.1.13) that

$$U(0) = \sum_{i=1}^{N} (d_{1i} - \frac{d_i n_{1i}}{n_i}), \qquad I(0) = \sum_{i=1}^{N} \frac{d_i n_{1i} n_{2i}}{n_i^2}, \tag{7.1.15}$$

where $n_i = n_{1i} + n_{2i}$. The test statistic is $Z = U(0)/I(0)^{1/2}$, and evidence against $S_1(t) = S_2(t)$ is provided by large values of $|Z|$ or Z^2.

Several points can be made about this test. First, we can think of the terms in $U(0)$ as being of the form "observed number minus expected number of failures at t_i that are from $S_1(t)$." To see this note that only times t_i at which a failure occurs (i.e., $d_i = 1$) contribute to $U(0)$ and $I(0)$, and that if $S_1(t) = S_2(t)$, then the conditional expectation of d_{1i}, given $d_i = 1$ and the numbers n_{1i}, n_{2i} at risk, is $d_i n_{1i}/n_i$. This shows directly that $E[U(0)] = 0$ under $H: \beta = 0$. A second point is that if ties can occur among the observed lifetimes, then (7.1.12)–(7.1.15) still apply. Finally, equivalent expressions are given by redefining the t_i as the distinct times at which failures occur across all n individuals, and d_i and d_{1i} as the total number of failures at t_i, and the number of these failures from $S_1(t)$, respectively.

Tied failure times are impossible under continuous lifetime distributions, but occur frequently in data sets because actual measurements are discrete. When there are substantial numbers of ties, it may be preferable to switch to a discrete model. This is discussed in Section 7.3, where a model is given that leads to the score statistic $U(0)$ in (7.1.15) for testing $S_1(t) = S_2(t)$, but with the information $I(0)$ replaced by

$$I_1(0) = \sum_{i=1}^{k} \frac{d_i(n_i - d_i) n_{1i} n_{2i}}{n_i^2 (n_i - 1)}. \tag{7.1.16}$$

In (7.1.16), i indexes the k distinct times at which failures occur, and d_i is the total number of failures at the ith time $t_{(i)}$.

There is a close connection between the two-sample test given here and two-sample rank tests discussed in Section 8.1. In particular, the test here is a generalization to censored data of a rank test proposed by Savage (1956), and because of this, it is often referred to as the log rank test. Many software packages include this test.

Example 7.1.1. The data below show remission times, in weeks, for 40 leukemia patients randomly assigned to two treatments A and B. Asterisks denote censoring times.

Treatment A	1, 3, 3, 6, 7, 7, 10, 12, 14, 15, 18, 19, 22, 26, 28*, 29, 34, 40, 48*, 49*
Treatment B	1, 1, 2, 2, 3, 4, 5, 8, 8, 9, 11, 12, 14, 16, 18, 21, 27*, 31, 38*, 44

Plots of Kaplan–Meier or Nelson–Aalen estimates for the two groups suggest that a proportionality assumption is reasonable for the two hazard functions. Let us therefore use (7.1.15) to test that the remission time distributions are the same for patients on the two treatments.

The statistics $U(0)$ and $I(0)$ in (7.1.15) are easily computed from the preceding raw data, or are available from software packages. We find $U(0) = -3.323$ and $I(0) = 8.409$, giving the test statistic $Z(0)^2 = U(0)^2/I(0) = 1.31$. The approximate $\chi^2_{(1)}$ distribution of $Z(0)^2$ under the hypothesis gives the p-value .25, so there is no evidence of a difference in distributions.

There are several ties in the data, and one might use (7.1.16) in place of $I(0)$; this gives 8.196 and $Z(0)^2 = 1.35$, very close to the previous value. The test could also be carried out by fitting the PH model (7.1.10) and using either the Wald statistic $Z^2 = \hat{\beta}^2/se(\hat{\beta})^2$ or the likelihood ratio statistic $\Lambda = 2\ell(\hat{\beta}) - 2\ell(0)$. We find here that $\hat{\beta} = -.388$, $se(\hat{\beta}) = I(\hat{\beta})^{-1/2} = .341$, $\ell(\hat{\beta}) = -103.30$, and $\ell(0) = -103.95$. These give $Z^2 = 1.30$ and $\Lambda = 1.30$. All four tests give essentially identical results.

Tests of the equality of three or more lifetime distributions are also readily obtained. To compare m distributions $S_1(t), \ldots, S_m(t)$ we define a vector of $m-1$ indicator covariates, $\mathbf{x} = (x_1, \ldots, x_{m-1})'$, where

$$x_r = I \text{ (individual's lifetime is from } S_r(t)), \qquad r = 1, \ldots, m-1.$$

If a PH model (7.1.2) is assumed, then $S_m(t) = S_0(t)$ and

$$S_r(t) = S_0(t)^{\exp(\beta_r)}, \qquad r = 1, \ldots, m-1. \tag{7.1.17}$$

The hypothesis $H: S_1(t) = \cdots = S_m(t)$ is equivalent to $H: \boldsymbol{\beta} = \mathbf{0}$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{m-1})$; a test of $\boldsymbol{\beta} = \mathbf{0}$ will be effective at detecting differences among survivor functions that are roughly of the form (7.1.17).

A score test of $\boldsymbol{\beta} = \mathbf{0}$ is very simple, as in the two-sample case. Assume that there are independent censored random samples of lifetimes from $S_1(t), \ldots, S_m(t)$ of sizes $N_1, \ldots, N_m$. Let $N = N_1 + \cdots + N_m$ be the size of the combined sample, which has observations $(t_i, \delta_i, \mathbf{x}_i)$, with $\mathbf{x}_i = (x_{i1}, \ldots, x_{i,m-1})'$, and

$$x_{ir} = I \text{ (individual } i\text{'s lifetime is from } S_r(t)).$$

Define for $r = 1, \ldots, m-1$

$$d_{ri} = \delta_i x_{ir} = I(t_i \text{ is a lifetime from } S_r(t))$$

$$n_{ri} = \sum_{\ell=1}^{N} Y_\ell(t_i) x_{\ell r} = \text{ number at risk from } S_r(t) \text{ at time } t_i,$$

and similarly define d_{mi} and n_{mi} as the numbers of deaths and individuals at risk from distribution $S_m(t)$. Also let $d_i = \sum d_{ri}$, $n_i = \sum n_{ri}$. The elements in the score vector (7.1.8) and information matrix (7.1.9) when $\boldsymbol{\beta} = \mathbf{0}$ can then be written as

$$U_r(\mathbf{0}) = \sum_{i=1}^{N} \left(d_{ri} - \frac{d_i n_{ri}}{n_i} \right) \qquad r = 1, \ldots, m-1 \tag{7.1.18}$$

$$I_{rs}(\mathbf{0}) = \sum_{i=1}^{N} \frac{d_i n_{ri}}{n_i} \left(\delta_{rs} - \frac{n_{si}}{n_i} \right) \qquad r, s = 1, \ldots, m-1, \tag{7.1.19}$$

where $\delta_{rs} = I(r = s)$.

Under the hypothesis $H: \boldsymbol{\beta} = \mathbf{0}$, $\mathbf{U}(\mathbf{0}) = [U_1(\mathbf{0}), \ldots, U_{m-1}(\mathbf{0})]'$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $I(\mathbf{0})$, by the results in Appendix C. A test of H can be based on the statistic

$$W = \mathbf{U}(\mathbf{0})' I(\mathbf{0})^{-1} \mathbf{U}(\mathbf{0}). \tag{7.1.20}$$

Large values of W provide evidence against the equality of the m distributions; under H the distribution of W is approximately $\chi^2_{(m-1)}$ for large samples, and p-values can be based on this.

When $m = 2$, this test is the same as that based on (7.1.15) and $Z^2 = U(0)^2/I(0)$. The comments following (7.1.15) apply here as well. In particular, the test can be used when there are substantial numbers of ties in the data, but it is advisable to replace $I_{rs}(\mathbf{0})$ in (7.1.19) with

$$\sum_{i=1}^{N} \frac{d_i(n_i - d_i)n_{ri}}{n_i(n_i - 1)} \left(\delta_{rs} - \frac{n_{si}}{n_i} \right), \qquad r, s = 1, \ldots, m-1. \tag{7.1.21}$$

An advantage of tests of distributional equality based on regression models is that the regression coefficients β_r provide measures of the differences among distribu-

tions. Under (7.1.17)

$$e^{\beta_r} = \frac{H_r(t)}{H_0(t)} = \frac{h_r(t)}{h_0(t)}, \qquad r = 1, \ldots, m-1 \tag{7.1.22}$$

is the ratio of the hazard or cumulative hazard functions for the distribution $S_r(t)$ and the baseline distribution $S_0(t) = S_m(t)$. These measures and the test based on (7.1.20) are, of course, useful only in so far as the model (7.1.17) on which they are based is satisfactory. This can be checked by using probability plots based on Kaplan–Meier or Nelson–Aalen estimates for the m distributions, as described in Section 6.2.1. Other tests of distributional equality based on a more general multiplicative hazards form are considered in Section 7.1.8.

7.1.3 Justification and Properties of the Likelihood Function $L(\boldsymbol{\beta})$

It was noted at the start of Section 7.1 that $L(\boldsymbol{\beta})$ in (7.1.3) is not in general an ordinary likelihood function, but that under quite mild conditions it can be treated as such for inferences about $\boldsymbol{\beta}$. In particular, the estimate $\hat{\boldsymbol{\beta}}$ obtained by maximizing $L(\boldsymbol{\beta})$ is consistent and in large samples can be treated as approximately p-variate normal with mean vector $\boldsymbol{\beta}$ and covariate matrix $I(\hat{\boldsymbol{\beta}})^{-1}$. Concomitantly, likelihood ratio statistics $\Lambda(\boldsymbol{\beta})$ based on $L(\boldsymbol{\beta})$ can be treated as approximately chi-squared in large samples, and score statistics $\mathbf{U}(\boldsymbol{\beta})$ as approximately normal with mean vector $\mathbf{0}$ and covariance matrix $I(\boldsymbol{\beta})$. In this section we outline some frameworks within which $L(\boldsymbol{\beta})$ and its properties can be studied.

7.1.3.1 $L(\boldsymbol{\beta})$ as a Marginal Likelihood

When there is no censoring, (7.1.3) can be derived as a marginal likelihood function based on the rank statistic for the data. Specifically, suppose lifetimes $t_1, \ldots, t_n$ of n individuals with covariates $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are observed. The probability density function (p.d.f.) of T_i given $\mathbf{x}_i$ under the PH model (7.1.1) is

$$f(t_i|\mathbf{x}_i) = h_0(t)e^{\boldsymbol{\beta}'\mathbf{x}_i} \exp[-H_0(t)e^{\boldsymbol{\beta}'\mathbf{x}_i}], \tag{7.1.23}$$

where $H_0(t)$ is the baseline cumulative hazard function. Let $\mathbf{r} = [(1), \ldots, (n)]$ denote the rank statistic for the data; that is, (i) is the label of the individual with the ith smallest lifetime. The possibility of ties is ignored, since they have probability 0 under the continuous model. The distribution of $\mathbf{r}$ is discrete, with $n!$ possible rank vectors. Its probability function is found as

$$Pr\{\mathbf{r} = [(1), \ldots, (n)]\} = Pr[T_{(1)} < T_{(2)} < \cdots < T_{(n)}] = \int_0^\infty \int_{t_{(1)}}^\infty \cdots \int_{t_{(n-1)}}^\infty f(t_{(1)}|\mathbf{x}_{(1)}) \cdots f(t_{(n)}|\mathbf{x}_{(n)})\, dt_{(n)} \cdots dt_{(1)}.$$

Straightforward integration of this expression using (7.1.23) yields

$$Pr\{\mathbf{r} = [(1) \cdots (n)]\} = \prod_{i=1}^{N} \left(e^{\boldsymbol{\beta}'\mathbf{x}_{(i)}} \Big/ \sum_{\ell \in R(t_{(i)})} e^{\boldsymbol{\beta}'\mathbf{x}_\ell} \right),$$

which is the likelihood function (7.1.3). In obtaining this result, we have used the fact that $R(t_{(i)}) = [(i), (i+1), \ldots, (n)]$, since there is no censoring.

In the simple noncensored case, therefore, (7.1.3) is a legitimate likelihood function arising from the probability distribution of the rank statistic. Under suitable assumptions concerning the $\mathbf{x}_i$ such as those given in Hajek and Sidak (1967), $L(\boldsymbol{\beta})$ behaves in the usual way, with $\hat{\boldsymbol{\beta}}$ being asymptotically normally distributed with mean $\boldsymbol{\beta}$ and covariance matrix $\mathcal{I}^{-1}$, where $\mathcal{I}$ has entries $\mathcal{I}_{rs} = E(-\partial^2 \log L/\partial\beta_r\, \partial\beta_s)$. $\mathcal{I}$ is consistently estimated by $\mathbf{I}(\hat{\boldsymbol{\beta}})$ as given by (7.1.9). It can be noted that $L(\boldsymbol{\beta})$ is a marginal likelihood in the sense of Fraser (1968) or Kalbfleisch and Sprott (1970).

If the data are subject to Type 2 censoring, an extension of the preceding argument shows that $L(\boldsymbol{\beta})$ is once again a legitimate marginal likelihood function based on a rank statistic. For more general types of censoring the argument breaks down, however. In general the rank statistic is in fact unknown, because censoring makes it impossible to know the exact ordering of the actual lifetimes.

7.1.3.2 $L(\boldsymbol{\beta})$ as a Partial Likelihood

Cox (1975) introduced the concept of partial likelihood and used it to obtain (7.1.3) and study its properties. Partial likelihoods are described in Appendix C. They are based on factoring a likelihood function via the multiplication rule for probabilities, and then discarding certain pieces that involve nuisance parameters. To develop (7.1.3) this way we return to the notation used in Sections 2.2.2 and 3.2.4, which describes the dynamic evolution of lifetime data.

As in Section 2.2.2, discretize time into short intervals $[t, t+\Delta t)$ and consider a random sample of n individuals with discrete hazard functions $h(t|\mathbf{x}_i)$, $i = 1, \ldots, n$. Let (t_i, δ_i), $i = 1, \ldots, n$ denote the observed times and status indicators. Define

$$dN_i(t) = I(T_i \in [t, t+\Delta t), \delta_i = 1)$$
$$Y_i(t) = I(t_i \geq t)$$
$$dN_\cdot(t) = \sum_{i=1}^{n} dN_i(t),$$

and let $d\mathbf{N}(t) = [dN_1(t), \ldots, dN_n(t)]$. Denote the history of failure and censoring that has occurred over $(0, t)$ as $\mathcal{H}(t)$. Following the arguments in Section 2.2.2 and assuming that the censoring process follows the independence requirements there, we have, analogous to (2.2.9), that

$$L = \prod_{t=0}^{\infty} Pr[d\mathbf{N}(t)|\mathcal{H}(t)] \tag{7.1.24}$$

provides a partial likelihood for estimation of the parameters specifying $h(t|\mathbf{x})$. We think of (7.1.24) as a discrete product across all times t defining the intervals $[t, t+\Delta t)$ that partition the time axis.

Suppose now that the hazard function $h(t|\mathbf{x})$ is of the PH form (7.1.1), and factor the terms in (7.1.24) as

$$Pr[d\mathbf{N}(t)|\,dN_\cdot(t), \mathcal{H}(t)]\,Pr[dN_\cdot(t)|\mathcal{H}(t)]. \tag{7.1.25}$$

Keeping only the first of these two probabilities yields another partial likelihood that to order Δt depends only on $\boldsymbol{\beta}$ and not on $h_0(t)$. To see this note that for Δt sufficiently small any interval contains either 0 or 1 failure with probability approaching 1. If $dN_\cdot(t) = 0$, then the first term in (7.1.25) is null, but if $dN_\cdot(t) = 1$, then $dN_i(t) = 1$ for some individual i. This gives the conditional probability

$$\begin{aligned} Pr[dN_i(t) = 1|dN_\cdot(t) = 1, \mathcal{H}(t)] &= \frac{Y_i(t)h(t|\mathbf{x}_i)\Delta t}{\sum_{\ell=1}^n Y_\ell(t)h_\ell(t|\mathbf{x}_\ell)\Delta t} + 0(\Delta t) \\ &= \frac{e^{\boldsymbol{\beta}'\mathbf{x}_i}Y_i(t)}{\sum_{\ell=1}^n Y_\ell(t)e^{\boldsymbol{\beta}'\mathbf{x}_\ell}}. \end{aligned}$$

The partial likelihood

$$L = \prod_{t=0}^{\infty} Pr[d\mathbf{N}(t)|\,dN_\cdot(t), \mathcal{H}(t)] \tag{7.1.26}$$

therefore gives exactly (7.1.5) in the limit as the Δt's approach 0.

7.1.3.3 $L(\boldsymbol{\beta})$ as a Profile Likelihood

The full likelihood function based on (7.1.24) is the standard censored data likelihood function (2.2.14). Under the PH model with hazard and survivor functions (7.1.1) and (7.1.2), this becomes

$$L_1(\boldsymbol{\beta}, H_0) = \prod_{i=1}^n [h_0(t_i)e^{\boldsymbol{\beta}'\mathbf{x}_i}]^{\delta_i} S_0(t_i)^{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}. \tag{7.1.27}$$

The H_0 in $L_1(\boldsymbol{\beta}, H_0)$ stands for the unknown baseline cumulative hazard function $H_0(t)$. We write $L_1(\boldsymbol{\beta}, H_0)$ rather than $L_1(\boldsymbol{\beta}, h_0)$ because it is simpler to work with H_0 in a maximum likelihood approach. Nonparametric maximum likelihood estimation of a cumulative hazard or survivor function was discussed in Section 3.2. This can be extended to deal with the semiparametric PH model, in which case we seek to maximize $L_1(\boldsymbol{\beta}, H_0)$ jointly for $\boldsymbol{\beta}$ and H_0. This is considered in Section 7.4, where it is shown that the likelihood (7.1.3) equals a profile likelihood function $L(\boldsymbol{\beta}, \tilde{H}_0(\boldsymbol{\beta}))$, where $\tilde{H}_0(\boldsymbol{\beta})$ is a nonparametric m.l.e. of H_0 with $\boldsymbol{\beta}$ given.

7.1.3.4 Properties of $L(\boldsymbol{\beta})$

Properties of (7.1.3), the estimator $\hat{\boldsymbol{\beta}}$ that maximizes it, and associated quantities can be studied using results from martingale theory (see Appendix F). With the counting process notation $N_i(t) = I(t_i \le t, \delta_i = 1)$, the score vector (7.1.8) can be rewritten as

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\infty [\mathbf{x}_i - \bar{\mathbf{x}}(t, \boldsymbol{\beta})]\,dN_i(t). \tag{7.1.28}$$

If we define

$$dM_i(t) = dN_i(t) - Y_i(t)e^{\boldsymbol{\beta}'\mathbf{x}_i}\,dH_0(t),$$

where $dH_0(t) = h_0(t)\,dt$, then we also find that

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\infty [\mathbf{x}_i - \bar{\mathbf{x}}(t, \boldsymbol{\beta})]\,dM_i(t). \tag{7.1.29}$$

To see this latter result, note that

$$\begin{aligned} &\sum_{i=1}^n \int_0^\infty [\mathbf{x}_i - \bar{\mathbf{x}}(t, \boldsymbol{\beta})]Y_i(t)e^{\boldsymbol{\beta}'\mathbf{x}_i}\,dH_0(t) \\ &= \int_0^\infty \left\{ \sum_{i=1}^n Y_i(t)\mathbf{x}_i e^{\boldsymbol{\beta}'\mathbf{x}_i} - \bar{\mathbf{x}}(t, \boldsymbol{\beta}) \sum_{i=1}^n Y_i(t)e^{\boldsymbol{\beta}'\mathbf{x}_i} \right\} dH_0(t), \end{aligned}$$

and by (7.1.7) the term inside the curly brackets is zero. Therefore, although (7.1.29) appears to involve $H_0(t)$, it does not. Note also that the equality of (7.1.28) and (7.1.29) does not extend to the individual terms for $i = 1, \ldots, n$. Moreover, the ith term in (7.1.29) has expectation $\mathbf{0}$, but that in (7.1.28) does not.

The terms for the different individuals $i = 1, \ldots, n$ in (7.1.29) are not independent, but the $M_i(t)$'s are mean 0 martingales, and the terms in square brackets are predictable processes. As indicated in Appendix F, it therefore follows immediately that $E[\mathbf{U}(\boldsymbol{\beta})] = \mathbf{0}$, and it is possible to use martingale central limit theory to show that $n^{-1/2}\mathbf{U}(\boldsymbol{\beta})$ is asymptotically p-variate normal. The covariance matrix for $\mathbf{U}(\boldsymbol{\beta})$ is given in Problem 7.2, and $I(\hat{\boldsymbol{\beta}})$ from (7.1.9) is shown to estimate it. The estimator $\hat{\boldsymbol{\beta}}$ can be shown to be consistent and asymptotically normal, and the asymptotic covariance matrices of $n^{-1/2}\mathbf{U}(\boldsymbol{\beta})$ and $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ are estimated consistently by $I(\hat{\boldsymbol{\beta}})/n$ and $nI(\hat{\boldsymbol{\beta}})^{-1}$, respectively. Finally, the likelihood ratio statistic $\Lambda(\boldsymbol{\beta}) = 2\ell(\hat{\boldsymbol{\beta}}) - 2\ell(\boldsymbol{\beta})$ is asymptotically $\chi^2_{(p)}$ when $\boldsymbol{\beta}$ is the true value of the parameter vector.

A rigorous development of this area is provied in the books by Andersen et al. (1993) and Fleming and Harrington (1991). Further references are given in the Bibliographic Notes at the end of the chapter.

7.1.4 Adjustments for Tied Lifetimes

In the preceding discussion it was assumed that if ties occur among the lifetimes, then we merely continue to use the likelihood function (7.1.5). If there is a substan-

tial number of ties, the discrete nature of the lifetimes should be considered. Ties can occur because the underlying lifetimes are discrete or because of rounding or grouping in continuous data. In either case, models that recognize the discrete measurements can be used; this topic is discussed in Section 7.3. If there are only a few ties, it is convenient to retain the continuous-time model and use a simple adjustment to the likelihood (7.1.5), because the discrete data likelihoods lead to more complex procedures; we consider this now.

Peto (1972) suggested that if $t_{(1)} < \cdots < t_{(k)}$ are the distinct lifetimes, and if d deaths are observed at some time $t_{(j)}$, then the times of these deaths are in reality distinct. Thus instead of the d terms in (7.1.5) corresponding to the deaths at $t_{(j)}$, we should have

$$\left(e^{\beta' \mathbf{x}_{\ell_1}} \Big/ \sum_{\ell \in R_j} e^{\beta' \mathbf{x}_\ell}\right) \times \left(e^{\beta' \mathbf{x}_{\ell_2}} \Big/ \sum_{\ell \in R_j - \{\ell_1\}} e^{\beta' \mathbf{x}_\ell}\right) \cdots \times \left(e^{\beta' \mathbf{x}_{\ell_d}} \Big/ \sum_{\ell \in R_j - \{\ell_1, \ldots, \ell_{d-1}\}} e^{\beta' \mathbf{x}_\ell}\right), \tag{7.1.30}$$

reflecting the fact that the individuals who died at $t_{(j)}$ did so in some order: first ℓ_1, then ℓ_2, and so on. We do not know the correct order, however, so cannot use (7.1.30). Peto and others suggest that it be replaced with its average across the $d!$ permutations of the labels for the individuals who died at $t_{(j)}$. This is computationally forbidding when a data set has many ties, and a simpler adjustment suggested by Efron (1977) is to weight the terms in the denominators of (7.1.5) as follows: if d deaths occur at the time $t_{(j)}$, then let $\ell_1, \ldots, \ell_d$ be the individuals who died at $t_{(j)}$, and replace the product of the d denominators for those individuals in (7.1.5) with

$$\prod_{r=0}^{d-1} \left\{ \sum_{\ell \in R(t_{(j)}) - \{\ell_1, \ldots, \ell_d\}} e^{\beta' \mathbf{x}_\ell} + \frac{d-r}{d} \sum_{i=1}^{d} e^{\beta' \mathbf{x}_{\ell_i}} \right\}. \tag{7.1.31}$$

The motivation for this is that each of the d individuals is in the risk set $R(t_{(j)})$, has probability $(d-1)/d$ of being in the next risk set with the individual dying first dropped, and so on.

If there are relatively few ties, then the use of (7.1.31) and of (7.1.5) as it stands give close to the same results, as does the ad hoc approach whereby ties are broken by randomly adding small values to certain lifetimes. Our preference is for (7.1.5), but some software packages use (7.1.31) as the default procedure. If the number of ties is large enough that the two approaches give substantially different estimates, then a discrete model as in Section 7.3 can be considered.

7.1.5 Estimation of $H_0(t)$ or $S_0(t)$

Estimates of $H_0(t)$ and $S_0(t)$ are generally wanted. For example, an estimate of $S_0(t)$ is needed to estimate $S(t|\mathbf{x})$ or quantiles for T via (7.1.2). In addition, nonparametric

estimates $\hat{H}_0(t)$ or $\hat{S}_0(t)$ can be examined with a view to parametric modeling. A simple intuitive estimate of $H_0(t)$ is

$$\hat{H}_0(t) = \sum_{i:\, t_i \le t} \left\{ \frac{\delta_i}{\sum_{\ell=1}^{n} Y_\ell(t_i) e^{\hat{\beta}' \mathbf{x}_\ell}} \right\}. \tag{7.1.32}$$

This is motivated by the fact that

$$E\left\{ \sum_{\ell=1}^{n} [dN_\ell(t) - Y_\ell(t) e^{\beta' \mathbf{x}_\ell} \, dH_0(t) | \mathcal{H}(t)] \right\} = 0,$$

suggesting the conditional moment estimate, with $\hat{\boldsymbol{\beta}}$ replacing $\boldsymbol{\beta}$:

$$d\hat{H}_0(t) = \sum_{\ell=1}^{n} dN_\ell(t) \Big/ \left[\sum_{\ell=1}^{n} Y_\ell(t) e^{\hat{\beta}' \mathbf{x}_\ell} \right]. \tag{7.1.33}$$

The estimate (7.1.32) equals $\int_0^t d\hat{H}_0(t)$. This is often referred to as the Breslow or generalized Nelson–Aalen estimate; note that when $\hat{\boldsymbol{\beta}} = \mathbf{0}$ (7.1.32) is just the Nelson–Aalen estimate (3.2.13).

A simple way to estimate $S_0(t)$ is to exploit the relationship $S_0(t) = \exp[-H_0(t)]$ and define

$$\hat{S}_0(t) = \exp[-\hat{H}_0(t-)]. \tag{7.1.34}$$

When there are no covariates, or $\hat{\boldsymbol{\beta}} = \mathbf{0}$, this does not give the Kaplan–Meier estimate (3.2.2), but another estimate, sometimes referred to as the Fleming–Harrington estimate. Both $\hat{H}_0(t)$ and $\hat{S}_0(t)$ are given by standard software for the PH model. Estimates

$$\hat{S}(t|\mathbf{x}) = \hat{S}_0(t)^{\exp(\hat{\beta}' \mathbf{x})} \tag{7.1.35}$$

for a specified vector $\mathbf{x}$ and value of t can also be obtained.

Standard errors for $\hat{S}(t|\mathbf{x})$ require the joint asymptotic distribution of $\hat{H}_0(t)$ and $\hat{\boldsymbol{\beta}}$. This can be obtained via martingale arguments that extend those referred to at the end of Section 7.1.3 (e.g., Andersen et al. 1993, pp. 503–506). We note a pair of useful results, recalling the definition of $\bar{x}(t, \beta)$ in (7.1.7) and defining

$$S^{(0)}(t, \boldsymbol{\beta}) = \sum_{i=1}^{n} Y_i(t) e^{\beta' \mathbf{x}_i}. \tag{7.1.36}$$

Then, $\sqrt{n}[\hat{H}_0(t) - H(t)]$ has a limiting normal distribution with mean 0 and variance estimated by n times

$$\widehat{\mathrm{Var}}[\hat{H}_0(t)] = \sum_{i:\, t_i \le t} \frac{\delta_i}{S^{(0)}(t_i, \hat{\boldsymbol{\beta}})^2} + \hat{\mathbf{W}}(t)' I(\hat{\boldsymbol{\beta}})^{-1} \hat{\mathbf{W}}(t), \tag{7.1.37}$$

where

$$\hat{\mathbf{W}}(t) = \sum_{i:t_i \le t} \frac{\delta_i \bar{\mathbf{x}}(t_i, \hat{\boldsymbol{\beta}})}{S^{(0)}(t_i, \hat{\boldsymbol{\beta}})} = \int_0^t \bar{\mathbf{x}}(u, \hat{\boldsymbol{\beta}})\, d\hat{H}_0(u). \tag{7.1.38}$$

In addition,

$$\widehat{\mathrm{Var}}\{\log[-\log \hat{S}(t|\mathbf{x})]\} = \hat{H}_0(t)^{-2}\left\{ \sum_{i:t_i \le t} \frac{\delta_i}{S^{(0)}(t_i, \hat{\boldsymbol{\beta}})^2} + \left[\hat{\mathbf{W}}(t) - \hat{H}_0(t)\mathbf{x}\right]' I(\hat{\boldsymbol{\beta}})^{-1}\left[\hat{\mathbf{W}}(t) - \hat{H}_0(t)\mathbf{x}\right]\right\} \tag{7.1.39}$$

estimates the asymptotic variance of $\log[-\log \hat{S}(t|\mathbf{x})]$. We give the form (7.1.39), since to obtain confidence intervals for $S(t|\mathbf{x})$, it is preferable to treat $\log[-\log \hat{S}(t|\mathbf{x})]$ as approximately normal rather than $\hat{S}(t|\mathbf{x})$. Note that the first terms in (7.1.37) and (7.1.39) correspond to the variance if $\boldsymbol{\beta}$ were known; the second terms reflect the fact that $\boldsymbol{\beta}$ is estimated by $\hat{\boldsymbol{\beta}}$.

7.1.6 Stratification

Sometimes the basic PH model is not adequate, but a model in which covariates affect the hazard function multiplicatively within certain strata may be. In particular, suppose that individuals can be assigned to one of J strata, defined in terms of one or more factors. Suppose also that for an individual in stratum j with covariate vector $\mathbf{x}$ the hazard function is

$$h_j(t|\mathbf{x}) = h_{0j}(t)e^{\boldsymbol{\beta}'\mathbf{x}} \qquad j = 1, \ldots, J, \tag{7.1.40}$$

where $h_{0j}(t)$ is a baseline hazard function. That is, individuals in the same stratum have proportional hazard functions, but those in different strata do not. In (7.1.40) it is assumed there is no stratum–covariate interaction; in some cases we will want to consider the model

$$h_j(t|\mathbf{x}) = h_{0j}(t)e^{\boldsymbol{\beta}_j'\mathbf{x}} \qquad j = 1, \ldots, J, \tag{7.1.41}$$

in which the covariate effects vary from stratum to stratum.

The model (7.1.41) can of course be handled by fitting separate PH models for each stratum. The simpler model (7.1.40) can be handled by obtaining a partial likelihood $L_j(\boldsymbol{\beta})$ of the form (7.1.5) for each stratum, then basing estimation on the combined likelihood function $L(\boldsymbol{\beta}) = L_1(\boldsymbol{\beta})L_2(\boldsymbol{\beta}), \ldots, L_J(\boldsymbol{\beta})$. Expressions for the log-likelihood, score vector, and information matrix in Section 7.1.1 merely require summation across strata. Standard PH model software deals with this model.

Once $\boldsymbol{\beta}$ in (7.1.40) is estimated, the baseline cumulative hazard functions can be estimated by (7.1.32), using for $\hat{H}_{0j}(t)$ only those individuals who are in stratum j.

7.1.7 Left Truncation and Delayed Entry

Left truncation occurs when it is a condition on the observation of an individual i that their lifetime exceeds some value, u_i. This phenomenon was discussed in Section 2.4, and it was subsequently shown in Sections 3.2 and 3.5 that nonparametric estimation of survivor or cumulative hazard functions easily accommodates left truncation. The same is true for the semiparametric analysis of multiplicative hazards models such as (7.1.1).

The key, as in Section 3.2.1, is to redefine $Y_i(t)$ when left truncation is present as

$$Y_i(t) = I(u_i \le t \le t_i), \qquad i = 1, \ldots, n,$$

where the observation on individual i now consists of $(u_i, t_i, \delta_i, \mathbf{x}_i)$, with t_i either a lifetime or right-censoring time, and $\delta_i = I(t_i \text{ is a lifetime})$. Provided that the left-truncation mechanism is independent in the sense described in Section 2.4.1, it then can be shown that the Cox likelihood (7.1.5) still applies exactly as given, as does the estimate (7.1.32) of $H_0(t)$ and all of the procedures described in preceding sections. An easy way to see that this is so is to consider the partial likelihood derivation in Section 7.1.3, with $Y_i(t)$ redefined as earlier, $dN_i(t) = I(T_i \in [t, t + \Delta t), T_i \ge u_i, \delta_i = 1)$, and $\mathcal{H}(t)$ in (7.1.24) defined so it includes the history of entry times u_i, as well as past failures and censoring. Then, (7.1.25) still holds and the partial likelihood (7.1.26) once again gives (7.1.5).

7.1.8 Time-Varying Covariates

The methods of analysis described here are mostly unchanged when the model (7.1.1) is extended to allow the covariates to be time-varying. For example, (7.1.5) is replaced by

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{e^{\boldsymbol{\beta}'\mathbf{x}_i(t_i)}}{\sum_{\ell=1}^{n} Y_\ell(t_i) e^{\boldsymbol{\beta}'\mathbf{x}_\ell(t_i)}} \right)^{\delta_i}, \tag{7.1.42}$$

and similar adjustments are made to the expressions for the score vector (7.1.8) and information matrix (7.1.9). It should be noted that (7.1.42) and procedures based on it require the values of time-varying covariates at the observed lifetime t_i, not just for individual i, but for all individuals at risk at t_i, that is, with $Y_\ell(t_i) = 1$. This may sometimes pose a problem; for example, in settings where covariates vary randomly over time, it may be possible to record covariate values only intermittently. In addition, data storage and computational demands in the evaluation of likelihood, score vector, and information matrix values are greater with time-varying covariates. Some software packages accept time-varying covariates only if they are piecewise-

constant; that is, for any individual i there is a set of times $0 = a_{i0} < a_{i1} < a_{i2} < \cdots$, such that

$$\mathbf{x}_i(t) = \mathbf{x}_{ij}, \qquad a_{i,j-1} \le t < a_{ij} \tag{7.1.43}$$

for $j = 1, 2, \ldots$.

Time-varying covariates were discussed briefly in Section 1.4, and in Section 6.4.3 for time transform models. When $\mathbf{x}(t)$ is external, the expression (1.4.6) for survival probabilities applies, giving

$$\begin{aligned} Pr(T > t|X) &= \exp\left\{-\int_0^t e^{\beta'\mathbf{x}(u)}\, dH_0(u)\right\} \\ &= S(t|X), \end{aligned} \tag{7.1.44}$$

where $X = \{\mathbf{x}(t),\ t \ge 0\}$ stands for a specific covariate history. It is possible to estimate $S(t|X)$ for a given X by inserting the estimates $\hat{\boldsymbol{\beta}}$ and

$$\hat{H}_0(t) = \sum_{l:\, t_l \le t} \frac{\delta_l}{\sum_{\ell=1}^n Y_\ell(t_l) e^{\hat{\beta}'\mathbf{x}_\ell(t_l)}} \tag{7.1.45}$$

into (7.1.44). Note that $\hat{H}_0(t)$ is the obvious generalization of (7.1.32) to the case of time-varying covariates.

Asymptotic variances for $\hat{H}_0(t)$ and for $\hat{S}(t|X)$ can be obtained as direct generalizations of (7.1.37) and (7.1.39), with $S^{(0)}(t, \boldsymbol{\beta})$ of (7.1.36) generalized to include $\mathbf{x}_l(t)$ and $\bar{\mathbf{x}}(t, \beta)$ of (7.1.7) generalized to

$$\bar{\mathbf{x}}(t, \beta) = \frac{\sum_{\ell=1}^n Y_\ell(t)\mathbf{x}_\ell(t) e^{\beta'\mathbf{x}_\ell(t)}}{S^{(0)}(t, \boldsymbol{\beta})}.$$

For the variance estimate for $\log\{-\log \hat{S}(t|X)\}$, we must replace the terms $\hat{H}_0(t)\mathbf{x}$ in (7.1.39) with

$$\int_0^t \mathbf{x}(u)\, d\hat{H}_0(u) = \sum_{i:\, t_i \le t} \frac{\delta_i \mathbf{x}(t_i)}{S^{(0)}(t_i, \hat{\boldsymbol{\beta}})}.$$

When covariates $\mathbf{x}_i(t)$ are piecewise-constant, as in (7.1.43), it is often convenient for data storage and computation to use as many lines of data for an individual as there are covariate values, in conjunction with the delayed entry notation of Section 7.1.7. For the jth interval in (7.1.43), therefore, we represent the data as $(a_{i,j-1}, a_{ij}, \delta_i, \mathbf{x}_{ij})$; $a_{i,j-1}$ and a_{ij} are the beginning and end of the risk interval, δ_i indicates whether individual i failed at a_{ij}, and $\mathbf{x}_{ij}$ is the (fixed) covariate vector for individual i over the interval. Some software packages require that data involving time-varying covariates be set up in this way. It is easily seen that for the methodology in this chapter or, indeed, Chapter 6, we can think of an individual's data for different time intervals as equivalent to data from separate individuals.

Example 7.1.2. Two-Sample Tests. The two-sample tests of Section 7.1.2 are based on the proportional hazards model and will lack power to detect distributional differences in certain settings. One such case is where the hazard or survivor functions for the two distributions cross somewhere near the medians of the distributions. Better tests can often be obtained by considering a multiplicative model with a defined time-varying covariate. In particular let x be a binary covariate that indicates which distribution an individual's lifetime comes from, and let $g(t)$ be a specified function. Then, instead of (7.1.11), we consider the multiplicative model

$$h(t|x) = h_0(t)e^{\beta'\mathbf{x}(t)}, \tag{7.1.46}$$

where

$$\begin{aligned} \boldsymbol{\beta}'\mathbf{x}(t) &= \beta_1 x_1(t) + \beta_2 x_2(t) \\ &= \beta_1 x + \beta_2 x g(t). \end{aligned} \tag{7.1.47}$$

This stipulates that the hazard ratio for the two distributions is

$$\frac{h(t|1)}{h(t|0)} = \exp[\beta_1 + \beta_2 g(t)],$$

and we can test equality of the two distributions by testing $H: \boldsymbol{\beta} = \mathbf{0}$.

The partial likelihood analysis provides simple tests. As in Section 7.1.2, it is especially simple to consider score tests, based on the score vector $\mathbf{U}(\boldsymbol{\beta})$ at $\boldsymbol{\beta} = \mathbf{0}$ for independent random samples of sizes N_1, N_2 from distributions 1 and 2. For the case of time-varying covariates, the score vector (7.1.8) generalizes to

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^N \delta_i [\mathbf{x}_i(t_i) - \bar{\mathbf{x}}(t_i, \boldsymbol{\beta})],$$

where $N = N_1 + N_2$. With $\mathbf{x}_i(t_l) = (x_l, x_l g(t))'$, this gives

$$\mathbf{U}(\mathbf{0}) = \sum_{l=1}^N \delta_l \left\{ x_l \begin{pmatrix} 1 \\ g(t_l) \end{pmatrix} - \frac{\sum_{\ell=1}^N Y_\ell(t_l) x_\ell \begin{pmatrix} 1 \\ g(t_l) \end{pmatrix}}{\sum_{\ell=1}^n Y_\ell(t_l)} \right\},$$

and in the notation of (7.1.15), this becomes

$$\mathbf{U}(\mathbf{0}) = \sum_{l=1}^N \left(d_{1l} - d_l \frac{n_{1l}}{n_l} \right) \begin{pmatrix} 1 \\ g(t_l) \end{pmatrix}, \tag{7.1.48}$$

with the convention that $x_l = I$ (lifetime is from distribution 1). The information matrix $I(\boldsymbol{\beta})$ generalizing (7.1.9) for time-varying covariates has $\mathbf{x}_\ell(t)$ replacing $\mathbf{x}_\ell$,

and we find that

$$I(0) = \sum_{i=1}^{N} \frac{d_i n_{1i} n_{2i}}{n_i^2} \begin{pmatrix} 1 & g(t_i) \\ g(t_i) & g(t_i)^2 \end{pmatrix}. \tag{7.1.49}$$

The statistic for testing distributional equality is then

$$W = \mathbf{U}(0)' I(0)^{-1} \mathbf{U}(0), \tag{7.1.50}$$

which is asymptotically $\chi^2_{(2)}$ under the null hypothesis. By a judicious choice of the function $g(t)$, we get a test with good power against "crossing" alternatives in which the hazard or survivor functions for the two distributions cross. If the two distributions happen to have proportional hazards, the test is still valid, but will be a little less powerful than the test based on (7.1.15) in Section 7.1.2. Problem 7.4 considers an application of the statistic (7.1.48).

7.1.9 Model Checking

The key assumptions in the PH model are the proportionality assumption concerning the effect of covariates, as represented by (7.1.1), and the specification of the covariate term $\exp(\boldsymbol{\beta}'\mathbf{x})$, or some other analogous function $r(\mathbf{x}; \boldsymbol{\beta})$ in (6.1.6). Most of the general remarks in Sections 6.2 and 6.5.1 on exploring and checking parametric models apply here. In particular, a combination of graphical methods based on residuals or stratification, and formal tests based on model expansion or on a comparison of model-based and empirical estimates is useful. As with parametric models, the presence of heavy censoring can make model assessment difficult.

Model expansion is a crucial tool for model assessment, but informal graphical methods can often provide insight, so we consider them first. In addition, formal test statistics can sometimes be based on residuals from a fitted model.

As discussed in Section 6.2.2, the model (7.1.2) implies that

$$\log[-\log S(t|\mathbf{x})] = \log[-\log S_0(t)] + \boldsymbol{\beta}'\mathbf{x},$$

so if there are very few distinct covariate vectors $\mathbf{x}$ in the data and a sufficient number of individuals with each, then one can stratify individuals according to covariate values and estimate $S(t|\mathbf{x}_j)$ for a specified value $\mathbf{x}_j$ using a Kaplan–Meier estimate, $\hat{S}_j(t)$. If the PH assumption is reasonable, then plots of $\log[-\log \hat{S}_j(t)]$'s versus t or $\log t$ should be roughly vertical translations of one another. Alternatively, Nelson–Aalen estimates $\hat{H}_j(t)$ can be used in place of $-\log \hat{S}_j(t)$. As illustrated in examples such as Example 6.2.2 and Example 6.5.2, such plots are subject to considerable variability and can be difficult to interpret.

7.1.9.1 Martingale and Exponential Residuals

Residuals for fitted PH models are conveniently introduced in either of two ways, which lead effectively to the same thing. The one approach is to consider the cumulative hazard or exponential residuals (6.2.7) or their adjusted version (6.2.8). Here, in the case of fixed covariates, these give

$$\hat{e}_i = \hat{H}(t_i|\mathbf{x}_i) = \hat{H}_0(t_i) e^{\hat{\boldsymbol{\beta}}'\mathbf{x}_i} \tag{7.1.51}$$

$$e_i^{adj} = \hat{e}_i + 1 - \delta_i, \tag{7.1.52}$$

where $\hat{\boldsymbol{\beta}}$ is the estimate from the likelihood (7.1.5), and we use the estimate (7.1.32) for $\hat{H}_0(t)$. As for parametric models, when there is no censoring the $\hat{e}_i$ should be roughly independent of the $\mathbf{x}_i$ if the model (7.1.1) is correct, and resemble a random sample of standard exponential random variables in large samples. However, although plots of $\hat{e}_i$ versus covariate values are helpful in assessing the regression specification, exponential probability plots of the $\hat{e}_i$ cannot be used for overall model assessment. This is because when $H_0(t)$ is treated nonparametrically, the empirical hazard function $\hat{H}_0(t)$ automatically conforms to an exponential model; Problem 7.5 considers this phenomenon. In addition to this difficulty, the residuals $\hat{e}_i$ or $\hat{e}_i^{adj}$ are positive, with a distribution that is highly skewed. Heavy censoring obscures much of the information on model fit and can create systematic patterns in the residuals, as with the parametric models of Chapter 6.

The second approach is called the martingale residual approach. As discussed in Sections 7.1.3 and 7.1.5, the processes $\{M_i(t),\ t \geq 0\}$, where

$$M_i(t) = \int_0^t dM_i(u) = N_i(t) - \int_0^t Y_i(u) e^{\boldsymbol{\beta}'\mathbf{x}_i(u)}\, dH_0(u), \tag{7.1.53}$$

are uncorrelated mean zero martingales. This suggests defining residuals by inserting estimates for $\boldsymbol{\beta}$ and $H_0(u)$ in $M_i(t)$. With $t = \infty$, we get $N_i(\infty) = \delta_i$ and, for the case of fixed covariates,

$$\hat{M}_i = \hat{M}_i(\infty) = \delta_i - \int_0^\infty Y_i(u) e^{\hat{\boldsymbol{\beta}}'\mathbf{x}_i}\, d\hat{H}_0(u) = \delta_i - \hat{H}_0(t_i) e^{\hat{\boldsymbol{\beta}}'\mathbf{x}_i}. \tag{7.1.54}$$

Note by comparison of (7.1.51) and (7.1.54) that

$$\hat{M}_i = 1 - \hat{e}_i^{adj}.$$

The random variables $M_i(\infty)$ are uncorrelated and have mean 0, and it can be argued that in large samples the residuals $\hat{M}_i$ should be approximately uncorrelated with mean 0, either conditionally on the $\mathbf{x}_i$ or unconditionally. Thus, plots of $\hat{M}_i$ versus covariate values should be consistent with a regression mean curve that is approximately $E(\hat{M}_i) = 0$, if (7.1.1) is correct. The distribution of the $\hat{M}_i$ is, however, highly skewed (the $\hat{M}_i$ lie between $-\infty$ and 1), and there can be systematic patterns in plots because of censoring; the situation is very similar to that for fully parametric regression models, as discussed and illustrated in Section 6.2.2.

The $\hat{M}_i$ can be seen to satisfy the two conditions

$$\sum_{i=1}^{n} \hat{M}_i = 0, \qquad \sum_{i=1}^{n} \hat{M}_i x_{ir} = 0 \qquad (r = 1, \ldots, p).$$

To see the first of these note that, using counting process notation,

$$\sum_{i=1}^{n} \hat{M}_i = \sum_{i=1}^{n} \int_0^\infty \left\{ dN_i(t) - \frac{Y_i(t) e^{\hat{\beta}' \mathbf{x}_i(t)} \, dN_\cdot(t)}{\sum_{\ell=1}^{n} Y_\ell(t) e^{\hat{\beta}' \mathbf{x}_\ell(t)}} \right\} = 0,$$

by an interchange of the sum and integral signs. To see the second result, note that

$$\sum_{i=1}^{n} \hat{M}_i \mathbf{x}_i = \mathbf{U}(\hat{\boldsymbol{\beta}}) = \mathbf{0},$$

where $\mathbf{U}(\boldsymbol{\beta})$ is the Cox likelihood score function (7.1.29).

Plots of exponential or martingale residuals $\hat{M}_i$ versus covariate values x_{ir} are often hard to interpret on their own, but if a scatterplot smooth is based on the points $(x_{ir}, \hat{M}_i)$, patterns can be discerned. As discussed in Section 6.2.2, however, heavy censoring limits model assessment from such plots. Another useful device when the covariate structure is not too complex is to group individuals into, say, J groups, $G_1, \ldots, G_J$, on the basis of their covariate values, and to define martingale residual sum processes for each group,

$$\hat{M}_j^+(t) = \sum_{l \in G_j} \hat{M}_l(t). \tag{7.1.55}$$

Plots of the $\hat{M}_j^+(t)$'s versus t can provide insight into model deficiencies if the groups are reasonably large.

7.1.9.2 *Score Residuals and Influence*

The partial likelihood score function $\mathbf{U}(\boldsymbol{\beta})$ is represented as a sum across individuals in two ways in (7.1.28) and (7.1.29). The latter gives a useful definition for a score residual, since the individual terms have expectation 0, whereas the former does not. This leads us to define

$$\hat{\mathbf{U}}_i = \mathbf{U}_i(\hat{\boldsymbol{\beta}}) = \int_0^\infty [\mathbf{x}_i(t) - \bar{\mathbf{x}}(t, \hat{\boldsymbol{\beta}})] \, d\hat{M}_i(t) \tag{7.1.56}$$

as a $p \times 1$ score vector residual for individual i ($i = 1, \ldots, n$). We will not explore the use of these residuals for direct model assessment, but instead use them for estimating the influence of observations.

As discussed in Section 6.2.2, one measure of the influence of the ith observation is given by (6.2.12). For $\hat{\boldsymbol{\beta}}$ this leads to $\hat{\boldsymbol{\beta}}_{(-i)} - \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}_{(-i)}$ is the m.l.e. from the

likelihood (7.1.5) obtained from the set of $n - 1$ observations that excludes the ith. When the data set is large we do not want to calculate $\hat{\boldsymbol{\beta}}_{(-i)}$ for every observation, so we seek approximations that can be computed quickly. The approach used to obtain (6.2.14) for parametric models can be applied here too. The situation is a little different than for the earlier parametric likelihood functions, because dropping the ith observation changes various terms in the log-likelihood (7.1.6). However, by the argument leading to (6.2.13) we have here that for $\boldsymbol{\beta}$ close to $\hat{\boldsymbol{\beta}}$,

$$\mathbf{U}_{(-i)}(\boldsymbol{\beta}) \doteq \mathbf{U}_{(-i)}(\hat{\boldsymbol{\beta}}) - I_{(-i)}(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}),$$

where $\mathbf{U}_{(-i)}(\boldsymbol{\beta})$ and $I_{(-i)}(\boldsymbol{\beta})$ denote the score function and information matrix based on all observations except the ith. Setting $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_{(-i)}$ and noting that $\mathbf{U}_{(-i)}(\hat{\boldsymbol{\beta}}_{(-i)}) = \mathbf{0}$, we have

$$\hat{\boldsymbol{\beta}}_{(-i)} - \hat{\boldsymbol{\beta}} \doteq I_{(-i)}(\hat{\boldsymbol{\beta}})^{-1} \mathbf{U}_{(-i)}(\hat{\boldsymbol{\beta}}).$$

Now,

$$\mathbf{U}_{(-i)}(\hat{\boldsymbol{\beta}}) = \sum_{j \neq i} \int_0^\infty [\mathbf{x}_j - \bar{\mathbf{x}}_{(-i)}(t, \hat{\boldsymbol{\beta}})] \, d\hat{M}_j(t),$$

and assuming that $\bar{\mathbf{x}}_{(-i)}(t, \hat{\boldsymbol{\beta}}) \doteq \bar{\mathbf{x}}(t, \hat{\boldsymbol{\beta}})$ and noting (7.1.56), we get that $\mathbf{U}_{(-i)}(\hat{\boldsymbol{\beta}}) \doteq -\mathbf{U}_i(\hat{\boldsymbol{\beta}})$. Upon making the additional assumption that $I_{(-i)}(\hat{\boldsymbol{\beta}}) \doteq I(\hat{\boldsymbol{\beta}})$, we get the approximation

$$\hat{\boldsymbol{\beta}}_{(-i)} - \hat{\boldsymbol{\beta}} \doteq -I(\hat{\boldsymbol{\beta}})^{-1} \mathbf{U}_i(\hat{\boldsymbol{\beta}}). \tag{7.1.57}$$

The entries D_{ir} of the $n \times p$ matrix D with rows given by (7.1.57) thus give approximations

$$D_{ir} \doteq \hat{\beta}_{r(-i)} - \hat{\beta}_r \qquad r = 1, \ldots, p \tag{7.1.58}$$

for the effect of observation i on each regression coefficient. These case deletion measures are given by several software packages, as are standardized values $D_{ir}/se(\hat{\beta}_r)$. Index plots of these values against i help to identify influential observations. The approximation tends to underestimate influence slightly compared to the true values $\hat{\boldsymbol{\beta}}_{(-i)} - \hat{\boldsymbol{\beta}}$, but is accurate enough for screening purposes. Deletion of single cases for which the values D_{ir} are large can then be examined in more detail. The joint effect of several observations may be greater than the individual deletion measures suggest, and it is wise to scrutinize residuals and covariate leverage values (6.2.16) as well as the D_{ir}.

7.1.9.3 *Model Expansion and Formal Tests*

The expansion of models within families of models like (7.1.1) is familiar as a method of checking on the specification of covariates. For example, interactions or terms representing nonlinear functions of explanatory variables are often added as a

way of checking on a linear specification $\boldsymbol{\beta}'\mathbf{x}$ based on variables $x_1, \ldots, x_k$. Another type of model expansion is to replace a linear term, say $\beta_1 x_1$, with a regression spline or with an arbitrary function $g(x_1)$ that can be estimated nonparametrically. This can be done with some PH software. An advantage of model expansion is that a formal hypothesis test of the baseline model can be made.

Departures from the proportional hazards structure of (7.1.1) are common and should be considered. In some situations the PH assumption may be unsupportable, but a multiplicative specification for $h(t|\mathbf{x})$ can still be reasonable. Semiparametric specifications of the form

$$h(t|\mathbf{x}) = h_0(t)r(\mathbf{x}; \boldsymbol{\beta})w(\mathbf{x}, t; \boldsymbol{\gamma}) \tag{7.1.59}$$

can be handled using time-varying covariates. As an illustration, a check on the effect of covariate x_1 might be made by comparing the models

$$h_1(t|x_1, \mathbf{x}_2) = h_0(t)e^{\beta_1 x_1 + \boldsymbol{\beta}_2'\mathbf{x}_2} \tag{7.1.60}$$

and

$$h(t|x_1, \mathbf{x}_2) = h_0(t)e^{\beta_1 x_1 + \boldsymbol{\beta}_2'\mathbf{x}_2 + \gamma g(t)x_1}, \tag{7.1.61}$$

where $g(t)$ is a known function and γ is a regression coefficient. A test of $H: \gamma = 0$ tests whether x_1 affects the hazard function multiplicatively. A more general expanded model is

$$h(t|\mathbf{x}) = h_0(t)e^{\boldsymbol{\beta}'\mathbf{x} + \boldsymbol{\gamma}'(\mathbf{g}(t)*\mathbf{x})}, \tag{7.1.62}$$

where $\mathbf{g}(t)*\mathbf{x} = (g_1(t)x_1, \ldots, g_p(t)x_p)'$, with the $g_j(t)$'s known functions. Another way to interpret (7.1.62) is to write the exponent as $\boldsymbol{\beta}(t)'\mathbf{x}$, where $\boldsymbol{\beta}(t) = \boldsymbol{\beta} + \mathbf{g}(t)*\boldsymbol{\gamma}$; in other words, the effect of $\mathbf{x}$ on the hazard function changes over time.

Procedures analogous to testing $\gamma = 0$ in (7.1.61) or $\boldsymbol{\gamma} = \mathbf{0}$ in (7.1.62) have been developed by various authors. A test due to Grambsch and Therneau (1994) is implemented in some software packages, and is closely related to a graphical procedure for assessing the proportionality assumption for any specific covariate. The graphical procedure is based on Schoenfeld residuals (Schoenfeld 1980), which are defined in terms of increments in the score function (7.1.28). This leads to $p \times 1$ vectors associated with individuals i whose lifetimes are observed (i.e., with $\delta_i = 1$),

$$\hat{\mathbf{s}}_i = \mathbf{x}_i - \bar{\mathbf{x}}(t_i, \hat{\boldsymbol{\beta}}). \tag{7.1.63}$$

Grambsch and Therneau show that if a model with

$$h(t|\mathbf{x}) = h_0(t)e^{\boldsymbol{\beta}(t)'\mathbf{x}}$$

is correct, but (7.1.1) with constant $\boldsymbol{\beta}$ is fitted, then for $r = 1, \ldots, p$

$$E(\hat{s}_{ir}^* + \hat{\beta}_r) \doteq \beta_r(t_i),$$

where $\hat{s}_{ir}^*$ is the rth element in $V(t_i, \hat{\boldsymbol{\beta}})^{-1}\hat{\mathbf{s}}_i$, with

$$V(t_i, \boldsymbol{\beta}) = \frac{\sum_{\ell=1}^n Y_\ell(t_i)e^{\boldsymbol{\beta}'\mathbf{x}_\ell}[\mathbf{x}_\ell - \bar{\mathbf{x}}(t_i, \boldsymbol{\beta})][\mathbf{x}_\ell - \bar{\mathbf{x}}(t_i, \boldsymbol{\beta})]'}{\sum_{\ell=1}^n Y_\ell(t_i)e^{\boldsymbol{\beta}'\mathbf{x}_\ell}}.$$

The recommended procedure is to plot for each covariate x_r ($r = 1, \ldots, p$) the points $(t_i, \hat{s}_{ir}^*)$, or $(g(t_i), \hat{s}_{ir}^*)$ for some function g, for i's such that $\delta_i = 1$. If the proportionality assumption is satisfactory for x_r (i.e., if $\beta_r(t) = \beta_r$), then a scatterplot smooth through the points should be approximately horizontal and at level $\hat{\beta}_r$; a trend in the plot suggests a time-varying effect for the covariate.

Statistics based on the $\hat{s}_{ir}^*$ have been developed for testing $\gamma = 0$ in (7.1.61) or $\boldsymbol{\gamma} = \mathbf{0}$ in (7.1.62), and are linked to the plots of the points $(g(t_i), s_{ir}^*)$. In addition to Grambsch and Therneau (1994), see Therneau and Grambsch (2000, Sec. 6.2).

Tests based on expanded models are useful to the extent that the expanded family provides an adequate representation of the data. In some cases, a multiplicative specification as in (7.1.62) may be unsatisfactory. In others, it may be satisfactory, but there may exist other types of model providing simpler or more plausible interpretations of the data. It is often worthwhile considering models of different types.

Other methods of assessing goodness of fit are discussed in Chapter 10, where the comparison of model-based and empirical estimates of distributional features is considered.

7.2 EXAMPLES

The examples in this section illustrate techniques discussed in Section 7.1. There are several good software packages that implement methodology for the semiparametric models (7.1.1). The examples here were handled using S-Plus and certain diagnostic checks reflect features that are available in S-Plus; some other software packages have features that are a little different.

Example 7.2.1. Data on the times to a first pulmonary exacerbation for persons with cystic fibrosis entered in a clinical trial were considered in Examples 6.2.3 and 6.3.4. The response variable was time to the first exacerbation, in days, and explanatory variables were treatment (rhDNase or Placebo) and forced expiratory volume (fev) at enrolment. Accelerated failure time models were found to fit the data well, and indicated that rhDNase was associated with longer times to first exacerbation, as was higher baseline fev.

Probability plots of Kaplan–Meier estimates for six treatment–fev (trt–fev) strata were considered in Example 6.2.3. The plots were reasonably consistent with either a PH or an AFT model, but the AFT model provided a little better description of the patterns in the plots, particularly with a normal error distribution in the location-scale form (6.1.4). The PH model is attractive, however, since for events such as infections the hazard function is physically meaningful as an infection intensity. We consider here the semiparametric model (7.1.1).

Table 7.1. Fitted Cox Model for Time to First Exacerbation

Covariate	$\hat{\beta}$	$se(\hat{\beta})$	Z
x_1 (trt)	−.383	.130	−2.95
x_2 (fevc)	−.0206	.0028	−7.44

The results of fitting (7.1.1) with covariates $x_1 = I$ (treatment = rhDNase) and x_2 = fevc (centered fev) are shown in Table 7.1. The values of Z for each covariate are $\hat{\beta}/se(\hat{\beta})$, and significant effects for both treatment and fev are indicated. The magnitude and significance of the effects agree closely with those for the accelerated failure time analysis in Example 6.3.4. In the PH framework here, the estimated effect of the rhDNase treatment is to reduce the hazard function for time to first exacerbation to exp(−.383) = .68 of what it is for the Placebo treatment. The baseline cumulative hazard function $\hat{H}_0(t)$ given by (7.1.32) is shown in Figure 7.1.

Addition of a trt–fevc interaction to the model resulted in a likelihood ratio statistic of virtually 0. Figures 7.2–7.4 show other diagnostic checks on the model: a plot of martingale residuals (7.1.53) against fev, with a scatterplot smooth (lowess in S-Plus) used to estimate trend (Figure 7.2); an index plot of approximations to standardized deletion measures $(\hat{\beta}_{2(-i)} - \hat{\beta}_2)/se(\hat{\beta}_2)$ as in (7.1.58), provided by dfbetas in the S-Plus function coxph (Figure 7.3); plots of Schoenfeld residuals provided by S-Plus function cox.zph (using the "identity" transform) and designed, as described

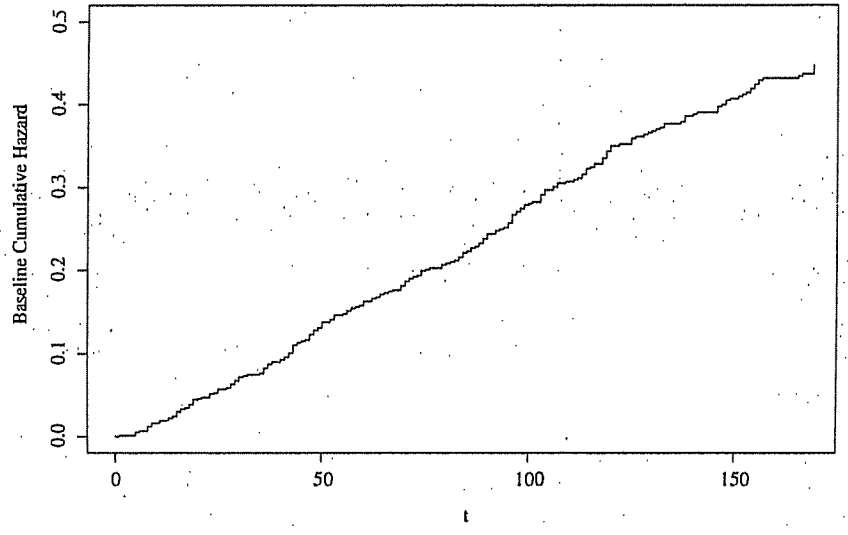

Figure 7.1. Estimated baseline cumulative hazard function (pulmonary exacerbations).

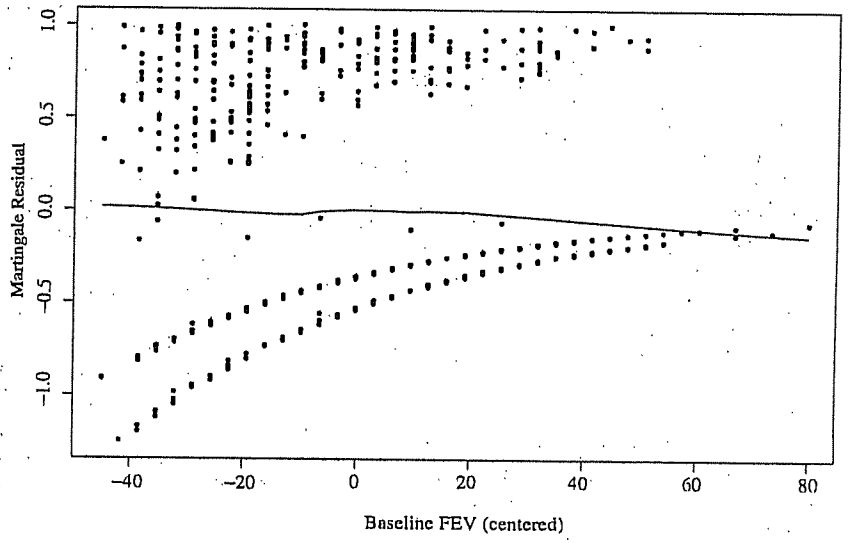

Figure 7.2. Martingale residuals vs. fevc (pulmonary exacerbations).

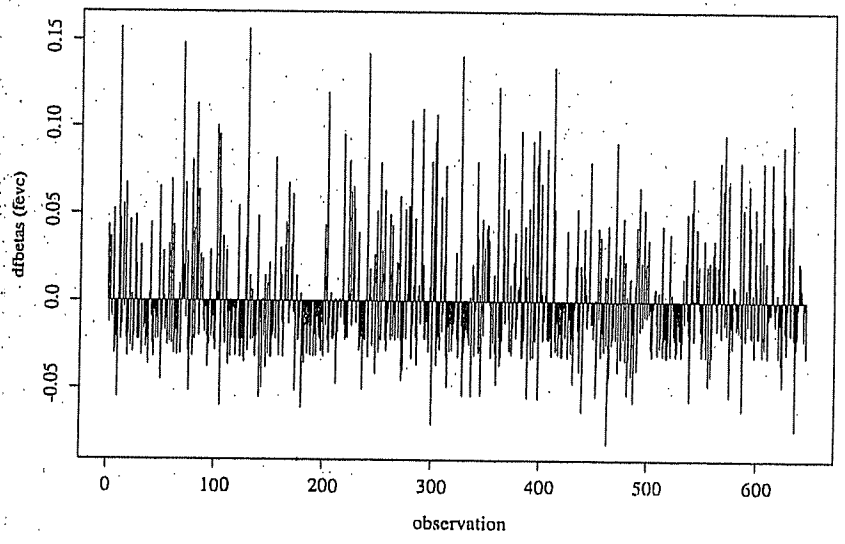

Figure 7.3. Index plot of case deletion measures (dfbetas) for fevc (pulmonary exacerbations).

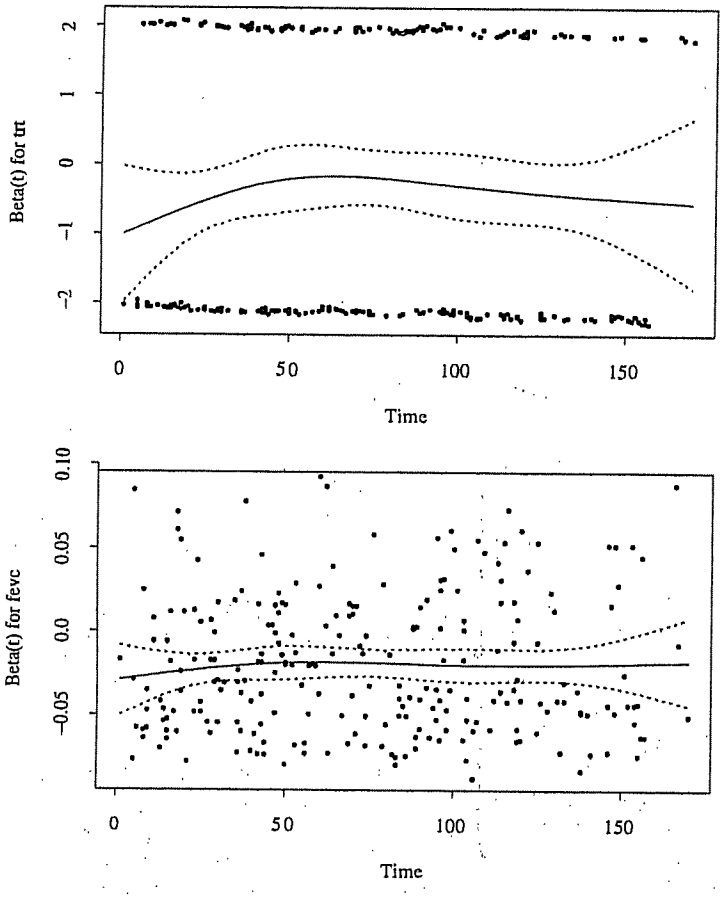

Figure 7.4. Schoenfeld residual plot for detection of time-dependent effects (pulmonary exacerbations).

in Section 7.1.9 and Therneau and Grambsch (2000, Sec. 6.2), to detect nonproportionality of covariate effects (Figure 7.4). None of these plots gives any indication of inadequacies in the model, or of remarkably influential observations. As discussed in Section 7.1.9 and previously in Example 6.2.3, residual plots like Figure 7.2 convey limited information due to the heavy censoring in the data.

Figure 7.4 shows no evidence of nonproportional hazard effects, and an associated global significance test provided by S-Plus function cox.zph gives a p-value of .81. Similar tests can be based on models such as (7.1.61), where time–covariate interactions are included. A good approach that is easily handled by most software packages is to make $g(t)$ in (7.1.61) piecewise constant; the simple model where $g(t) = I(t > a)$ for some specified value a is especially useful. For the current

setting, about half of the observed failure times are 60 days or less, so for illustration we consider $g(t) = I(t > 60)$, which is equivalent to defining additional covariates

$$x_3(t) = x_1 I(t > 60), \qquad x_4(t) = x_2 I(t > 60).$$

The likelihood ratio statistic generated by testing the Cox model with covariates x_1, x_2 against that with covariates $x_1, x_2, x_3(t), x_4(t)$ is $\Lambda = .2$, corresponding to a $\chi^2_{(2)}$ p-value of .90. The estimates and standard errors of β_3 and β_4 in the expanded model are .055(.262) and .0021(.0056). There is no evidence of nonproportionality.

A PH model thus fits the data quite well, and gives results about covariate effects that are very similar to those from the log-normal and log-logistic accelerated failure time (AFT) models in Example 6.3.4. Hazard functions $h(t|\mathbf{x})$ for the PH and AFT models are rather different, but the estimated cumulative hazard functions $H(t|\mathbf{x})$ and survivor functions $S(t|\mathbf{x})$ for the two types of model agree quite closely over the data window (0,169) days. Nonparametric hazard function estimates are not sufficiently precise to provide much further guidance, though we can note that within the AFT family the Weibull model (which is also PH and which gives estimates very close to those for the semiparametric PH model here) is somewhat less well supported than the log-normal and log-logistic models.

As a final illustration of the PH methodology, we give confidence intervals for the probability of no exacerbation within the first 160 days, or $S(160|\mathbf{x})$. The S-Plus function survfit, based on the asymptotic normality of $\log \hat{S}(t|\mathbf{x})$ and variance estimate from (7.1.39), gives approximate .95 confidence intervals for $\mathbf{x} = (0, 0)'$ and $\mathbf{x} = (1, 0)$, respectively, as (.54, .65) and (.65, .75).

Problem 7.1.8 examines another aspect of model specification and treatment comparison for this example.

Example 7.2.2. Data on the survival times of patients with advanced lung cancer were introduced in Example 1.1.9 and considered in Examples 6.3.3 and 6.4.3, where accelerated failure time models were used as a basis for analysis. Here we consider an analysis using multiplicative hazards models.

Table 7.2 shows the results of a PH model (7.1.1) fitted to the full data set ($n = 137$) described in Example 6.4.3, with covariates I(prior therapy = yes), performance status (PS), I(treatment = Standard), I(cell type = Squamous), I(cell type = Small), I(cell type = Adeno), age, and months since diagnosis. Separate PH models fitted to the groups with ($n = 40$) and without ($n = 97$) prior therapy are also shown. The covariate effects suggested by these results are in broad agreement with those from the AFT analysis in Table 6.6 of Example 6.4.3, and the analysis for the therapy = yes group in Example 6.3.3. In particular, PS is an important factor, with higher values associated with longer survival, cell types Small and Adeno are associated with shorter survival times than are types Squamous and Large, and neither age nor months since diagnosis is significant. There is a slight suggestion of a treatment effect from the full data, but the separate analyses for the two therapy groups strongly suggest an interaction, with a significantly higher risk of death associated with the standard treatment for the group not receiving prior therapy.

Table 7.2. PH Models Fitted to VA Lung Cancer Data

Parameter	Full Data ($n = 137$) Estimate	*se*	Therapy ($n = 40$) Estimate	*se*	No Therapy ($n = 97$) Estimate	*se*
β (PS)	−.033	.006	−.060	.014	−.027	.006
β (Squamous)	−.400	.671	−.322	.486	−.467	.361
β (Small)	.457	.266	−.026	.506	.664	.320
β (Adeno)	.789	.303	1.048	.628	.728	.354
β (trt)	.290	.207	−.407	.408	.706	.268
β (age)	−.009	.009	−.012	.021	−.013	.012
β (diagnosis)	.000	.009	.001	.012	.018	.020
β (therapy)	.072	.232	—		—	

Checks on the PH assumption indicate problems, however. The plots and hypothesis tests in S-Plus function cox.zph indicate that the assumption is violated for the important PS covariate, the problem arising mainly with the group not receiving prior therapy. The diagnostic plots suggest that higher PS is associated with a smaller hazard function only up to about 60–80 days, after which there is no significant effect. This can be investigated further by fitting a model with a time-varying covariate, as in (7.1.61). Table 7.3 shows the results for a model with the same covariates as Table 7.2, except with age, diagnosis, and therapy dropped, and the covariate PS $* I(t > 60)$ added. This last covariate is strongly significant, with a Z^2 value of 12.6; the corresponding likelihood ratio statistic for testing this effect is 12.5. The positive estimate $\hat{\beta} = .039$ for the regression coefficient roughly cancels the PS estimate $\hat{\beta}(\text{PS}) = -.044$; this results in almost no effect due to PS after 60 days, as suggested by the diagnostic plots. An expanded model in which therapy and a therapy–treatment interaction are added gives a likelihood ratio statistic $\Lambda = 2.7$ with 2 degrees of freedom; the interaction effect has a regression coefficient estimate (and *se*) of −.66(.52). Thus, we find no evidence of a therapy effect or an interaction with treatment.

The decreasing relevance of PS (measured at entry to the study) as time on study increases is perhaps plausible, and the nature of this effect agrees qualitatively with that in the AFT model fitted in Example 6.4.3, where hazard ratios are not constant

Table 7.3. Model for Lung Cancer Data With Time-Varying PS Effect

Parameter	Estimate	*se*	*Z*
β (PS)	−.044	.007	−6.87
β(PS $* I(t > 60)$)	.039	.011	3.48
β (Squamous)	−.334	.283	−1.18
β (Small)	.543	.267	2.03
β (Adeno)	.741	.300	2.47
β (trt)	.093	.204	.45

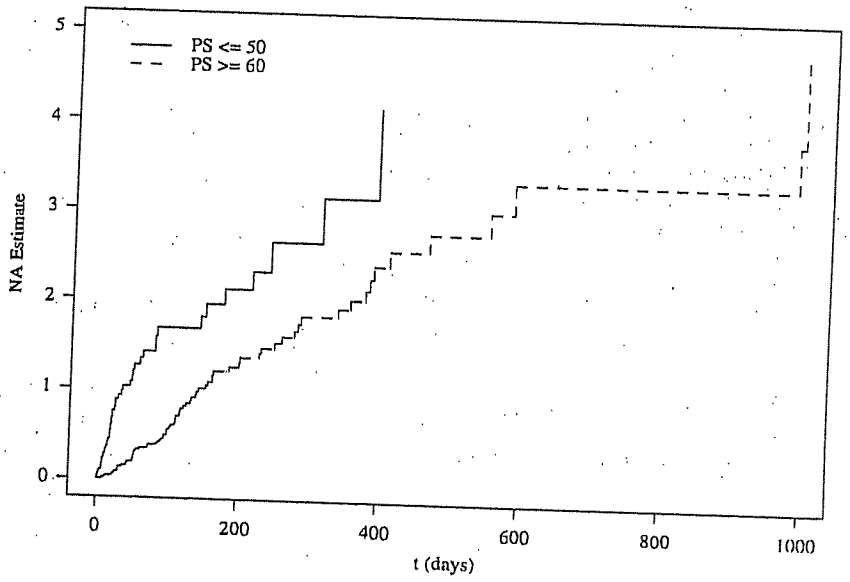

Figure 7.5. Estimated cumulative hazard functions for patients stratified by PS (lung cancer survival).

for individuals with different covariates. Figure 7.5 displays the marginal effect of performance status. It shows Nelson–Aalen cumulative hazard function estimates (3.2.13) for individuals with (a) PS $\leq$ 50 and (b) PS $\geq$ 60. It is seen that after about 60–80 days the two estimates $\hat{H}(t)$ have close to the same slope, suggesting that the hazard functions $h(t)$ are roughly the same. For $t < 60$, the hazard function for individuals with PS $\leq$ 50 is markedly larger. A complicating factor, however, is the association between cell type and PS. Table 7.4 shows a breakdown, and we see that Squamous and Large cell types are associated with higher PS values. The analysis giving Table 7.3 shows an association between these two cell types and longer survival. The effects of PS and cell type are therefore confounded to some degree in this study.

We remark that these data have been discussed many times in the literature. Sources that can be consulted for additional insights are Kalbfleisch and Prentice

Table 7.4. Numbers of Individuals by PS and Cell Type

		Cell Type			
		Squamous	Small	Adeno	Large
PS	10–50	12	22	12	6
	60–90	23	26	15	21

(1980, pp. 60, 89), Farewell and Prentice (1977), Bennett (1983a), Younes and Lachin (1997), and Therneau and Grambsch (2000, p. 135).

7.3 METHODS FOR GROUPED OR DISCRETE LIFETIMES

Discrete response models can be used in two settings. The first is when lifetimes are inherently discrete, for example, when the time to failure of a switch is measured in terms of the number of on-off cycles. The second is when an underlying continuous lifetime can only be observed to lie in certain intervals; this is commonly referred to as grouped or interval-censored data.

Grouped lifetimes were considered in Section 4.3 and in Problem 6.17. Handling them with a fully parametric model poses no significant problems; the main complication is when covariates are time varying, in which case they are usually assumed constant over intervals. Semiparametric models such as (7.1.1) are more problematic. When lifetimes are all grouped into a single set of intervals, the best that can be done is to fit a parametric model in which one group of parameters represents baseline survivor or cumulative hazard function values at the interval endpoints. It is convenient to consider such models here because of their relationship to models like those in Section 7.1, even though they are not strictly semiparametric. This is done in Sections 7.3.1 and 7.3.2. The more complex case of arbitrary interval censoring is deferred to Section 7.4.

Discrete-time regression models are considered in Section 7.3.3. It will be seen that both they and the models used for grouped data are closely related to generalized linear-regression models for discrete responses.

7.3.1 Regression Analysis of Grouped Lifetimes

The situation considered here is that of Section 4.3.1, except that fixed covariates are present. Observations are taken on n individuals, with a lifetime T and a $p \times 1$ vector $\mathbf{x}$ of covariates associated with each individual. The exact $\mathbf{x}$ is known for each individual, but lifetimes are grouped. In particular, it is assumed, as in Section 4.3, that time is partitioned into $k+1$ intervals $I_i = [a_{i-1}, a_i)$, $i = 1, \ldots, k+1$, where $a_0 = 0$ and $a_{k+1} = \infty$, and that we know only in what interval an individual died or was censored.

Define the quantities

$$\begin{aligned} P_l(\mathbf{x}) &= Pr(T \geq a_l | \mathbf{x}) \\ p_l(\mathbf{x}) &= \frac{P_l(\mathbf{x})}{P_{l-1}(\mathbf{x})} \\ &= Pr(T \geq a_l | T \geq a_{l-1}, \mathbf{x}), \end{aligned} \tag{7.3.1}$$

with $P_0(\mathbf{x})$ equal to one for all $\mathbf{x}$. As in Sections 3.6 and 4.3.1, there is a need for explicit assumptions regarding censoring times. For the time being we assume that all censoring takes place at the ends of the intervals; modifications will be mentioned later for occasions where this assumption is unreasonable. Let R_l be the risk set at time a_{l-1}, D_l the set of individuals observed to die in $I_l = [a_{l-1}, a_l)$, and C_l the set of individuals censored in I_l. Under assumptions like those in Section 2.3.1, the likelihood function is then

$$L = \prod_{i=1}^{k} \left(\prod_{l \in D_i} [1 - p_l(\mathbf{x}_l)] \prod_{l \in R_l - D_l} p_l(\mathbf{x}_l) \right), \tag{7.3.2}$$

remembering that all censoring in I_i is assumed to take place just prior to a_i, after all deaths in I_i have occurred.

Fully parametric models such as those in Chapter 6 give parametric expressions $P_l(\mathbf{x}; \boldsymbol{\theta})$ and $p_l(\mathbf{x}; \boldsymbol{\theta})$. Here we consider models with weaker assumptions, analogous to the semiparametric models of Section 7.1. Two models will be discussed in some detail; they are as follows.

7.3.1.1 *A Model Based on Grouped Lifetimes from (7.1.5)*

If lifetimes T given $\mathbf{x}$ from a PH model (7.1.2) are grouped into intervals in the way described here, then a grouped regression model is obtained, for which

$$\begin{aligned} P_l(\mathbf{x}) &= S(a_l | \mathbf{x}) \\ &= S_0(a_l)^{\exp(\boldsymbol{\beta}'\mathbf{x})} = P_l^{\exp(\boldsymbol{\beta}'\mathbf{x})}, \end{aligned} \tag{7.3.3}$$

where

$$P_l = S_0(a_l), \qquad l = 1, \ldots, k,$$

with $P_0 = 1$. This gives

$$p_l(\mathbf{x}) = \frac{P_l(\mathbf{x})}{P_{l-1}(\mathbf{x})} = p_l^{\exp(\boldsymbol{\beta}'\mathbf{x})}, \tag{7.3.4}$$

where

$$p_l = \frac{P_l}{P_{l-1}}, \qquad l = 1, \ldots, k.$$

This model produces a likelihood (7.3.2) that can be used for inferences about $\boldsymbol{\beta}$ and $p_1, \ldots, p_k$. Note that although the underlying continuous-time model is semiparametric, the discrete model involves only $S_0(a_1), \ldots, S_0(a_k)$, so is essentially parametric. Unlike the parametric models in Chapter 6, however, no parametric relationship is assumed for the $S_0(a_j)$'s.

7.3.1.2 *A Logistic Model*

Model (7.3.3) is appealing because it is based on the continuous PH model. We also consider a second model that is useful for analyzing grouped data, though it cannot

be obtained by grouping from a continuous-time model. This models takes $p_i(\mathbf{x})$ of the form

$$p_i(\mathbf{x}) = \left(1 + \gamma_i e^{\beta'\mathbf{x}}\right)^{-1} \qquad i = 1, \ldots, k, \tag{7.3.5}$$

where

$$\gamma_i = \frac{1 - p_i(\mathbf{0})}{p_i(\mathbf{0})}, \qquad i = 1, \ldots, k.$$

This is a logistic model; note that

$$\log\left(\frac{1 - p_i(\mathbf{x})}{p_i(\mathbf{x})}\right) = \log \gamma_i + \boldsymbol{\beta}'\mathbf{x}.$$

This model is flexible and convenient, as logistic models are in many other discrete data situations (e.g., Cox and Snell 1989). As the number of intervals increases and interval lengths decrease, (7.3.5) and (7.3.4) agree more and more closely. In the limit as interval lengths approach 0 it is easily seen that if the p_i approach 1, then

$$\frac{1 - p_i^{\exp(\beta'\mathbf{x})}}{p_i^{\exp(\beta'\mathbf{x})}} \sim \frac{1 - p_i}{p_i} e^{\beta'\mathbf{x}}.$$

7.3.1.3 Maximum Likelihood Estimation

The models (7.3.4) and (7.3.5) are of generalized linear form: defining $\alpha_i = \log(-\log p_i)$ for (7.3.4) and $\alpha_i = \log \gamma_i$ for (7.3.5), we can write them as

$$p_i(\mathbf{x}) = \exp\left(-e^{\alpha_i + \beta'\mathbf{x}}\right), \qquad i = 1, \ldots, k \tag{7.3.6}$$

and

$$p_i(\mathbf{x}) = \left(1 + e^{\alpha_i + \beta'\mathbf{x}}\right)^{-1}, \qquad i = 1, \ldots, k \tag{7.3.7}$$

respectively. Furthermore, the likelihood function (7.3.2) has the same form as for a binary response regression model, and so software for discrete generalized linear models (g.l.m.'s) can be used to fit (7.3.6) or (7.3.7). The link functions, which in g.l.m. terminology correspond to the inverses of $\exp(-e^z)$ and $(1 + e^z)^{-1}$, are the log-log and logit functions, respectively, (e.g., McCullagh and Nelder 1989).

Generalized linear model software can be used by defining k indicator covariates and then associating a separate observation $(y, \mathbf{x}^*)$ with every interval for which each individual is at risk. That is, an individual ℓ is at risk over $[a_{i-1}, a_i)$ if they are alive and uncensored at a_{i-1}, and the observation $(y, \mathbf{x}^*)$ associated with that individual

and interval can be taken as

$$y = I \text{ (individual } \ell \text{ did not die in } [a_{i-1}, a_i))$$
$$\mathbf{x}^* = (\mathbf{x}_\ell, \mathbf{z}_i),$$

where $\mathbf{z}_i$ is a $k \times 1$ vector with 1 in position i and 0 elsewhere ($i = 1, \ldots, k$).

Although it is generally convenient to use g.l.m. software, we will give expressions for the log-likelihood and score functions for the grouped PH and logistic models; they will be used in Section 7.3.2. Some issues concerning estimation of survival probabilities, censoring, and time-varying covariates will also be considered.

The log-likelihood from (7.3.2) and the grouped PH model (7.3.6) is

$$\ell(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{k} \left(\sum_{\ell \in D_i} \log[e^{\exp(\alpha_i + \beta'\mathbf{x}_\ell)} - 1] - \sum_{\ell \in R_i} e^{\alpha_i + \beta'\mathbf{x}_\ell} \right). \tag{7.3.8}$$

To write down the first derivatives of $\ell(\boldsymbol{\beta}, \boldsymbol{\alpha})$ define

$$z_{i\ell} = e^{\alpha_i + \beta'\mathbf{x}_\ell} \qquad i = 1, \ldots, k; \qquad \ell = 1, \ldots, n.$$

Then

$$\frac{\partial \ell}{\partial \beta_r} = \sum_{i=1}^{k} \left(\sum_{\ell \in D_i} \frac{x_{\ell r} z_{i\ell}}{1 - e^{-z_{i\ell}}} - \sum_{\ell \in R_i} x_{\ell r} z_{i\ell} \right) \qquad r = 1, \ldots, p \tag{7.3.9}$$

$$\frac{\partial \ell}{\partial \alpha_i} = \sum_{\ell \in D_i} \frac{z_{i\ell}}{1 - e^{-z_{i\ell}}} - \sum_{\ell \in R_i} z_{i\ell} \qquad i = 1, \ldots, k. \tag{7.3.10}$$

The m.l.e.'s $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$ are easy to obtain with standard optimization or equation-solving software. When $\boldsymbol{\beta} = \mathbf{0}$, equations (7.3.10) give estimates

$$\hat{p}_i = e^{-\exp(\hat{\alpha}_i)} = \frac{n_i - d_i}{n_i},$$

where $d_i = |D_i|$ and $n_i = |R_i|$ are the number of deaths and the number at risk in the ith interval, respectively. These are the life table estimates of p_i in Section 3.6, when censoring is at the ends of intervals. They can be used as initial estimates of $\alpha_1, \ldots, \alpha_k$ for a maximum likelihood iteration procedure.

With $p_i(\mathbf{x})$ given by the logistic model (7.3.7), the log-likelihood function from (7.3.2) can be written as

$$\ell(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{k} \left(\sum_{\ell \in D_i} (\alpha_i + \boldsymbol{\beta}'\mathbf{x}_\ell) - \sum_{\ell \in R_i} \log(1 + e^{\alpha_i + \beta'\mathbf{x}_\ell}) \right). \tag{7.3.11}$$

First derivatives of $\ell(\boldsymbol{\beta}, \boldsymbol{\alpha})$ are

$$\frac{\partial \log L}{\partial \beta_r} = \sum_{l=1}^{k} \left(\sum_{\ell \in D_l} x_{\ell r} - \sum_{\ell \in R_l} \frac{x_{\ell r} e^{\alpha_l + \boldsymbol{\beta}' \mathbf{x}_\ell}}{1 + e^{\alpha_l + \boldsymbol{\beta}' \mathbf{x}_\ell}} \right) \qquad r = 1, \ldots, p \qquad (7.3.12)$$

$$\frac{\partial \log L}{\partial \alpha_l} = d_l - \sum_{\ell \in R_l} \frac{e^{\alpha_l + \boldsymbol{\beta}' \mathbf{x}_\ell}}{1 + e^{\alpha_l + \boldsymbol{\beta}' \mathbf{x}_\ell}} \qquad l = 1, \ldots, k. \qquad (7.3.13)$$

The m.l.e.'s $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}$ are once again easily obtained. When $\boldsymbol{\beta} = \mathbf{0}$, (7.3.13) yields the life table estimates $\hat{p}_l = (1 + e^{\hat{\alpha}_l})^{-1} = (n_l - d_l)/n_l$, which can be used as initial estimates in an iteration procedure.

Tests about $\boldsymbol{\beta}$ are of particular interest in many situations. In the case of the logistic model, a partial likelihood for $\boldsymbol{\beta}$ is also available and can be used for inferences about $\boldsymbol{\beta}$ in the absence of knowledge of $p_1, \ldots, p_k$, though this requires excessive computation if the number of deaths per interval is substantial. This approach is described later. More generally, standard large-sample procedures based on normal approximations for $\hat{\boldsymbol{\beta}}$ or likelihood ratio statistics can be used for tests or interval estimation. The constancy of $\boldsymbol{\beta}$ across intervals can also be checked, by allowing $\boldsymbol{\beta}$ in (7.3.6) or (7.3.7) to depend on $i = 1, \ldots, k$. This is analogous to checking the proportional hazards assumption in continuous PH models.

One may also want estimates or tests for survival probabilities,

$$P_l(\mathbf{x}) = \prod_{j=1}^{l} p_j(\mathbf{x}).$$

A satisfactory procedure in most instances is to treat $\log \hat{P}_l(\mathbf{x})$ as normally distributed, with mean $\log P_l(\mathbf{x})$ and variance obtained as follows. Let $\hat{z}_j(\mathbf{x}) = \hat{\alpha}_j + \hat{\boldsymbol{\beta}}' \mathbf{x}$ for $j = 1, \ldots, k$ and let $\mathbf{z}(\mathbf{x}) = (z_1(\mathbf{x}), \ldots, z_k(\mathbf{x}))'$. It follows from Theorem B2 in Appendix B that the asymptotic covariance matrix of $\hat{\mathbf{z}}(\mathbf{x})$ is given by

$$\text{Asvar}[\hat{\mathbf{z}}(\mathbf{x})] = L(\mathbf{x}) \text{Asvar}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) L(\mathbf{x})',$$

where $\text{Asvar}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \mathcal{I}(\boldsymbol{\alpha}, \boldsymbol{\beta})^{-1}$ is the asymptotic covariance matrix for $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ and $L(\mathbf{x})$ is a $k \times (k + p)$ matrix with the $k \times k$ identity matrix giving the first k columns and the last p columns given by placing the vector $\mathbf{x}$ in each row. An application of the asymptotic variance formula (B2) then gives the asymptotic variance of $\log \hat{P}_l(\mathbf{x})$ as

$$\text{Asvar}[\log \hat{P}_l(\mathbf{x})] = \mathbf{a}(\mathbf{x})' L(\mathbf{x}) \text{Asvar}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) L(\mathbf{x})' \mathbf{a}(\mathbf{x})$$

where $\mathbf{a}(\mathbf{x})$ is a $k \times 1$ vector with elements

$$a_j(\mathbf{x}) = \partial \log \hat{P}_l(\mathbf{x}) / \partial z_j(\mathbf{x}).$$

A variance estimate for $\log \hat{P}_l(\mathbf{x})$ is obtained by replacing $\mathcal{I}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ with $I(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ with $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ in all expressions.

For the grouped PH model, for example,

$$\log \hat{P}_l(\mathbf{x}) = -\sum_{j=1}^{l} \exp[\hat{z}_j(\mathbf{x})]$$

so $a_j(\mathbf{x}) = -I(j \le i) \exp[\hat{z}_j(\mathbf{x})]$.

7.3.1.4 *A Partial Likelihood for the Logistic Model*

It turns out that for the logistic model a partial likelihood function for $\boldsymbol{\beta}$ can be obtained; this is not possible for the grouped PH model. This partial likelihood is developed by considering for each interval I_l the probability of the observed set of individuals who die, conditional on the observed number of deaths. For $d_l > 0$, we have

$$\begin{aligned} &Pr(\text{the individuals in } D_l \text{ die in } I_l | d_l \text{ individuals die in } I_l) \\ &= \left(\prod_{\ell \in D_l} e^{\boldsymbol{\beta}' \mathbf{x}_\ell} \right) \Big/ \sum_{\text{all } D_j} \left(\prod_{\ell \in D_j} e^{\boldsymbol{\beta}' \mathbf{x}_\ell} \right) \\ &= e^{\boldsymbol{\beta}' \mathbf{s}_l} \Big/ \sum_{\text{all } D_j} e^{\boldsymbol{\beta}' \mathbf{s}_j}. \end{aligned} \qquad (7.3.14)$$

In (7.3.14) the sum in the denominator is over all possible d_l-subsets D_j of R_l, and $\mathbf{s}_j = \sum_{\ell \in D_j} \mathbf{x}_\ell$ is the sum of the covariate vectors for individuals in D_j. Multiplying the factors for different intervals gives the partial likelihood

$$L_1(\boldsymbol{\beta}) = \prod_{l=1}^{k} \left(e^{\boldsymbol{\beta}' \mathbf{s}_l} \Big/ \sum_{\text{all } D_j} e^{\boldsymbol{\beta}' \mathbf{s}_j} \right). \qquad (7.3.15)$$

This is a partial likelihood according to the definition in Appendix C, and can be employed as though it is an ordinary likelihood. It usually requires excessive computation, unless the d_l are small, because of the large number of d_l-subsets of R_l. This likelihood is useful, however, for testing that $\boldsymbol{\beta} = \mathbf{0}$, where a statistic based on the score function from (7.3.15) is easily computed. This is discussed in Section 7.3.2 and is used to test the equality of two or more lifetime distributions on the basis of grouped data.

7.3.1.5 *Assumptions About Censoring*

It has been assumed thus far that all censoring takes place at the ends of intervals. If this assumption is unreasonable, modifications can be made to the preceding methods; these will usually have to be somewhat ad hoc, as discussed in Section 3.6. If censoring times are more or less uniformly distributed across intervals, for example,

Thompson (1977) suggests the following procedure, which was introduced in Problem 3.18. Partition R_i into three groups: D_i is the set of individuals who die in I_i, C_i is the set of individuals censored in I_i, and $G_i = R_i - D_i - C_i$ is the set of individuals surviving beyond I_i. The likelihood (7.3.2) is then replaced by

$$\prod_{i=1}^{k}\left(\prod_{\ell\in D_i}[1-p_i(\mathbf{x}_\ell)]\prod_{\ell\in G_i}p_i(\mathbf{x}_\ell)\prod_{\ell\in C_i}p_i(\mathbf{x}_\ell)^{1/2}\right). \tag{7.3.16}$$

With (7.3.16), the only difference in the maximum likelihood formulas given earlier is that sums $\sum_{\ell\in R_i}$ are replaced by

$$\sum_{\ell\in R_i-C_i}+\frac{1}{2}\sum_{\ell\in C_i}.$$

That is, the contribution to the log-likelihood of individuals censored in I_i is halved. If g.l.m. software allows case weights for individual observations, this modification is easily accommodated. This is an ad hoc but reasonable approach which, when $\boldsymbol{\beta}=\mathbf{0}$, leads to the standard life table estimates of p_i given by one minus (3.6.3).

7.3.1.6 *Time-Varying Covariates*

Time-varying covariates are accommodated in a straightforward way if their values can be assumed constant over intervals I_i. In that case individual ℓ has covariate vectors

$$\mathbf{x}_\ell(t)=\mathbf{x}_{\ell i} \qquad a_{i-1}\le t<a_i.$$

Since multiple terms are associated with individuals in the likelihood function (7.3.2), corresponding to the number of intervals I_i over which they are at risk, no essential complication in methodology ensues. The same is true if g.l.m. software for binary response regression models is used to handle (7.3.2), as described earlier.

7.3.2 Testing the Equality of Distributions with Grouped Data

Tests for the equality of two or more lifetime distributions, analogous to those in Section 7.1.2, can be obtained from the regression models (7.3.6) or (7.3.7). The objective is to test the equality of $m\ge 2$ distributions, represented by the hypothesis

$$H: S_1(t)=\cdots=S_m(t). \tag{7.3.17}$$

In the grouped data scenarios considered here, we assume that the lifetimes in m independent samples from $S_1(t),\ldots,S_m(t)$ are subject to the same grouping intervals $I_1,\ldots,I_{k+1}$. The best that we can do is to compare the $S_r(t)$'s at the interval endpoints $a_1,\ldots,a_k$, and so we can replace (7.3.17) with the hypothesis

$$H: p_{1i}=\cdots=p_{mi} \qquad \text{for all} \qquad i=1,\ldots,k, \tag{7.3.18}$$

where

$$p_{ri}=S_r(a_i)/S_r(a_{i-1}), \qquad i=1,\ldots,k. \tag{7.3.19}$$

Tests can be based on either of the regression models (7.3.6) or (7.3.7) by defining $m-1$ indicator covariates that identify which distribution an individual's lifetime comes from. The models imply a relationship for the p_{ri}, analogous to the relationship (7.1.17) among the survivor functions in the case of continuous observation of lifetimes. In particular, consider the case $m=2$ and define the covariate

$$x = I \text{ (individual's lifetime is from } S_1(t)).$$

Then (7.3.6) implies that $p_{1i}=p_i(1)$ and $p_{2i}=p_i(0)$ for $i=1,\ldots,k$ are of the form

$$p_{1i}=\exp(-e^{\alpha_i+\beta}), \qquad p_{2i}=\exp(-e^{\alpha_i}), \tag{7.3.20}$$

whereas (7.3.7) implies that

$$p_{1i}=(1+e^{\alpha_i+\beta})^{-1}, \qquad p_{2i}=(1+e^{\alpha_i})^{-1}. \tag{7.3.21}$$

The relationships (7.3.20) and (7.3.21) represent quite strong assumptions, but are reasonable in certain situations. Checks on the assumptions can be made by estimating the conditional probabilities p_{1i} and p_{2i} nonparametrically, as in Section 3.6. If all censoring is at the ends of the intervals I_i, then

$$\hat{p}_{ri}=1-d_{ri}/n_{ri} \qquad i=1,\ldots,k; \qquad r=1,2,$$

where

$$d_{ri} = \text{number of lifetimes in } I_i \text{ from } S_r(t)$$
$$n_{ri} = \text{number of individuals at risk in } I_i \text{ from } S_r(t).$$

Under (7.3.21),

$$\log\left(\frac{1-p_{1i}}{p_{1i}}\right)=\log\left(\frac{1-p_{2i}}{p_{2i}}\right)-\beta, \tag{7.3.22}$$

so the values $\log((1-\hat{p}_{1i})/\hat{p}_{1i})$ and $\log((1-\hat{p}_{2i})/\hat{p}_{2i})$ can be compared numerically or graphically to assess the validity of (7.3.22). Similarly, (7.3.20) can be assessed by comparing $\log(-\log\hat{p}_{1i})$ and $\log(-\log\hat{p}_{2i})$ for $i=1,\ldots,k$.

Tests of (7.3.17) or (7.3.18) can be carried out by testing that $\boldsymbol{\beta} = \mathbf{0}$ in the regression models (7.3.6) or (7.3.7); for the case $m = 2$, this amounts to a test of $\beta = 0$ in (7.3.20) or (7.3.21). This can be done using g.l.m. software and test statistics based on $\hat{\boldsymbol{\beta}}$ or on log-likelihoods. As in Section 7.1.2, it is also possible to derive simple score tests that do not require the calculation of $\hat{\boldsymbol{\beta}}$. These are obtained by considering the score vector $\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\alpha})$ for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ from the likelihood (7.3.2) in partitioned form

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{bmatrix} = \begin{bmatrix} \partial \log L/\partial \boldsymbol{\beta} \\ \partial \log L/\partial \boldsymbol{\alpha} \end{bmatrix},$$

with the information matrix $I(\boldsymbol{\beta}, \boldsymbol{\alpha})$ partitioned similarly as

$$I(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix} = \begin{bmatrix} -\partial^2 \log L/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}' & -\partial^2 \log L/\partial \boldsymbol{\beta} \partial \boldsymbol{\alpha}' \\ -\partial^2 \log L/\partial \boldsymbol{\alpha} \partial \boldsymbol{\beta}' & -\partial^2 \log L/\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}' \end{bmatrix}.$$

As discussed in Appendix C, a score test of $H: \boldsymbol{\beta} = \mathbf{0}$ can be based on the statistic

$$\mathbf{U}_1[\mathbf{0}, \hat{\boldsymbol{\alpha}}(\mathbf{0})] = \left(\frac{\partial \log L}{\partial \boldsymbol{\beta}} \right)_{[\mathbf{0}, \hat{\boldsymbol{\alpha}}(\mathbf{0})]}, \tag{7.3.23}$$

where $\hat{\boldsymbol{\alpha}}(\mathbf{0})$ is the m.l.e. of $\boldsymbol{\alpha}$ when $\boldsymbol{\beta} = \mathbf{0}$; recall that $\hat{\boldsymbol{\alpha}}(\mathbf{0})$ has a simple closed-form expression. If $\boldsymbol{\beta} = \mathbf{0}$, $\mathbf{U}_1[\mathbf{0}, \hat{\boldsymbol{\alpha}}(\mathbf{0})]$ is asymptotically normal with mean vector $\mathbf{0}$ and covariance matrix estimated by

$$V_1 = I_{11}(\mathbf{0}, \hat{\boldsymbol{\alpha}}(\mathbf{0})) - I_{12}(\mathbf{0}, \hat{\boldsymbol{\alpha}}(\mathbf{0}))I_{22}(\mathbf{0}, \hat{\boldsymbol{\alpha}}(\mathbf{0}))^{-1} I'_{12}(\mathbf{0}, \hat{\boldsymbol{\alpha}}(\mathbf{0})), \tag{7.3.24}$$

and $W = \mathbf{U}_1[\mathbf{0}, \hat{\boldsymbol{\alpha}}(\mathbf{0})]' V_1^{-1} \mathbf{U}_1[\mathbf{0}, \hat{\boldsymbol{\alpha}}(\mathbf{0})]$ is asymptotically $\chi^2_{(m-1)}$. Large values of W provide evidence against the hypothesis that $\boldsymbol{\beta} = \mathbf{0}$.

The score vectors for the models (7.3.6) and (7.3.7) are given by expressions (7.3.9), (7.3.10), (7.3.12), and (7.3.13), and the information matrices $I(\boldsymbol{\beta}, \boldsymbol{\alpha})$ can be obtained by differentiating the expressions. In the case of the logistic model (7.3.7), an alternative procedure is to base a score test on the partial likelihood function (7.3.15); this has the advantage that $\boldsymbol{\alpha}$ does not need to be considered at all. Procedures of both types will be described briefly. It will be assumed that all censoring occurs at interval endpoints; if this is not the case, then a modification to the likelihood (7.3.2) can be considered, such as (7.3.16).

7.3.2.1 A Score Test Based on the Grouped PH Model (7.3.6)

Let us consider a test of equality of $m \geq 2$ distributions, represented by the hypothesis (7.3.18). Define a vector $\mathbf{x}$ of $m - 1$ indicator covariates, such that individuals from populations (distributions) $1, \ldots, m$ have $\mathbf{x}$ vectors $(1, 0, \ldots, 0)$, $(0, 1, \ldots, 0), \ldots, (0, \ldots, 0, 1)$, and $(0, \ldots, 0)$, respectively. That is, the rth covariate ($r = 1, \ldots, m - 1$) equals 1 if and only if an individual is from population r. We assume that the p_{ri} in (7.3.18) are given by the model (7.3.6), and test that $\boldsymbol{\beta} = \mathbf{0}$.

It is readily found from (7.3.10) that

$$\hat{\alpha}_i(0) = \log\left[-\log\left(\frac{n_i - d_i}{n_i} \right) \right], \qquad i = 1, \ldots, k,$$

where $d_i = \sum d_{ri}$ and $n_i = \sum n_{ri}$ are the total number of lifetimes and total number of individuals at risk in I_i, across the samples from all m populations. It then follows from (7.3.9) that $\mathbf{U}_1 = \mathbf{U}_1[\mathbf{0}, \hat{\boldsymbol{\alpha}}(\mathbf{0})]$ has entries

$$U_{1r} = \sum_{i=1}^{k} \left(n_{ri} - \frac{n_i d_{ri}}{d_i} \right) \log\left(\frac{n_i - d_i}{n_i} \right) \qquad r = 1, \ldots, m-1. \tag{7.3.25}$$

Note that (7.3.25) can be rewritten as a weighted sum of terms $(d_{ri} - n_{ri} d_i/n_i)$, which takes the "expected–observed" deaths form seen in (7.1.18). The covariance matrix V_1 in (7.3.24) can be obtained after some algebra. Prentice and Gloeckler (1978) noted that V_1 can be replaced with a slightly simpler version V, which is obtained by taking conditional expectations of terms in $I(\mathbf{0}, \hat{\boldsymbol{\alpha}}(\mathbf{0}))$: specifically, the expectations $E(d_{ri} | d_i, n_{ri}, n_i) = n_{ri} d_i / n_i$ replace d_{ri} for $i = 1, \ldots, k$ and $r = 1, \ldots, m - 1$. The matrix V has entries

$$V_{rs} = \sum_{i=1}^{k} \frac{n_i - d_i}{d_i} \left[\log\left(\frac{n_i - d_i}{n_i} \right) \right]^2 \left(n_{ri} \delta_{rs} - \frac{n_{ri} n_{si}}{n_i} \right) \tag{7.3.26}$$

for $r, s = 1, \ldots, m - 1$, where $\delta_{rs} = I(r = s)$.

Under the hypotheses (7.3.17) or (7.3.18) that the m distributions are the same, the statistic

$$W = \mathbf{U}_1' V \mathbf{U}_1, \tag{7.3.27}$$

is approximately $\chi^2_{(m-1)}$; large values of W provide evidence against the hypothesized equality.

7.3.2.2 A Partial Likelihood Score Test Based on the Logistic Model

Consider the regression model (7.3.7), with the $(m-1) \times 1$ covariate vector $\mathbf{x}$ defined in the same way as for the grouped PH model in the preceding subsection. The log partial likelihood for $\boldsymbol{\beta}$ is, from (7.3.15),

$$\ell_1(\boldsymbol{\beta}) = \sum_{i=1}^{k} \boldsymbol{\beta}' \mathbf{s}_i - \sum_{i=1}^{k} \log\left(\sum_{\text{all } D_j} e^{\boldsymbol{\beta}' \mathbf{s}_j} \right),$$

where $\mathbf{s}_j = \sum_{\ell \in D_j} \mathbf{x}_\ell$ is the sum of the covariate vectors for all individuals in a particular set D_j and the sum $\sum_{\text{all } D_j}$ is taken over all d_i-subsets of R_i. The first and second derivatives of $\ell_1(\boldsymbol{\beta})$ are

$$\frac{\partial \ell_1}{\partial \beta_r} = \sum_{i=1}^{k} s_{ir} - \sum_{i=1}^{k} \left(\frac{\sum_{\text{all } D_j} s_{jr} e^{\beta' s_j}}{\sum_{\text{all } D_j} e^{\beta' s_j}} \right) \qquad r = 1, \ldots, m-1 \tag{7.3.28}$$

$$\frac{\partial^2 \ell_1}{\partial \beta_r \partial \beta_t} =$$

$$- \sum_{i=1}^{k} \frac{\left(\sum_{\text{all } D_j} s_{jr} s_{jt} e^{\beta' s_j}\right)\left(\sum_{\text{all } D_j} e^{\beta' s_j}\right) - \left(\sum_{\text{all } D_j} s_{jr} e^{\beta' s_j}\right)\left(\sum_{\text{all } D_j} s_{jt} e^{\beta' s_j}\right)}{\left(\sum_{\text{all } D_j} e^{\beta' s_j}\right)^2}$$

$$r, t = 1, \ldots, m-1. \tag{7.3.29}$$

The score vector is $\mathbf{U}(\boldsymbol{\beta}) = (\partial \ell_1 / \partial \beta_r)$ and the information matrix is $\mathbf{I}(\boldsymbol{\beta}) = (-\partial^2 \ell_1 / \partial \beta_r \partial \beta_t)$. Under $H: \boldsymbol{\beta} = \mathbf{0}$, $\mathbf{U}(\mathbf{0})$ is approximately normal with mean $\mathbf{0}$ and covariance matrix $I(\mathbf{0})$, and $W = \mathbf{U}(\mathbf{0})' I(\mathbf{0})^{-1} \mathbf{U}(\mathbf{0})$ is asymptotically $\chi^2_{(m-1)}$. Large values of W provide evidence against the hypothesis, and p-values can be computed using the χ^2 approximation.

Although $\mathbf{U}(\boldsymbol{\beta})$ and $I(\boldsymbol{\beta})$ require a lot of computation in the general case, $\mathbf{U}(\mathbf{0})$ and $I(\mathbf{0})$ take simple forms, particularly for the m-sample problem. Noting that in (7.3.28) and (7.3.29) the sums over "all D_j" are over all d_i-subsets of R_i, we have the following results for $i = 1, \ldots, k$ and $r = 1, \ldots, m-1$:

$$s_{ir} = \sum_{\ell \in D_i} x_{\ell r} = d_{ri}$$

$$\sum_{\text{all } D_j} 1 = \binom{n_i}{d_i}$$

$$\sum_{\text{all } D_j} s_{jr} = \sum_{\text{all } D_j} \sum_{\ell \in D_j} x_{\ell r}$$

$$= \sum_{\ell \in R_i} x_{\ell r} \text{ (number of } d_i\text{-subsets that contain } \ell) = n_{ri} \binom{n_i - 1}{d_i - 1}$$

$$\sum_{\text{all } D_j} s_{jr} s_{jt} = \sum_{\text{all } D_j} \sum_{\ell \in D_j} x_{\ell r} x_{\ell t} + \sum_{\text{all } D_j} \sum_{\ell \neq m \in D_j} x_{\ell r} x_{mt}$$

$$= \delta_{rt} \left[n_{ri} \binom{n_i - 1}{d_i - 1} + n_{ri}(n_{ri} - 1) \binom{n_i - 2}{d_i - 2} \right].$$

Inserting these expressions into (7.3.28) and (7.3.29) with $\boldsymbol{\beta} = \mathbf{0}$, we obtain the score vector and information matrix components

$$U_r(\mathbf{0}) = \sum_{i=1}^{k} \left(d_{ri} - d_i \frac{n_{ri}}{n_i} \right) \qquad r = 1, \ldots, m-1 \tag{7.3.30}$$

$$I_{rt}(\mathbf{0}) = \sum_{i=1}^{k} \frac{(n_i - d_i) d_i}{n_i (n_i - 1)} \left(n_{ri} \delta_{rt} - \frac{n_{ri} n_{ti}}{n_i} \right) \qquad r, t = 1, \ldots, m-1. \tag{7.3.31}$$

The statistic for testing equality of the distributions is

$$W = \mathbf{U}(\mathbf{0})' I(\mathbf{0})^{-1} \mathbf{U}(\mathbf{0}), \tag{7.3.32}$$

which is asymptotically $\chi^2_{(m-1)}$ when the hypothesis of equality is true. Note that (7.3.30) has the same form as the test statistic (7.1.18) based on the continuous PH model, though in the case of (7.1.18) the terms refer to the times t_i for the individuals in the sample rather than to time intervals I_i. It is possible to think of the continuous-time case as giving a sequence of very short intervals that bracket the t_i, and early derivations of the test based on (7.1.18) were made on that basis. Note further that the information matrix (7.3.31) for the grouped data problem is the form (7.1.21) proposed for use in the continuous case when there are numerous ties among the lifetimes.

The tests of distributional equality based on the grouped PH and on the logistic model have good power to detect alternatives in which any two conditional survival probabilities p_{ri}, p_{ti} in (7.3.18) are related roughly as in (7.3.20) or (7.3.21). They can have poor power in settings where survival or hazard functions for the different distributions cross. More powerful tests can be devised for such settings by considering time-varying covariate effects in models of the form (7.3.6) or (7.3.7); this is analogous to the approach taken in Example 7.1.2 of Section 7.1.5 for the case of continuous observation. Cook and Lawless (1991) consider such tests; see also Problem 7.12.

Example 7.3.1. To illustrate the tests for the equality of distributions, consider the data given in Table 7.5, which represent failure time (in weeks) for three types of electrical components subject to constant use. A total of 140 components are involved, with 42, 50, and 48 components of Types A, B, and C, respectively.

There are 128 failures and 12 censoring times; the number of withdrawals in the different time intervals is not shown in the table, but can be deduced from the numbers at risk and the numbers of failures for the intervals. To test for the equality of

Table 7.5. Grouped Data on Component Failures

	Total		Type A		Type B		Type C	
Interval	n_i	d_i	n_{1i}	d_{1i}	n_{2i}	d_{2i}	n_{3i}	d_{3i}
[0, 10)	140	21	42	4	50	6	48	11
[10, 20)	119	24	38	3	44	11	37	10
[20, 30)	94	25	35	3	32	10	27	12
[30, 40)	68	21	31	5	22	8	15	8
[40, 50)	44	16	26	6	12	6	6	4
[50, 60)	26	7	20	4	5	3	1	0
[60, 70)	17	5	15	3	1	1	1	1
[70, 80)	11	4	11	4	0	0	0	0
[80, ∞)	5	5	5	5	0	0	0	0

failure time distributions using the logistic score statistic (7.3.32) it is necessary to calculate $\mathbf{U} = [U_1(\mathbf{0}), U_2(\mathbf{0})]'$ and $\mathbf{I}(\mathbf{0})$, given in (7.3.30) and (7.3.31). The values are easily calculated as $\mathbf{U} = (-24.10, 8.32)'$ and

$$\mathbf{I}(\mathbf{0}) = \begin{pmatrix} 19.8157 & -11.3917 \\ -11.3917 & 19.1190 \end{pmatrix}.$$

Note that $k = 8$ in the expressions involved in (7.3.32), the ninth interval being $[80, \infty)$, wherein all remaining individuals must die. This gives the observed score statistic value $\mathbf{U}'\mathbf{I}(\mathbf{0})^{-1}\mathbf{U} = 32.74$, which gives a p-value of under 10^{-7} on $\chi^2_{(2)}$. As is pretty clear from a look at the data, there is strong evidence of a difference in the failure time distributions.

When intervals are long, tests based on (7.3.32) and those based on the grouped proportional hazards model can give slightly different results, but the conclusions emerging from the two tests will usually be similar. In the present situation we find from (7.3.25) and (7.3.26) that $\mathbf{U} = (U_1, U_2)' = (-28.08, 8.72)'$ and

$$\mathbf{V} = \begin{pmatrix} 26.0903 & -15.0919 \\ -15.0919 & 24.9656 \end{pmatrix},$$

which gives a score statistic value (7.2.27) of 33.70, very close to the value for the logistic model.

7.3.3 Discrete-Time Hazard-Based Models

In many settings where lifetimes are discrete, it is convenient and satisfactory to treat them as continuous. However, it is sometimes useful to consider T as a discrete variable, for example, when the number of observed values for T is small or when there are large numbers of tied or equal lifetimes. Time-transform models are not natural for discrete-time, and so there is a strong emphasis on the hazard function for modeling and analysis. Assume without loss of generality that the range of T is $\{0, 1, 2, \ldots\}$ and let $h(t|\mathbf{x}(t))$ denote a discrete-time hazard function for T, given a vector $\mathbf{x}(t)$ of possibly time-varying covariates. We restrict the discussion to external covariates, and so assume that for individual i the hazard function satisfies

$$h(t|\mathbf{x}(t)) = Pr(T = t|T \geq t, X), \tag{7.3.33}$$

where $X = \{\mathbf{x}(t), t \geq 0\}$ is the covariate history.

The survivor function corresponding to (7.3.33) is

$$Pr(T \geq t|X) = \prod_{u=0}^{t-1}\{1 - h(u|x(u))\}, \tag{7.3.34}$$

and so the contribution to the likelihood function from an individual i with time and status indicator (t_i, δ_i) is, by (2.2.12),

$$L_i = h(t_i|\mathbf{x}_i(t_i))^{\delta_i}\{1 - h(t_i|x_i(t_i))\}^{1-\delta_i}\prod_{u=0}^{t_i-1}\{1 - h(u|\mathbf{x}_i(u))\}. \tag{7.3.35}$$

Note that (7.3.35) has the same form as the likelihood for a set of $t_i + 1$ independent binary responses; in particular, we can write L_i as

$$L_i = \prod_{u=0}^{t_i} h(u|\mathbf{x}_i(u))^{dN_i(u)}[1 - h(u|\mathbf{x}_i(u))]^{1-dN_i(u)}, \tag{7.3.36}$$

where $dN_i(u) = I(t_i = u, \delta_i = 1)$. This shows that software for binary response regression can be utilized, provided of course that it accommodates response probabilities of the specified form $h(u|\mathbf{x}_i(u))$. As for the grouped data models of Section 7.3.1, it is necessary to create multiple observations $dN_i(u)$, $\mathbf{x}_i(\mathbf{u})$, $u = 0, 1, \ldots, t_i$ for each individual as input to the binary response software.

The discrete-time modeling framework is formally the same as for the grouped data setting; if we associate times, t, with intervals, I, in the former setting, then $h(t|\mathbf{x}(t))$ in (7.3.33) corresponds to conditional probabilities $1 - p_i(\mathbf{x}(i))$ in the grouped data case, where $\mathbf{x}(i)$ indicates that $\mathbf{x}$ in (7.3.1) may vary from interval to interval. It is convenient to use g.l.m.'s analogous to (7.3.6) and (7.3.7) for discrete-time modeling, because discrete-time hazard functions are probabilities that must take values in (0, 1). The discrete-time models that correspond to (7.3.6) and (7.3.7) are, respectively,

$$h(t|\mathbf{x}(t)) = 1 - \exp\left\{-e^{\alpha(t)+\beta'\mathbf{x}(t)}\right\} \tag{7.3.37}$$

and

$$h(t|\mathbf{x}(t)) = \frac{e^{\alpha(t)+\beta'\mathbf{x}(t)}}{1 + e^{\alpha(t)+\beta'\mathbf{x}(t)}}. \tag{7.3.38}$$

Unlike the case of grouped data, where the number of intervals is generally rather small, the general discrete case may involve a large number of t-values, and so it is not usually feasible to leave $\alpha(t)$ in (7.3.37), (7.3.38), and other similar models completely arbitrary. One option is to specify $\alpha(t)$ parametrically; if this is done additively, as $\alpha(t) = \gamma_0 + \gamma_1 g_1(t) + \cdots + \gamma_k g_k(t)$ for specified functions $g_j(t)$, then models like the preceding ones take the convenient form

$$h(t|\mathbf{x}(t)) = \psi[\boldsymbol{\gamma}'\mathbf{z}(t) + \boldsymbol{\beta}'\mathbf{x}(t)], \tag{7.3.39}$$

where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_k)'$, $\mathbf{z}(t) = (1, g_1(t), \ldots, g_k(t))'$, and ψ is a function mapping $(-\infty, \infty)$ to $(0, 1)$. It is easy to use g.l.m. software with such models.

A second option is to leave $\alpha(t)$ arbitrary, but to impose some form of smoothing, along the lines of the hazard-function estimation in Section 3.4. Some g.l.m. software has the capability to fit models like (7.3.37) or (7.3.38) with this approach. The Bibliographic Notes provide a few references on discrete-time modeling.

7.4 SEMIPARAMETRIC MAXIMUM LIKELIHOOD

Semiparametric models can often be addressed by more or less standard maximum likelihood procedures. Although complex mathematical issues arise in a rigorous treatment, practical implementations of methodology are in many cases straightforward. In this section we examine maximum likelihood for multiplicative hazards models.

7.4.1 Estimation from Continuous Observation

Consider the PH model (7.1.1) with fixed covariate vector $\mathbf{x}$. The likelihood function from a fully observed censored random sample of n lifetimes, conditional on covariate values $\mathbf{x}_1, \ldots, \mathbf{x}_n$ was given in (7.1.27). Written in terms of the parameters $\boldsymbol{\beta}$ and $H_0 = \{H_0(t), t \geq 0\}$, this is

$$L(\boldsymbol{\beta}, H_0) = \prod_{i=1}^{n} \left[dH_0(t_i) e^{\boldsymbol{\beta}'\mathbf{x}_i} \right]^{\delta_i} \exp\left[-H_0(t_i) e^{\boldsymbol{\beta}'\mathbf{x}_i} \right]. \tag{7.4.1}$$

The objective of semiparametric maximum likelihood is to maximize $L(\boldsymbol{\beta}, H_0)$ jointly with respect to $\boldsymbol{\beta}$ and H_0 and, insofar as possible, to apply standard methodology from parametric likelihood theory, as presented in Appendix C. Mathematical and procedural issues arise, however, because we want to conceive of $H_0(t)$ as a continuous function, but rigorous analysis (e.g., Kiefer and Wolfowitz 1956) shows that there is no unique definition of likelihood for functional parameters such as H_0. It is clear that if we extend $H_0(t)$ to allow both discrete and continuous functions, then $L(\boldsymbol{\beta}, H_0)$ is always maximized by a discrete function, because for continuous models $dH_0(t) = h_0(t)\,dt \to 0$ as $dt \to 0$. The problem is that for discrete models (7.4.1) is not quite the correct expression for the likelihood function, since $S_0(t)$ in (7.1.27) is not equal to $\exp[-H_0(t)]$; see (1.2.8). Furthermore, there is no unique way to discretize the continuous model (7.1.1). Note that (7.1.1) and (7.1.2) cannot hold simultaneously for a discrete-time model.

Fortunately, it turns out there are practical ways to proceed. If we approximate the model (7.1.1) ever more closely through a sequence of models with finite parameter spaces, then under fairly mild conditions we can apply standard maximum likelihood methodology to make inferences about $\boldsymbol{\beta}$, $H_0(t)$, and $S_0(t)$. By choosing different approximating sequences we get slightly different procedures that, however, agree more and more closely as sample size increases.

We can use discrete-lifetime models to approximate (7.1.1), but it is a little simpler to use piecewise-continuous models. Consider the model discussed in Example 6.5.1, in which $h_0(t)$ is piecewise constant and $H_0(t)$ is piecewise linear. That is, for a specified set of cut points $0 = a_0 < a_1 < \cdots < a_k = \infty$ we have

$$h_0(t) = \alpha_j \qquad a_{j-1} \leq t < a_j \tag{7.4.2}$$

$$H_0(t) = \sum_{j=1}^{k} \alpha_j \Delta_j(t), \tag{7.4.3}$$

where (see (1.3.26))

$$\Delta_j(t) = \int_{a_{j-1}}^{a_j} I(t \geq u)\,du.$$

Maximum likelihood for this model was developed in Example 6.5.1. Let us consider these results in the case where k becomes large; in doing this we will assume that a_{k-1} is fixed at some large value beyond which failures are essentially impossible, and that as k increases the values $a_j - a_{j-1}$ for $j = 1, \ldots, k-1$ approach 0. It follows from (6.5.11) that the profile likelihood function for $\boldsymbol{\beta}$ is proportional to

$$L_p(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ \frac{e^{\boldsymbol{\beta}'\mathbf{x}_i}}{\sum_{j=1}^{k} I(a_{j-1} \leq t_i < a_j) \sum_{\ell=1}^{n} \Delta_j(t_\ell) e^{\boldsymbol{\beta}'\mathbf{x}_\ell}} \right\}^{\delta_i}.$$

When $k \to \infty$ as specified, $L_p(\boldsymbol{\beta})$ normalized by a factor accounting for the decreasing interval widths approaches

$$\prod_{i=1}^{n} \left\{ \frac{e^{\boldsymbol{\beta}'\mathbf{x}_i}}{\sum_{\ell=1}^{n} Y_\ell(t_i) e^{\boldsymbol{\beta}'\mathbf{x}_\ell}} \right\}^{\delta_i}, \tag{7.4.4}$$

which is the Cox likelihood function (7.1.5). Thus, as claimed in Section 7.1.3, (7.1.5) can be obtained as a profile likelihood or, more specifically, as a limiting profile likelihood for a sequence of models approaching (7.1.1).

Note in addition that the m.l.e. of $H_0(t)$ is, by (6.5.10) and (7.4.3),

$$\hat{H}_0(t) = \sum_{j=1}^{k} \left\{ \frac{d_j \Delta_j(t)}{\sum_{\ell=1}^{n} \Delta_j(t_\ell) e^{\hat{\boldsymbol{\beta}}'\mathbf{x}_\ell}} \right\},$$

where $d_j = \sum_i \delta_i I(a_{j-1} \leq t_i < a_j)$ is the number of lifetimes in $[a_{j-1}, a_j)$. In the limit as k increases, this approaches

$$\int_0^t \frac{dN_\cdot(u)}{\sum_{\ell=1}^{n} Y_\ell(u) e^{\hat{\boldsymbol{\beta}}'\mathbf{x}_\ell}} = \sum_{i:t_i \leq t} \left\{ \frac{\delta_i}{\sum_{\ell=1}^{n} Y_\ell(t_i) e^{\hat{\boldsymbol{\beta}}'\mathbf{x}_\ell}} \right\}, \tag{7.4.5}$$

which is the Breslow or generalized Nelson–Aalen estimate (7.1.32). The preceding development also goes through when covariates $\mathbf{x}_\ell(t)$ are time-varying. Thus, we can produce the estimates for the multiplicative models given in Section 7.1 as limits of estimates for a sequence of approximating models. Since the approximating models can approach the purely semiparametric model (7.1.1) arbitrarily closely, we can think of this maximum likelihood treatment as essentially nonparametric.

Note that asymptotic properties of these types of procedures need to be determined. For the PH model and the approach just considered, the limiting profile

likelihood for $\boldsymbol{\beta}$ and the estimate of $H_0(t)$ coincide with those from the partial likelihood–martingale analysis of Section 7.1.3, and therefore their properties have been established. Moreover, variance estimates for $\hat{H}_0(t)$, $\hat{S}_0(t)$, and $\hat{S}(t|x) = \exp[-\hat{H}_0(t)\exp(\hat{\boldsymbol{\beta}}'\mathbf{x})]$ obtained from the information matrix $I(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ in Example 6.5.1 according to standard maximum likelihood large-sample theory approach the expressions given in Section 7.1.5 as k increases and interval lengths become small. More generally, such properties are not always known. However, it is usually a satisfactory practical procedure to use standard parametric likelihood methods with an approximating model that has a moderate value of k. Experience indicates that reduction of the grid fineness beyond a certain point in models like (7.1.1) combined with (7.4.2) produces little change in inferences.

Different families of approximating models generally lead to slightly different procedures in finite samples. Instead of models with piecewise-constant hazard (7.4.2), we could, for example, consider models with piecewise-constant probability density functions; this is slightly less tractable than using (7.4.2). Another approach is to use discrete approximating models where it is assumed that lifetime T takes on values in some discrete set $\{a_1, a_2, \ldots\}$, which must include all the observed failure times. An idea that has been considered with fixed covariates is to retain the PH survivor function relationship (7.1.2),

$$S(t|\mathbf{x}) = S_0(t)^{\exp(\boldsymbol{\beta}'\mathbf{x})},$$

where $S_0(t)$ is now a left-continuous step function with jumps possible only at points $a_1, a_2, \ldots$. The likelihood function from a censored random sample of lifetimes is, by (2.2.14),

$$L(\boldsymbol{\beta}, S_0) = \prod_{i=1}^{n} [S(t_i|\mathbf{x}_i) - S(t_i + |\mathbf{x}_i)]^{\delta_i} S(t_i + |\mathbf{x}_i)^{1-\delta_i}. \tag{7.4.6}$$

This likelihood involves $S_0(t)$ only at the observed times $t_1, \ldots, t_n$, so maximizing it will give estimates $\hat{S}_0(t)$ that jump only at these times. In fact, $\hat{S}_0(t)$ has jumps only at the observed lifetimes, $\{t_i : \delta_i = 1\}$. The estimate $\hat{S}_0(t, \boldsymbol{\beta})$ maximizing $L(\boldsymbol{\beta}, S_0)$ with $\boldsymbol{\beta}$ fixed has a closed form only if there are no tied lifetimes, so it is only in this case that the profile likelihood function $L_p(\boldsymbol{\beta})$ has a closed form. Problem 7.14 considers this approach.

It turns out that if we apply an approach similar to that in the preceding paragraph to the likelihood (7.4.1) expressed in terms of H_0 and $\boldsymbol{\beta}$, then closed-form results are obtained. In fact, the estimate of $H_0(t)$ is precisely (7.4.5), and the profile likelihood for $\boldsymbol{\beta}$ is (7.4.4). To see this, associate parameters $\alpha_j = dH_0(t_{(j)})$ with the k distinct times $t_{(1)} < \cdots < t_{(k)}$ at which failures occur. It is then easy to show that (7.4.1) is maximized with $\boldsymbol{\beta}$ fixed when

$$d\tilde{H}_0(t_{(j)}) = \frac{dN_\cdot(t_{(j)})}{\sum_{\ell=1}^{n} Y_\ell(t_{(j)}) e^{\boldsymbol{\beta}'\mathbf{x}_\ell}},$$

and that $L(\boldsymbol{\beta}, \tilde{H}_0)$ reduces to (7.4.4). As previously noted, this approach is not exactly proper, since with a discrete-time model, the likelihood is not exactly (7.4.1). However, as the sample size increases, the difference between (7.4.1) and the proper likelihood becomes negligible.

7.4.2 Estimation from Incomplete Data

An advantage of semiparametric maximum likelihood is that it can be applied when data features such as interval censoring or arbitrary truncation occur, whereas the partial likelihood and martingale methods of Section 7.1.3 break down. However, problems with implementation and theoretical properties can arise with some approaches. For example, in the case of interval-censored data and an underlying continuous PH model (7.1.1), the likelihood function takes the form

$$L(\boldsymbol{\beta}, S_0) = \prod_{i=1}^{n} [S_0(L_i)^{\exp(\boldsymbol{\beta}'\mathbf{x}_i)} - S_0(R_i)^{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}], \tag{7.4.7}$$

where the information available is that $L_i \le T_i < R_i$. One approach is to identify parameters $\alpha_1, \ldots, \alpha_k$ with the values of $S_0(t)$ at the distinct values in the set $\{L_i, R_i, \quad i = 1, \ldots, n\}$ and to maximize (7.4.7) to obtain $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$. The parameter values α_j are ordered, and the global maximum can be difficult to find. In addition, the asymptotic properties of $\hat{\boldsymbol{\beta}}$ and the resulting $\hat{S}_0(t)$ are not completely known, though it appears that inferences about $\boldsymbol{\beta}$ may be based on its profile likelihood function in the standard parametric way.

In cases where the number of distinct intervals $[L_i, R_i]$ is fixed and not large, both numerical implementation and theoretical results are straightforward with the preceding approach; we are then dealing with a finite parameter situation. For example, in the case of current status observation, as in Example 4.3.2, the likelihood function takes the form

$$L(\boldsymbol{\beta}, S_0) = \prod_{i=1}^{n} \left[1 - S_0(C_i)^{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}\right]^{\delta_i} \left[S_0(C_i)^{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}\right]^{1-\delta_i}, \tag{7.4.8}$$

where C_i is the inspection time for individual i and $\delta_i = I(T_i \le C_i)$. If there are only finitely many distinct times $C_1, \ldots, C_k$ no matter how large n is, then (7.4.8) becomes $L(\boldsymbol{\beta}, \boldsymbol{\alpha})$, where $\alpha_j = S_0(C_j)$ for $j = 1, \ldots, k$. The baseline survivor function is estimable only at the times $C_1, \ldots, C_k$, and the problem is parametric. To consider pure semiparametric estimation, we require data for which the C_i become dense on some interval $(0, \tau)$ as n increases. In that case only can we consider nonparametric estimation of a continuous $S(t)$ for $0 \le t \le \tau$. Theoretical treatments of these types of problems are rather involved (e.g., Huang 1996).

A better practical approach for interval-censored data when the number of observational intervals $[L_i, R_i]$ is large is to employ piecewise-continuous models; the same is true for data involving different truncation patterns. Regression splines or piecewise-constant forms (7.4.2) for baseline hazard functions are both flexible and

reasonably easy to handle computationally. Problems can arise if the number of parameters is large, or if the pieces are not oriented suitably relative to the data. For example, some of the $\hat{\alpha}_j$ in (7.4.2) may equal zero, and it may not be possible to base inferences on normal approximations for $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$. However, once a moderate number (usually less than 10) of pieces is used, further increases in the number have little effect on estimation of $\boldsymbol{\beta}$ or $S_0(t)$; this means that inference can rely on ordinary finite-dimensional maximum likelihood procedures. Carstensen (1996) provides a good discussion of PH and additive hazards models using piecewise-constant baseline hazard functions.

BIBLIOGRAPHIC NOTES

The semiparametric proportional hazards models and analogous multiplicative models involving time-varying covariates were introduced by Cox (1972a), and rapidly became standard bases for analysis, particularly of medical, epidemiological, and demographic data. In the discussion following Cox's paper and in subsequent articles the validity of the likelihood (7.1.3) was discussed. Kalbfleisch and Prentice (1973) provided the rank-based marginal likelihood justification in the case of complete or Type 2 censored data, and Breslow (1974) and Holford (1976) gave heuristic maximum likelihood justifications. The partial likelihood justification introduced in Cox (1972a) was placed on a firmer theoretical basis by Cox (1975). More rigorous examinations of semiparametric maximum likelihood procedures and the connection between the Cox likelihood (7.1.3) and profile likelihood functions were given by Bailey (1983, 1984), Jacobsen (1984), and Johansen (1983). Estimation of the baseline cumulative hazard function $H_0(t)$ and survivor function $S_0(t)$ was addressed by Breslow (1974), who gave (7.1.32), and by Kalbfleisch and Prentice (1973); see also Tsiatis (1981a), Bailey (1983), Jacobsen (1984), and Link (1984).

Rigorous developments of asymptotic properties for the Cox likelihoods (7.1.5) and (7.1.42) and associated estimates for $H_0(t)$ or $S_0(t)$ are given by Tsiatis (1981a), Andersen and Gill (1982), Naes (1982), Bailey (1983), Prentice and Self (1983), and others. These papers take a variety of approaches, ranging from martingale-counting process methods to approaches that represent score functions as sums of independent components. Early discussions of issues associated with time-varying covariates can be found in Cox (1975), Efron (1977), Kalbfleisch and MacKay (1978), and Kalbfleisch and Prentice (1980). Peace and Flora (1978), Lee et al. (1983), and others studied finite sample behavior for tests and estimation via simulation. The books by Kalbfleisch and Prentice (1980, 2002), Fleming and Harrington (1991), and Andersen et al. (1993) contain extensive discussions of the multiplicative model, including questions of efficiency for semiparametric inference.

Two sample tests based on the Cox model, as in Section 7.1.2, were considered by Cox (1972a). Mantel (1966) first proposed the tests, following work by Mantel and Haenszel (1959), and they were also developed by Peto and Peto (1972) and others as rank tests. The test based on (7.1.15) is sometimes referred to as either the Mantel–Haenszel test or the log rank test. Peto et al. (1976, 1977) give an extensive discussion of the design and analysis of randomized clinical trials in which these tests, and extensions involving stratification, feature prominently. Collett (1994, Ch. 9) discusses sample size requirements for two-sample log rank tests, and provides references. Two- and m-sample tests are considered further in Chapter 8.

Diagnostic checks for multiplicative models have been discussed in many papers. Early discussion of exponential or cumulative hazard residuals (7.1.51) and of graphical methods were given by Kay (1977), Crowley and Hu (1977), Cox (1979), Kalbfleisch and Prentice (1980, Ch. 4), and Lagakos (1981). Lagakos (1981) and Crowley and Storer (1983) warned against the use of exponential residuals from semiparametric models like (7.1.1) in probability plots. Barlow and Prentice (1988) and Therneau et al. (1990) proposed martingale residuals. Goodness-of-fit tests for multiplicative models were considered by Andersen (1982), Kay (1984), and Nagelkerke et al. (1984). Kalbfleisch and McIntosh (1977), Schoenfeld (1982), Moreau et al. (1985, 1986), Gore et al. (1984), O'Quigley and Pessione (1989), Chappell (1992), and others considered tests of proportional hazards based on model expansion with time-varying covariates. Grambsch and Therneau (1994) and Lin et al. (1993) developed tests based on residuals. Deletion measures such as (7.1.57) for influence analysis were considered by Cain and Lange (1984), Reid and Crépeau (1985), and Storer and Crowley (1985). Pettitt and Bin Daud (1989) link case influence measures and residuals and discuss the usefulness of various plots.

The regression models for grouped data in Section 7.3 were first discussed by Cox (1972a) for the logistic case and Kalbfleisch and Prentice (1973) for the PH case. More extensive treatments were given by Thompson (1977) and Prentice and Gloeckler (1978), respectively. Pierce et al. (1979) considered alternative methods for discrete data. Aranda-Ordaz (1983) considered generalized models that include the discrete PH and logistic cases; Tibshirani and Ciampi (1983) considered another extended family. Two-sample tests were considered by Mantel and Haenszel (1959), but the formal model-based development in Section 7.3.2 started with Cox (1972a). Cook and Lawless (1991) consider two-sample tests that are effective in situations where survivor or hazard functions cross. Fahrmeir and Tutz (1994) and Tutz and Pritscher (1996) discuss the smoothing of parameters $\alpha(t)$ in discrete-time regression models such as (7.3.37) and (7.3.38).

Semiparametric maximum likelihood for multiplicative intensity models has been considered by Bailey (1983, 1984), Jacobsen (1984), and Johansen (1983). The use of piecewise-constant hazard functions was considered by Holford (1976) and Laird and Olivier (1981), among others. Friedman (1982) gave a rigorous discussion and considered the case where the number of pieces becomes large, as discussed in Section 7.4.1. Carstensen (1996) and Lindsey and Ryan (1998) provide illustrations of piecewise-constant hazards models with interval-censored data. Finkelstein (1986) considered semiparametric maximum likelihood; Huang (1996) provides a rigorous investigation for the special case of current-status observation. Lindsey and Ryan (1998), Kooperberg and Clarkson (1997), and Betensky et al. (1999) use smoothing and splines with interval-censored lifetimes.

Prentice (1986) considers settings where both lifetimes and censoring times may be unobserved by design for some individuals in a follow-up study, and introduces

what is known as a case-cohort design. Lawless et al. (1999) consider such problems in a broader but parametric context.

Theoretical aspects of semiparametric estimation that apply to the models in this chapter are discussed by Gill and Van der Vaart (1993), Bickel et al. (1993), Murphy and Van der Vaart (1999), and others. One aspect concerns the efficiency of estimators, for example, $\hat{\boldsymbol{\beta}}$ obtained from the Cox likelihood (7.1.3). Early studies of the efficiency of $\hat{\boldsymbol{\beta}}$ were carried out by Kalbfleisch (1974), Efron (1977), Oakes (1977), and Kay (1979). Lee et al. (1983), Peace and Flora (1978), and others provided empirical studies of efficiency.

Semiparametric methods for additive hazards models have also been developed. Andersen et al. (1993, Sec. 7.4), Klein and Moeschberger (1997, Ch. 10), Lin and Ying (1997), and Oakes (2001, Sec. 5.6) may be consulted for methodology and references.

COMPUTATIONAL NOTES

Methodology for proportional or multiplicative hazards model is included in many software packages. In S-Plus the relevant functions include coxph and cox.zph, which were used in examples in this chapter. Various two- and m-sample tests are also available. Therneau and Grambsch (2000) provide detailed lifetime data illustrations involving S-Plus and SAS. Generalized linear model software for binary or multinomial responses can be used to deal with the grouped or discrete-lifetime models of Section 7.3. The likelihood methods in Section 7.4 for dealing with interval-censored or truncated data are parametric, and are conveniently implemented by using general-purpose optimization software.

PROBLEMS AND SUPPLEMENTS

7.1 Consider the information matrix $\mathbf{I}(\boldsymbol{\beta})$ defined by (7.1.9) in the case in which there is no censoring and no ties.

(a) Prove that the exact covariance matrix for $\mathbf{U}(\mathbf{0})$ is

$$E[\mathbf{I}(\mathbf{0})] = \left(\sum_{\ell=1}^{n}(e_{\ell,n}-1)^2/(n-1)\right)\mathbf{X}_c'\mathbf{X}_c \qquad (7.5.1)$$

where $\mathbf{X}_c'\mathbf{X}_c$ is the corrected sum of squares matrix with (r, s) entry

$$\sum_{i=1}^{n}(x_{ir}-\bar{x}_r)(x_{is}-\bar{x}_s) \qquad r, s = 1, \ldots, p$$

and

$$e_{\ell,n} = \sum_{i=1}^{\ell}(n-i+1)^{-1}.$$

(b) Specialize (7.5.1) to the two-sample problem discussed in Section 7.1.2 to obtain Var[$\mathbf{U}(\mathbf{0})$].

(c) Specialize (7.5.1) to the m-sample problem to obtain an alternative to (7.1.19) as the covariance matrix for $\mathbf{U}(\mathbf{0})$ in the noncensored no-ties situation.

(Section 7.1.2)

7.2 Consider the score function $\mathbf{U}(\boldsymbol{\beta})$ for the Cox partial likelihood, written in the form (7.1.29).

(a) Use the martingale–stochastic integral results (F13)–(F16) in Appendix F to obtain the covariance matrix for $\mathbf{U}(\boldsymbol{\beta})$ as

$$\sum_{i=1}^{n}\int_0^{\infty}[\mathbf{x}_i-\bar{\mathbf{x}}(t,\boldsymbol{\beta})][\mathbf{x}_i-\bar{\mathbf{x}}(t,\boldsymbol{\beta})]'Y_i(t)e^{\boldsymbol{\beta}'\mathbf{x}_i}\,dH_0(t). \qquad (7.5.2)$$

(b) Show that insertion of the estimator (7.1.33) for $dH_0(t)$ and $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ in (7.5.2) gives $I(\hat{\boldsymbol{\beta}})$, where $I(\boldsymbol{\beta})$ is the observed information matrix (7.1.9).

(Sections 7.1.3, 7.1.5)

7.3 *Matched pairs.* Consider n pairs of individuals, and suppose that two treatments A and B represented by an indicator covariate $x = 0$ or 1 are randomly assigned to the individuals in each pair, so that one individual gets treatment A and one treatment B. Suppose that response times T_{j1} and T_{j2} in the jth pair are independent, and that their hazard functions are

$$h_{j1}(t) = h_{0j}(t)e^{\beta x_{j1}}, \qquad h_{j2}(t) = h_{0j}(t)e^{\beta x_{j2}}.$$

For convenience, suppose that individual 1 in each pair is labeled as the one getting treatment A, so that $x_{j1} = 0$, $x_{j2} = 1$. Assume that all response times are uncensored.

(a) Use the stratified PH model of Section 7.1.6 to give a partial likelihood for the estimation of β. Show that this likelihood is equivalent to one based on binary observations $Y_j = I(T_{j2} > T_{j1})$, $j = 1, \ldots, n$.

(b) Give both score and Wald tests for the hypothesis $\beta = 0$, noting their equivalence to tests for the probability $p = Pr(Y_j = 1)$.

(c) Suppose that T_{j1} and T_{j2} are subject to the same potential censoring time, C_j. Discuss why the partial likelihood, which now is based on pairs with at least one of T_{j1} or T_{j2} uncensored, is still valid.

(d) The leukemia remission time data given in Example 1.1.7 actually arose in matched pairs, as described there. The data are given below, with the first time in each pair being for the subject on drug 6-MP and the second being for the subject on the Placebo treatment. Asterisks denote censoring times. Carry out a test of no treatment effect using the stratified model given earlier. Compare the result of this with a two-sample log rank test, based on

(7.1.15), that ignores the pairing. Discuss the validity of the latter test in this context.

(10,1)	(7,22)	(32*,3)	(23,12)	(22,8)	(16,17)
(16,2)	(34*,11)	(32*,8)	(25*,12)	(11*,2)	(20*,5)
(19*,4)	(6,15)	(17*,8)	(35*,23)	(6,5)	(13,11)
(9*,4)	(6*,1)	(10*,8)			

(e) This study was actually terminated early, after all Placebo subjects' remission had ended. Using the material in Section 2.2.2, explain why the paired analysis in part (d) is still valid under these conditions.

(Sections 7.1.6, 7.1.2; Holt and Prentice 1974; Kalbfleisch and Prentice 1980, Sec. 8.1)

7.4 The data below are survival times for patients with bile duct cancer who took part in a study to determine whether a combination of radiation treatment (R_0R_X) and the drug 5-fluorouracil (5-FU) prolonged survival (Fleming et al., 1980). Survival times, in days, are given for a group of patients given the radiation–drug therapy and for a control group of patients. Asterisks denote censored observations.

$R_0R_X + 5 - FU$	30, 67, 79*, 82*, 95, 148, 170, 171, 176, 193, 200, 221, 243, 261, 262, 263, 399, 414, 446, 446*, 464, 777
Control	57, 58, 74, 79, 89, 98, 101, 104, 110, 118, 125, 132, 154, 159, 188, 203, 257, 257, 431, 461, 497, 723, 747, 1313, 2636

(a) Plot the Kaplan–Meier estimates based on the two groups. Is the test based on (7.1.15) liable to be effective in this situation?

(b) Consider a multiplicative model with survivor function of the form (7.1.47) in Example 7.1.2, where $x = I$ (individual received R_0R_X+5-FU) and $g(t)$ is a specified function. Carry out a test of equality of the survival distributions for the Treatment ($x = 1$) and Control ($x = 0$) populations, using the score statistic (7.1.48) with $g(t) = t - 200$. For comparison, also carry out the test based on (7.1.15).

(Sections 7.1.2, 7.1.8)

7.5 Consider the adjusted exponential residuals $\hat{e}_i^{adj}$ of (7.1.52), in the case of models with fixed covariates. As stated in Section 7.1.9, these residuals are not suitable for PH model checks based on exponential probability plots. To help see why, consider the PH model (7.1.1) with no covariates and suppose there is no censoring in the data. Show that the values $\hat{e}_i^{adj}$ then consist of the values $(\alpha_{(1)}, \ldots, \alpha_{(n)})$ in some order, where

$$\alpha_{(i)} = \sum_{j=1}^{i} \frac{1}{n-j+1} \qquad i = 1, \ldots, n.$$

Now show using Theorem 4.1.1 that $\alpha_{(i)}$ is the expected value of the ith smallest observation in a random sample of n standard exponential random variables.

Thus, an exponential probability plot of the adjusted exponential residuals from the model with $\boldsymbol{\beta} = \mathbf{0}$ will automatically conform to the exponential distribution. Try to extend the discussion to the case of a model with a single indicator covariate.

(Section 7.1.9; Crowley and Storer 1983)

7.6 *Efficiency of partial likelihood in a simple situation.* Suppose in the model (7.1.1) that $h_0(t) = \lambda$; that is, the distribution is actually exponential. Consider the case in which there is a single covariate x. Suppose that $t_1, \ldots, t_n$ are observed lifetimes in a random sample of n, corresponding to covariates $x_1, \ldots, x_n$; assume that the x_i are centered so that $\sum x_i = 0$.

(a) Show that the joint p.d.f. of $a_2, \ldots, a_n$, where $a_i = t_i/t_1$, is

$$(n-1)!\left(\sum_{i=1}^{n} a_i e^{\beta x_i}\right)^{-n} \qquad a_i > 0,$$

where $a_1 = 1$. This can be used for inference about β when λ is unknown. Determine the expected information $I_1(\beta)$ based on this distribution and show that

$$I_1(0) = \frac{n}{n+1}\sum_{i=1}^{n} x_i^2 = \frac{n\mu_2}{n+1}.$$

(b) Consider the partial likelihood (7.1.5). Determine the expected information $I_2(0)$, noting that $t_1, \ldots, t_n$ are independent and identically distributed (i.i.d.) when $\beta = 0$, whereupon each of the $n!$ possible rank vectors has probability $(n!)^{-1}$. Show that

$$I_2(0) = \frac{n\mu_2}{n-1}\sum_{i=1}^{n} \frac{n-i}{n-i+1}.$$

(c) Examine $I_2(0)/I_1(0)$ for various values of n; this represents the efficiency of the partial likelihood method at $\beta = 0$.

(Section 7.1.1; Kalbfleisch, 1974)

7.7 Consider the electrical insulation failure time data of Example 5.4.2.

(a) Assuming that the failure voltages for the two types of insulation have distributions with proportional hazard functions, use the methods of Section 7.1.2 to test that the two distributions are identical.

(b) Obtain a confidence interval for $\delta = \exp(\beta)$ in the PH model (7.1.1), which you used implicitly in part (a). Compare this with the confidence interval for δ that is obtained under a Weibull model (see Problem 5.13).

(Sections 7.1.1, 7.1.2, 7.2)

7.8 Consider the data on survival times for patients with multiple myeloma, discussed in Problem 6.9. Assess the relationship of the five covariates to survival time by using the semiparametric PH model (7.1.1), supplemented by appropriate checks on assumptions. Compare your conclusions with those based on the parametric models of Problem 6.9.

(Sections 7.1, 7.2)

7.9 Wei et al. (1989) discussed data on the times to recurrence of bladder cancer in 85 patients. The data were described in Problem 6.16 and examined there using AFT models. Some individuals had several recurrences over a period of time, but we consider only the time to first recurrence, measured from entry to the study in question. The full data set is given in Wei et al. (1989) and is contained in the S-Plus data frame "bladder." Covariates include, as in Problem 6.16,

rx: Treatment group (1 = Placebo, 2 = Drug Thiotepa)
number: The number of tumors present at initial diagnosis
size: The size of the largest initial tumor.

Investigate the relationship of these covariates to the time to first recurrence using multiplicative hazards models; include appropriate model checks.

(Sections 7.1, 7.2)

7.10 Abrahamowicz et al. (1996) presented survival data on 87 persons with lupus nephritis who underwent a renal biopsy during the years 1967–1983. Individuals were followed until death or the end of 1990; by that time 35 of the patients had died. Abrahamowicz et al. discuss the effect of a covariate termed "duration," which is the duration of the individual's untreated renal disease prior to biopsy. The data are discussed in Appendix G, and are available electronically.

Use multiplicative hazards models to explore the relationship of duration to survival time, with special attention to the possibility that the effect of the covariate on the hazard function may change over time.

(Sections 7.1, 7.2)

7.11 *Examination of trend across several distributions.* Suppose there are several lifetime distributions and that it is suspected there is a trend among the hazard functions for the distributions. Trends can be examined with the methods of this chapter. For example, suppose there are m distributions, corresponding to levels $0, d_1, \ldots, d_{m-1}$ of a covariate d, and that it is suspected that hazard functions for the m distributions are proportional and vary monotonically with d.

(a) Consider the model (7.1.1) in which distribution j has hazard function

$$h_j(t) = e^{\beta d_j} h_0(t) \qquad j = 0, 1, \ldots, m-1. \tag{7.5.3}$$

Derive the score function statistic $U(0)^2 I(0)^{-1}$ for testing $\beta = 0$ from (7.1.8) and (7.1.9). Show that $U(0)$ and $I(0)$ can be expressed as $\mathbf{d}'\mathbf{U}^*(\mathbf{0})$ and $\mathbf{d}'\mathbf{I}^*(\mathbf{0})\mathbf{d}$, respectively, where $\mathbf{d}' = (d_1, \ldots, d_{m-1})$ and $\mathbf{U}^*(\mathbf{0})$ and $\mathbf{I}^*(\mathbf{0})$ are the score vector (7.1.18) and information matrix (7.1.19) used in the m-sample test of Section 7.1.2. Tarone (1975) has shown that $X_D^2 = \mathbf{U}^*(\mathbf{0})'\mathbf{I}^*(\mathbf{0})^{-1}\mathbf{U}^*(\mathbf{0}) - U(0)^2 I(0)^{-1}$ is asymptotically $\chi^2_{(m-2)}$ under the hypothesis that the m distributions are equal. This can be used to test for departures from trend.

(b) Consider the insulation data of Examples 1.1.5, 5.4.1, and 6.3.2. Let $d_j = \log(v_j/32)$. Carry out a test for equality of the lifetime distributions at the seven voltage levels, using (7.1.20). Also carry out a test for trend by considering the model (7.5.3) and testing $\beta = 0$ as suggested in part (a). Finally, use the statistic X_D^2 of part (a) to examine departures from the trend represented by (7.5.2). Compare your results with those of the fully parametric analysis of Example 5.4.1.

(Section 7.1; Tarone 1975)

7.12 *Two-sample tests with grouped data.* Consider tests for the equality of two lifetime distributions, based on grouped data as in Section 7.3.2. Define p_{ri} $(r = 1, 2; \quad i = 1, \ldots, k)$ as in (7.3.19) and let $q_{ri} = 1 - p_{ri}$. Let $f(q)$, $0 < q < 1$ be a continuous, monotonic, twice-differentiable function mapping $(0, 1)$ onto $(-\infty, \infty)$, and suppose that q_{1i} and q_{2i} are related by

$$f(q_{2i}) = f(q_{1i}) + \beta \qquad (i = 1, \ldots, k). \tag{7.5.4}$$

Note that the choices $f(q) = \log[q/(1-q)]$ and $f(q) = \log[-\log(1-q)]$ give the grouped logistic and PH models given by (7.3.21) and (7.3.20), respectively. Testing $\beta = 0$ amounts to testing the equality of the two distributions, under the assumed model (7.5.4)

(a) Derive the partial score test of $\beta = 0$ for the model (7.5.4), as described in Section 7.3.2. Show that the score statistic (7.3.23) and variance estimate (7.3.24) are, respectively,

$$U_1 = \sum_{i=1}^{k} \hat{g}_i' \hat{u}_i$$

$$V_1 = \sum_{i=1}^{k} (\hat{g}_i')^2 \hat{v}_i,$$

where

$$\hat{u}_i = \frac{d_{2i} - \hat{q}_i}{\hat{q}_i \hat{p}_i}, \qquad \hat{q}_i = 1 - \hat{p}_i = \frac{d_i}{n_i},$$

$$\hat{v}_i = \frac{n_{1i} n_{2i}}{n_i \hat{q}_i \hat{p}_i}, \qquad \hat{g}_i' = g'[f(\hat{q}_i)], \qquad g(\theta) = f^{-1}(\theta).$$

(b) Let $A_i = (1, r_i)'$ be a 2×1 vector of scores associated with the intervals $i = 1, \ldots, k$, and consider the family of models

Table 7.6. Survival Data for Cancer Patients Classified by Two Factors

	A_1B_1		A_1B_2		A_2B_1		A_2B_2	
Interval	n_{1l}	d_{1l}	n_{2l}	d_{2l}	n_{3l}	d_{3l}	n_{4l}	d_{4l}
[0, 3)	75	15	87	18	59	12	64	9
[3, 6)	60	3	69	14	46	7	55	8
[6, 9)	56	14	55	8	38	8	47	11
[9, 12)	41	17	47	16	30	7	34	8
[12, 15)	22	7	30	11	21	9	25	5
[15, 18)	13	6	19	12	10	5	18	9
[18, 21)	5	4	7	4	4	2	8	3
[21, ∞)	1	1	2	2	2	2	4	4

$$f(q_{2i}) = f(q_{1i}) + \mathbf{A}_i'\boldsymbol{\beta},$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. Derive the partial score test of $\boldsymbol{\beta} = \mathbf{0}$ as described in Section 7.3.2.

(Section 7.3.2, Cook and Lawless 1991)

7.13 The data in Table 7.6 show the survival experience of a certain type of cancer patient in life table form. Time is given in months from treatment. The 285 patients fall into four categories, which correspond to two levels for each of two factors, A and B. The data give the survival experience for each of the four patient groups A_1B_1, A_1B_2, A_2B_1, and A_2B_2.

Examine whether the survival distributions for the four patient groups might be the same. Assess the effect of factors A and B on the distributions.

(Section 7.3)

7.14 Consider the observed semiparametric likelihood function $L(\boldsymbol{\beta}, S_0)$ given by (7.4.6).

(a) Show that when this is maximized with respect to $\boldsymbol{\beta}$ and $S_0(t)$, the estimate $\hat{S}_0(t)$ has jumps only at the observed lifetimes.

(b) Find the function $\tilde{S}_0(t)$ that maximizes $L(\boldsymbol{\beta}, S_0)$ when $\boldsymbol{\beta}$ is fixed, by first rewriting the likelihood in terms of $\boldsymbol{\beta}$ and the parameters $\alpha_1, \ldots, \alpha_k$, where

$$\alpha_j = S_0(t_{(j)}+)/S_0(t_{(j)}) \qquad j = 1, \ldots, k$$

and $t_{(1)} < \cdots < t_{(k)}$ are the distinct observed lifetimes. Show that if there is only a single lifetime equal to $t_{(i)}$ then $\tilde{\alpha}_i$ satisfies the equation

$$\tilde{\alpha}_i \exp(\boldsymbol{\beta}'\mathbf{x}_{(i)}) = 1 - e^{\boldsymbol{\beta}'\mathbf{x}_{(i)}} \Big/ \sum_{\ell \in R(t_{(i)})} e^{\boldsymbol{\beta}'\mathbf{x}_\ell},$$

but that if there is more than one death at $t_{(i)}$, there is no closed-form solution.

(c) In the case where there are no tied lifetimes, compare the profile likelihood function for $\boldsymbol{\beta}$ with (7.1.5). For the two-sample data below, compare the two log profile likelihood functions for the model (7.1.1) with $\boldsymbol{\beta}'\mathbf{x} = \beta x$ ($x = 0, 1$) by plotting them on the same graph. Also plot the estimate $\tilde{S}_0(t)$ described in part (b) and the estimate (7.1.34) on the same graph.

Sample 1 ($x = 0$): 60, 204, 29, 48, 366*, 26, 10, 255, 18, 74
Sample 2 ($x = 1$): 366*, 103, 364, 26, 81, 62, 84, 366*, 366*, 366*

(*Note:* Asterisks denote censoring times.)

(Section 7.4; Kalbfleisch and Prentice 1973; Bailey 1984)

7.15 *Piecewise-constant hazards and semiparametric inference.* The information matrix $I(\boldsymbol{\alpha}, \boldsymbol{\beta})$ for the piecewise-constant hazards model in Section 7.4.1 is given in Example 6.5.1.

(a) Consider the covariance matrix $I(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})^{-1}$ and obtain an asymptotic variance estimate for $\hat{H}_0(t)$ of (7.4.3). Show that as the number of intervals k increases and interval lengths shrink to zero, this variance estimate approaches (7.1.37).

(b) Extend the development to deal with the variance estimate for $\log[-\log \hat{S}(t|\mathbf{x})]$, as in (7.1.39).

(Sections 7.1.5, 7.4)

7.16 *Right-truncated lifetimes.* Right-truncated data were discussed in Sections 2.4, 3.5.2, and 4.3.3. In this setting the data consist of pairs $(t_i, v_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ where v_i is the truncation time, $t_i \le v_i$ is the observed lifetime, and $\mathbf{x}_i$ is a covariate vector for individual i. There is no censoring in this case. In the no-covariate case it was shown in Section 3.5.2 how to obtain a nonparametric estimate of the conditional distribution $P(T \le t | T \le v_{\max})$, where $v_{\max} = \max(v_i)$ is the largest truncation time in the data set. This was conveniently done by introducing the reverse time hazard function $h_{RT}(t) = f(t)/F(t)$ for lifetime T; we note that by reversing the time scale right truncation turns into left truncation.

(a) For the covariate case, let

$$h_{RT}(t|\mathbf{x}) = f(t|\mathbf{x})/F(t|\mathbf{x})$$

denote the reverse time hazard function, conditional on $\mathbf{x}$. Consider the multiplicative model

$$h_{RT}(t|\mathbf{x}) = h_{RT0}(t)\exp(\boldsymbol{\beta}'\mathbf{x}). \tag{7.5.5}$$

Motivate the likelihood function

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{e^{\boldsymbol{\beta}'\mathbf{x}_i}}{\sum_{\ell=1}^{n} Y_\ell(t_i) e^{\boldsymbol{\beta}'\mathbf{x}_\ell}} \right),$$

where $Y_\ell(t) = I(t_\ell \le t \le v_\ell)$, as a partial or other type of likelihood function.

(b) Show that

$$\frac{F(t|\mathbf{x})}{F(v|\mathbf{x})} = \exp\left\{-\int_t^v h_{RT}(u|\mathbf{x})\,du\right\} \qquad 0 \le t \le v,$$

and thus consider the effect of $\mathbf{x}$ on failure probabilities.

(Sections 7.1, 7.4; Kalbfleisch and Lawless 1991)

7.17 *A test for quasi-independent truncation.* For the setting in Problem 7.16, it is sometimes of interest to assess independence of the lifetime variable T and the right-truncation variable, V. Because only cases with $t \le v$ are observed, it is possible only to test what we might call quasi-independence. In the case with no covariates, this is expressed as

$$H: h_{RT}(t|v) = h_{RT}(t), \qquad 0 \le t \le v.$$

(a) By defining a model (7.5.5) with a covariate that is a function of v, develop a score test of H by considering a score test of $\boldsymbol{\beta} = \mathbf{0}$ in (7.5.5), as in Section 7.1.2.

(b) Indicate how to use a similar approach to test that the ordinary hazard function for T given a left-truncation time $U = u$, is quasi-independent of u.

(Section 7.1.7; Tsai 1990; Kalbfleisch and Lawless 1991)

7.18 *Sensitivity of PH models to covariate misspecification.* Suppose that the hazard function of a lifetime T given covariates x_1 and x_2 is of PH form (7.1.1):

$$h(t|x_1, x_2) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2). \qquad (7.5.6)$$

Suppose now that covariate x_2 is not observed, but has the conditional distribution function $G(x_2|x_1)$, given x_1, in the population observed.

(a) Obtain the hazard function $h(t|x_1)$ for T given x_1 by first obtaining the survivor function (s.f.) for T, given x_1,

$$S(t|x_1) = \int_{-\infty}^{\infty} \exp[-H_0(t)\exp(\beta_1 x_1 + \beta_2 x_2)]\,dG(x_2|x_1).$$

Show that $h(t|x_1)$ is not of PH form (7.1.1) in general.

(b) Show that $h(t|x_1)$ in part (a) is not of PH form even when $G(x_2|x_1) = G(x_2)$, that is, x_1 and x_2 are independent. Consider the implications of this when x_1 is a binary treatment indicator in a randomized clinical trial.

(c) Consider the randomized clinical trial involving the treatment rhDNase in Example 7.2.1. Discuss why PH analyses and estimates of treatment effect

using (7.1.1) with and without the fev covariate x_2 included are not strictly compatible. Assess whether a PH model with only the treatment covariate x_1 present is consistent with the data. What do you conclude?

(d) Use the results of Section 6.3.4 to discuss why accelerated failure time models are insensitive to this type of covariate misspecification.

(Sections 7.1, 7.2, 7.4; Struthers and Kalbfleisch 1986; Appendix C.2)

CHAPTER 8

Rank-Type and Other Semiparametric Procedures for Log-Location-Scale Models

Log-location-scale or accelerated failure time models were discussed extensively in Chapters 5 and 6. There, the emphasis was on fully parametric models that involved distributions such as the extreme value, logistic, and normal families. This chapter considers semiparametric methodology in which the location-scale format is used, but no particular family of distributions is assumed. Applications to accelerated failure time (AFT) regression models and to tests of distributional equality will be considered. This chapter can be regarded as the (log-) location-scale analog to Chapter 7, which dealt with semiparametric methods for proportional hazards models.

As in Chapter 6, we consider models for which the distribution of log-lifetime Y, given a vector of fixed covariates $\mathbf{x}$, is of the form

$$Y = \boldsymbol{\beta}'\mathbf{x} + Z, \tag{8.0.1}$$

where Z has a distribution that does not involve $\mathbf{x}$. The distribution of Z on $(-\infty, \infty)$ is here left unspecified; in this case, we do not include an intercept term in $\boldsymbol{\beta}'\mathbf{x}$, since it can be subsumed in the "error" term Z. Sometimes we may wish to assume that Z is unspecified aside from having $E(Z) = 0$; in that case we would include an intercept term in $\boldsymbol{\beta}'\mathbf{x}$. This, for example, is the standard approach when least-square methods are applied to (8.0.1). It is also sometimes convenient to express Z as bZ_1, where Z_1 has a specified mean and variance (usually $E(Z_1) = 0$, $\mathrm{Var}(Z_1) = 1$) and $b > 0$ is a scale parameter.

Two main approaches to semiparametric methods for location-scale models have been taken. One uses rank-based procedures (e.g., Hajek and Sidak 1967), and the other uses robust estimating functions. The primary focus in this chapter is on rank-based procedures, which are more readily adapted to handle censoring.

A major application of rank-based procedures is to tests for the equality of two or more distributions. Section 8.1 considers this topic and gives tests that are analogous to those of Section 7.1.2, developed under multiplicative hazards assumptions. In

fact, a close connection between the two types of tests can be established. Section 8.2 considers inference for $\boldsymbol{\beta}$ in the model (8.0.1), and an extension of this model to incorporate time-varying covariates.

8.1 RANK TESTS FOR COMPARING DISTRIBUTIONS

8.1.1 Linear Rank Tests for the m-Sample Problem

Tests for the equality of two or more lifetime distributions are often required. When it is not convenient or appropriate to adopt a parametric family of models within which to carry out tests (as is done in Section 5.4, for example), distribution-free methods can be used. Some such procedures were discussed in Section 7.1.2; here we consider linear rank tests based on the model (8.0.1). Interestingly, the log rank test of Section 7.1.2 can be obtained from (8.0.1). Some other rank tests also will be examined, and their connection to counting process test formulations in Chapter 7 demonstrated.

The basic ideas of rank tests and their extension to censored data are described later. Several books provide extended treatments of rank tests and of asymptotic properties of the censored data procedures in this section; see the Bibliographic Notes at the end of the chapter.

The framework used to develop the rank tests is one in which any two distributions are assumed to differ only with respect to location. That is, the two distributions are assumed to have probability density functions (p.d.f.'s) $g(y)$ and $g(y-\theta)$, and they are identical if and only if $\theta = 0$. The resulting tests are therefore good at detecting whether two or more distributions with the same general shape are different; they can be poor at detecting certain other types of differences.

The tests can be formulated in terms of the model (8.0.1). Suppose that the m distributions of interest have p.d.f.'s for log-lifetime Y of the forms

$$g_1(y) = g(y-\theta_1), \ldots, g_{m-1}(y) = g(y-\theta_{m-1}), g_m(y) = g(y), \qquad (8.1.1)$$

where $-\infty < y < \infty$ and $-\infty < \theta_j < \infty$ for the parameters $\theta_1, \ldots, \theta_{m-1}$. As in Section 7.1.2, we let $\mathbf{x} = (x_1, \ldots, x_{m-1})'$ be a vector of distribution indicator variables, defined so that individuals from the distributions $1, \ldots, m-1, m$ have $\mathbf{x}'$ vectors $(1, 0, \ldots, 0) \cdots (0, \ldots, 0, 1), (0, 0, \ldots, 0)$. With $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{m-1})'$, the regression model

$$f(y|\mathbf{x}) = g(y - \mathbf{x}'\boldsymbol{\theta}) \qquad (8.1.2)$$

then gives (8.1.1), and the equality of the m distributions is represented by the hypothesis $\boldsymbol{\theta} = \mathbf{0}$. We now consider rank tests for this hypothesis; we begin with the case of uncensored data and then we discuss adjustments to deal with censoring.

8.1.1.1 Tests with Uncensored Data

The construction of rank tests of $\boldsymbol{\theta} = \mathbf{0}$ will be described first for the case of uncensored data. Briefly, a rank test is one for which the test statistic is a function of the

ranks of the observations and not their actual values. Such a test is distribution-free in the sense that significance levels calculated from the distribution of the ranks are valid for arbitrary distributions. The power of a rank test depends on the alternative hypothesis and the underlying distribution, but tests can be selected to have good power against specific types of alternatives. In the present context let $y_1, \ldots, y_n$ be a sample from (8.1.2), selected as a set of independent random samples from each of distributions $1, \ldots, m$; let N_i be the number of observations from distribution i $(N_1 + \cdots + N_m = n)$. Let $\mathbf{r} = [(1), \ldots, (n)]$ denote the rank vector based on the y_i; that is, (i) is the label of the individual with the ith smallest y value. The ordered observations $y_{(1)} < \cdots < y_{(n)}$ are assumed to be distinct; this entails no loss of generality under a continuous model. Rank tests of $\boldsymbol{\theta} = \mathbf{0}$ can be constructed by considering a score test based on the distribution of $\mathbf{r}$. This approach is outlined here; a more detailed presentation can be found, for example, in Hajek and Sidak (1967).

The probability function of $\mathbf{r} = [(1), \ldots, (n)]$ is

$$p(\mathbf{r}; \boldsymbol{\theta}) = \int \cdots \int_A \prod_{j=1}^{n} g[y_{(j)} - \mathbf{x}'_{(j)}\boldsymbol{\theta}] dy_{(1)} \cdots dy_{(n)},$$

where A is the region $\{(y_{(1)}, \ldots, y_{(n)}) : -\infty < y_{(1)} < \cdots < y_{(n)} < \infty\}$ and $\mathbf{x}_{(j)}$ is the regression vector associated with (j). The first derivatives of the log-likelihood based on $p(\mathbf{r}; \boldsymbol{\theta})$ are thus

$$U_\ell(\boldsymbol{\theta}) = \frac{\partial \log p(\mathbf{r}; \boldsymbol{\theta})}{\partial \theta_\ell}$$

$$= \frac{-1}{p(\mathbf{r}; \boldsymbol{\theta})} \sum_{i=1}^{n} \int \cdots \int_A x_{(i)\ell} \frac{g'(y_{(i)} - \mathbf{x}'_{(i)}\boldsymbol{\theta})}{g(y_{(i)} - \mathbf{x}'_{(i)}\boldsymbol{\theta})} \prod_{j=1}^{n} g(y_{(j)} - \mathbf{x}'_{(j)}\boldsymbol{\theta})\, dy_{(1)} \cdots dy_{(n)}$$

$$\ell = 1, \ldots, m-1. \qquad (8.1.3)$$

A test of $\boldsymbol{\theta} = \mathbf{0}$ can be based on the score statistic

$$\mathbf{U}(\mathbf{0}) = [U_1(\mathbf{0}), \ldots, U_{m-1}(\mathbf{0})]'.$$

When $\boldsymbol{\theta} = \mathbf{0}$, all individuals have the same distribution and $p(\mathbf{r}; \mathbf{0}) = (n!)^{-1}$ for each possible rank vector $\mathbf{r}$. Thus

$$U_\ell(\mathbf{0}) = -n! \sum_{i=1}^{n} \int \cdots \int_A x_{(i)\ell} \frac{g'(y_{(i)})}{g(y_{(i)})} \prod_{j=1}^{n} g(y_{(j)})\, dy_{(1)} \cdots dy_{(n)}.$$

But $n! \prod_{j=1}^{n} g(y_{(j)})$ is the joint p.d.f. of $y_{(1)}, \ldots, y_{(n)}$ when $\boldsymbol{\theta} = \mathbf{0}$ [see (B15)], thus

$$U_\ell(\mathbf{0}) = \sum_{i=1}^{n} x_{(i)\ell} a_i \qquad \ell = 1, \ldots, m-1, \qquad (8.1.4)$$

where

$$\alpha_i = E\left(-\frac{g'(y_{(i)})}{g(y_{(i)})};\ \boldsymbol{\theta} = \mathbf{0}\right). \tag{8.1.5}$$

The α_i are called the scores associated with $y_{(1)}, \ldots, y_{(n)}$. Note that if one uses the likelihood function based on the actual observations $y_1, \ldots, y_n$, and not just their ranks, the score statistic at $\boldsymbol{\theta} = \mathbf{0}$ has components

$$\sum_{i=1}^{n} x_{i\ell}\left[\frac{-g'(y_i)}{g(y_i)}\right],$$

so that the effect of considering only the ranks is to replace $g'(y_i)/g(y_i)$ with a score that is a function of the rank of y_i. Specific tests are obtained by choosing a function $g(y)$ on which to base the scores. The choice of $g(y)$ affects the power of the tests, but not its distribution-free nature. We consider choices of scores after first developing some general results.

The mean and variance of $\mathbf{U}(\mathbf{0})$ can be derived by standard permutation theory arguments, since under H_0: $\boldsymbol{\theta} = \mathbf{0}$ all $n!$ possible rank vectors $\mathbf{r}$ are equally likely. The necessary formulas are given by the following result.

LEMMA 8.1.1. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be given vectors with $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$ and let $\alpha_1, \ldots, \alpha_n$ be given constants such that $\sum \alpha_i = 0$. Let $[(1), \ldots, (n)]$ be a random permutation of $(1, \ldots, n)$ and define

$$U_\ell = \sum_{i=1}^{n} x_{(i)\ell}\alpha_i \qquad \ell = 1, \ldots, p.$$

Then $E(U_\ell) = 0$ and

$$E(U_\ell U_s) = \left(\sum_{i=1}^{n} \alpha_i^2/(n-1)\right)\left(\sum_{i=1}^{n} x_{i\ell}x_{is} - n\bar{x}_\ell\bar{x}_s\right) \qquad \ell, s = 1, \ldots, p, \tag{8.1.6}$$

where $\bar{x}_\ell = \sum x_{i\ell}/n$ $(\ell = 1, \ldots, p)$.

Proof. First,

$$E(U_\ell) = \sum_{i=1}^{n} \alpha_i E_p(x_{(i)\ell}),$$

where E_p denotes expectation over the set of permutations of $1, 2, \ldots, n$. This gives

$$\sum_{i=1}^{n} \alpha_i\left(\sum_{j=1}^{n} x_{j\ell}\frac{1}{n}\right) = \left(\sum_{i=1}^{n} \alpha_i\right)\bar{x}_\ell = 0.$$

Also, for each ℓ and $s = 1, \ldots, p$,

$$\begin{aligned}
E(U_\ell U_s) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j E_p(x_{(i)\ell}x_{(j)s}) \\
&= \sum_{i=1}^{n} \alpha_i^2\left(\sum_{k=1}^{n} x_{k\ell}x_{ks}\frac{1}{n}\right) + \sum\sum_{i\neq j} \alpha_i\alpha_j\left(\sum\sum_{k\neq t} x_{k\ell}x_{ts}\frac{1}{n(n-1)}\right) \\
&= \left(\sum_{i=1}^{n} \alpha_i^2/n\right)\sum_{k=1}^{n} x_{k\ell}x_{ks} - \left(\sum_{i=1}^{n} \alpha_i^2/n(n-1)\right) \\
&\quad \times\left[\left(\sum_{k=1}^{n} x_{k\ell}\right)\left(\sum_{t=1}^{n} x_{ts}\right) - \sum_{k=1}^{n} x_{k\ell}x_{ks}\right] \\
&= \left(\sum_{i=1}^{n} \alpha_i^2/(n-1)\right)\left(\sum_{k=1}^{n} x_{k\ell}x_{ks} - n\bar{x}_\ell\bar{x}_s\right).
\end{aligned}$$

□

We observe that for the scores (8.1.5) $\sum \alpha_i = 0$, since

$$\sum \alpha_i = E\left(-\sum \frac{d}{dy}\log g(y_{(i)})\right) = E\left(-\sum \frac{d}{dy}\log g(y_i)\right) = 0.$$

The mean and covariance matrix for the score vector $\mathbf{U}(\mathbf{0})$ are thus given by Lemma 8.1.1 as $E[\mathbf{U}(\mathbf{0})] = \mathbf{0}$ and

$$\begin{aligned}
\mathbf{V} &= E[\mathbf{U}(\mathbf{0})\mathbf{U}(\mathbf{0})'] \\
&= \left(\sum_{i=1}^{n} \alpha_i^2/(n-1)\right)\mathbf{X}_c'\mathbf{X}_c,
\end{aligned} \tag{8.1.7}$$

where $\mathbf{X}_c$ is the $n \times p$ matrix of x's centered about their means. For the m-sample problem, where $\mathbf{x}' = (x_1, \ldots, x_{m-1})$, the score vector $\mathbf{U}(\mathbf{0})$ in (8.1.4) and covariance matrix $\mathbf{V}$ in (8.1.7) have components

$$U_\ell(0) = \sum_{i\in S_\ell} \alpha_{(i)} \qquad \ell = 1, \ldots, m-1$$

$$V_{\ell s} = \left(\sum_{i=1}^{n} \alpha_i^2/(n-1)\right)\left(N_\ell\delta_{\ell s} - \frac{N_\ell N_s}{n}\right) \qquad \ell, s = 1, \ldots, m-1, \tag{8.1.8}$$

where $\delta_{\ell s} = I(\ell = s)$ and S_ℓ denotes the individuals in the sample who are from distribution ℓ.

Under quite general conditions (e.g., Hajek and Sidak 1967, p. 159), the distribution of $\mathbf{U}(\mathbf{0})$ is asymptotically normal, and the equality of the m distributions can be

tested with the statistic

$$X^2 = \mathbf{U}(\mathbf{0})' V^{-1} \mathbf{U}(\mathbf{0}), \tag{8.1.9}$$

where $\mathbf{U}(\mathbf{0})$ and V are given by (8.1.8). Under the hypothesis that the distributions are identical, X^2 is distributed approximately as $\chi^2_{(m-1)}$. Large values of X^2 indicate evidence against the hypothesis of equality. With small samples, it is also feasible to compute exact test properties or p-values from the permutation distribution of $\mathbf{U}(\mathbf{0})$.

In order to obtain a rank test we have to define scores. Usually scores are selected by basing the α_i in (8.1.5) on a specific p.d.f., $g(y)$. If the data actually arise from a model (8.1.2) with this p.d.f., then the rank test is asymptotically fully efficient relative to the parametric procedure based on the actual observations, y_j. In addition, the rank test generally retains substantially higher efficiency than the corresponding parametric test when the model is of the form (8.1.2), but with a different p.d.f. Finally, p-values calculated from the rank test are valid regardless of the common underlying distribution of the observations. This is not true for parametric tests based on a specific model.

We consider two examples of rank tests, both of which will be later extended to the censored data case.

Example 8.1.1. Exponential Ordered Scores (Log Rank) Test. If scores are generated by letting $g(y)$ in (8.1.5) be the extreme value p.d.f. $\exp(y - e^y)$, $-\infty < y < \infty$, then $g'(y)/g(y) = 1 - e^y$ and

$$\alpha_i = E(e^{y_{(i)}} - 1).$$

Since $v = e^y$ has a standard exponential distribution with p.d.f. e^{-v}, $v \geq 0$, $E(\exp y_{(i)})$ is the expected value of the ith-order statistic in a random sample of size n from the standard exponential distribution. From Theorem 4.1.1 it follows that

$$\begin{aligned}\alpha_i &= \sum_{\ell=1}^{i} \frac{1}{n-\ell+1} - 1 \\ &= e_{i,n} - 1.\end{aligned}$$

The α_i are sometimes called exponential ordered scores (Cox, 1964).

In the two-sample test, for example, the rank statistic and its variance are, from (8.1.8),

$$U(0) = \sum_{i \in S_1} e_{(i),n} - N_1$$

$$V = \frac{N_1 N_2}{n(n-1)} \sum_{i=1}^{n} (e_{i,n} - 1)^2.$$

The variance can be simplified slightly by using the easily proved relations

$$\sum_{i=1}^{n} e_{i,n} = n \qquad \text{and} \qquad \sum_{i=1}^{n} e_{i,n}^2 = 2n - e_{n,n}$$

to give

$$V = \frac{N_1 N_2}{n(n-1)} (n - e_{n,n}).$$

If the data arise from two extreme value distributions differing only with respect to location, the asymptotic relative efficiency of this rank test is one. If the data come from normal distributions differing only with respect to location, the asymptotic relative efficiency turns out to be .82. The efficiency of the test is discussed further in Section 8.1.5.

Example 8.1.2. Wilcoxon Test. If scores are generated by taking the logistic p.d.f. $g(y) = e^y/(1+e^y)^2$, $-\infty < y < \infty$, the Wilcoxon test (Wilcoxon, 1945) is obtained. In this case $g'(y)/g(y) = 1 - 2e^y/(1+e^y) = 1 - 2G(y)$, where $G(y)$ is the distribution function corresponding to $g(y)$. Since $G(y)$ is uniformly distributed on (0, 1), we find

$$\begin{aligned}\alpha_i &= E[2G(y_{(i)}) - 1] \\ &= \frac{2i}{n+1} - 1,\end{aligned} \tag{8.1.10}$$

using the well-known fact that the mean of the ith-order statistic from a random sample from Uniform(0, 1) is $i/(n+1)$.

For the two-sample test (8.1.8) gives the rank statistic and its variance as

$$U(0) = \frac{2}{n+1} \sum_{i \in S_1} (i) - N_1$$

$$\begin{aligned}V &= \frac{N_1 N_2}{n(n-1)} \sum_{i=1}^{n} \left(\frac{2i}{n+1} - 1 \right)^2 \\ &= \frac{N_1 N_2}{3(n+1)}.\end{aligned}$$

The two-sample Wilcoxon statistic is often considered in different but equivalent forms. The test is asymptotically fully efficient for detecting location shifts when the underlying distributions are logistic. In view of the similarity of the logistic and normal distributions, it also would be expected to have high efficiency when the underlying distributions are normal distributions differing only in mean. In fact, the asymptotic relative efficiency of the Wilcoxon test in this case is .95. If the under-

lying distributions are extreme value distributions, on the other hand, the asymptotic relative efficiency is .75.

8.1.1.2 *Tests with Censored Data*

When the data are censored, some modification of the procedures just described is needed. Early approaches to this problem focused on adjustments to score statistics (8.1.4) in which rank-based weights are determined for both censored and uncensored observations. Later work has tended to emphasize the counting process framework, described in Section 8.1.4. A brief description of rank tests for censored data will be given, followed by a more detailed discussion of generalizations of the log rank and Wilcoxon tests. Numerous references to this area are provided in the Bibliographic Notes at the end of the chapter.

Consider the linear regression model (8.1.2) once again, and suppose that from a sample involving n individuals with covariate vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ there arise k distinct observed log lifetimes $y_{(1)} < \cdots < y_{(k)}$ and $n - k$ censoring times. In addition, suppose that m_i log censoring times lie in the interval $[y_{(i)}, y_{(i+1)})$, for $i = 0, 1, \ldots, k$, where we define $y_{(0)} = 0$ and $y_{(k+1)} = \infty$. Let $\mathbf{x}_{(i)}$ be the covariate vector associated with the individual whose y-value is $y_{(i)}$, and let $\mathbf{s}_{(i)}$ be the sum of these vectors for the m_i individuals with log censoring times in $[y_{(i)}, y_{(i+1)})$. To construct rank tests of the hypothesis H_0: $\boldsymbol{\theta} = \mathbf{0}$ in this situation, Prentice (1978) and others proposed the use of a pseudoscore statistic that has components of the form

$$U_\ell(\mathbf{0}) = \sum_{i=1}^{k} (x_{(i)\ell}\alpha_i + s_{(i)\ell}a_i) \qquad \ell = 1, \ldots, m-1. \tag{8.1.11}$$

That is, individuals whose lifetimes are censored are given scores a_i that are different from the scores of those whose lifetimes are observed. All individuals censored in $[t_{(i)}, t_{(i+1)})$ are given the same score, regardless of their respective censoring times.

Prentice (1978) suggested a general method of obtaining scores α_i and a_i for (8.1.11) and discussed estimation of the covariance matrix of $\mathbf{U}(\mathbf{0}) = [U_1(\mathbf{0}), \ldots, U_{m-1}(\mathbf{0})]'$. This will not be considered here, except to note that scores can be defined so that $E[\mathbf{U}(\mathbf{0})] = \mathbf{0}$ and that a kind of permutation variance for $\mathbf{U}(\mathbf{0})$ can be obtained, with entries for $\ell, s = 1, \cdots, m-1$ of

$$\begin{aligned} V_{\ell s} &= E[U_\ell(\mathbf{0})U_s(\mathbf{0})] \\ &= \left(\sum_{i=1}^{k} (\alpha_i^2 + m_i a_i^2)/(n-1)\right)\left(N_\ell \delta_{\ell s} - \frac{N_\ell N_s}{n}\right). \end{aligned} \tag{8.1.12}$$

This permutation variance is conditional on the particular assignment of scores in the situation at hand, and is thus conditioned on $m_0, m_1, \ldots, m_k$. Consequently, formula (8.1.12) should be used only when it can safely be assumed that censoring is independent of $\mathbf{x}$. Alternative variance estimates are developed in Sections 8.1.2–8.1.4.

An approximation to the linear rank statistics (8.1.4) provides another choice of weights. Asymptotically equivalent tests to those based on (8.1.4) in the uncensored data case are obtained if we replace the weights α_i in (8.1.5) with ones that are appropriate functions of expected Uniform(0, 1) order statistics $i/(n+1)$. Thus, for m-sample tests we consider weights α_i in (8.1.8) that are of the form $\phi[i/(n+1)]$. With censored data, the values $i/(n+1)$ are not known for all (i), $i = 1, \ldots, k$, and the usual procedure is to define the weight for an uncensored observation $t_{(i)}$ as $\phi[1 - \tilde{S}(t_{(i)})]$, where $\tilde{S}(t)$ is an estimate of the common survivor function $S(t) = S_1(t) = \cdots = S_m(t)$ under the hypothesis of equality of the m lifetime distributions. The weights for censored observations are more complicated (e.g., see Andersen et al. 1993, p. 351), but have the same property as those in (8.1.11): all of the censored lifetimes between any two observed lifetimes $t_{(i)}$ and $t_{(i+1)}$ are given the same weight. See Problem 8.1 for further details concerning this.

The rank test procedures are not strictly distribution-free when there is censoring: the distribution and properties of the test statistics depend on the censoring and lifetime distributions involved. In this respect, observe that the assignment of scores in (8.1.11) is not, in general, prespecified, but depends on the observed data. It is prespecified, however, for various kinds of Type 2 censoring.

We now consider extensions of the log rank and Wilcoxon tests to the censored-data situation.

8.1.2 The Exponential Ordered Scores (Log Rank) Test with Censored Data

With uncensored data the exponential ordered scores test for the equality of two or more distributions employs the scores α_i given in Example 8.1.1. To discuss the case of censored data it is convenient to use the notation of earlier chapters: specifically, suppose that n_i is the total number of individuals at risk across all m distributions just prior to $t_{(i)}$, where $t_{(i)} = \exp[y_{(i)}]$ is the ith observed lifetime ($i = 1, \ldots, k$). Let d_i be the number of deaths at $t_{(i)}$; for now d_i is taken to be one, since lifetimes are assumed to be distinct, but we later allow d_i to be greater than one to handle ties in the data. Let S_ℓ be the set of individuals from distribution ℓ and define, for $\ell = 1, \ldots, m$ and $i = 1, \ldots, k$,

$$\begin{aligned} d_{\ell i} &= \text{Number of deaths at } t_{(i)} \text{ among individuals in } S_\ell; \\ n_{\ell i} &= \text{Number of individuals from } S_\ell \text{ at risk just prior to } t_{(i)}. \end{aligned}$$

Of course,

$$\sum_{\ell=1}^{m} d_{\ell i} = d_i \qquad \text{and} \qquad \sum_{\ell=1}^{m} n_{\ell i} = n_i.$$

Prentice (1978) and Peto and Peto (1972) have suggested the following scores for use with (8.1.11):

$$\left.\begin{aligned} \alpha_i &= \sum_{j=1}^{i} \frac{1}{n_j} - 1 \\ a_i &= \sum_{j=1}^{i} \frac{1}{n_j} \end{aligned}\right\} \qquad i = 1, \ldots, k. \tag{8.1.13}$$

It is easily seen that when there is no censoring, the α_i in (8.1.13) are identical to those in Example 8.1.1. To motivate (8.1.13) in the censored case, note that for the extreme value distribution that generated the scores in Example 8.1.1, $-g'(y)/g(y) = e^y - 1 = H(y) - 1$, where $H(y)$ is the distribution's cumulative hazard function. In (8.1.13) α_i is seen to be $\tilde{H}(y_{(i)}+) - 1$, where $\tilde{H}(y)$ is the empirical cumulative hazard function (3.2.13). The score a_i is $\alpha_i + 1$; this is motivated by the observation that led to (6.2.8), namely, that $H(y)$ has a standard exponential distribution, suggesting an adjustment of $+1$ to a censored observation.

For the m-sample problem, (8.1.11) in conjunction with the scores (8.1.13) gives the rank statistic

$$\begin{aligned} U_\ell(\mathbf{0}) &= \sum_{i=1}^{k} [d_{\ell i}\alpha_i + (n_{\ell i} - d_{\ell i} - n_{\ell,i+1})a_i] \\ &= \sum_{i=1}^{k} \left[-d_{\ell i} + (n_{\ell i} - n_{\ell,i+1}) \left(\sum_{j=1}^{i} \frac{1}{n_j} \right) \right] \\ &= -\sum_{i=1}^{k} d_{\ell i} + \sum_{i=1}^{k} \frac{n_{\ell i}}{n_i}. \end{aligned}$$

Since $d_i = 1$ $(i = 1, \ldots, k)$ here, $U_\ell(\mathbf{0})$ can be rewritten as

$$U_\ell(\mathbf{0}) = -\sum_{i=1}^{k} \left(d_{\ell i} - \frac{n_{\ell i} d_i}{n_i} \right) \qquad \ell = 1, \ldots, m-1. \tag{8.1.14}$$

This is, aside from sign, the statistic (7.1.18) produced in Section 7.1.2 by the partial likelihood arguments for the proportional hazards model. From the expression (7.1.21) obtained in Chapter 7 the estimated covariance matrix for $\mathbf{U}(\mathbf{0})$ can be taken to have entries

$$V_{\ell s} = \sum_{i=1}^{k} \frac{d_i(n_i - d_i)n_{\ell i}}{n_i(n_i - 1)} \left(\delta_{\ell s} - \frac{n_{si}}{n_i} \right) \qquad \ell, s = 1, \ldots, m-1. \tag{8.1.15}$$

As discussed in Section 7.1.2, (8.1.14) and (8.1.15) can also be used if there is a small number of ties in the data, in which case some of the d_i will be greater than one.

The equality of the m distributions is tested with the statistic

$$X^2 = \mathbf{U}(\mathbf{0})' V^{-1} \mathbf{U}(\mathbf{0}),$$

which is distributed approximately as $\chi^2_{(m-1)}$ under the hypothesis of equality. The variance estimate (8.1.15) was derived from the observed information matrix for the partial likelihood of Section 7.1.2; an alternative is the permutation variance given by (8.1.12). This should be used, however, only if the censoring pattern is roughly the same in each of the m samples. When there is no censoring, the permutation variance is exact, as indicated in Example 8.1.1. Unless censoring differs somewhat across the m samples, or samples are rather small, the two variance formulas usually give results that are in close agreement.

The log rank test has been derived from two different points of view, first as a test based on the proportional hazards model of Chapter 7 and, here, as a linear rank test for location differences. As noted in Example 8.1.1, when there is no censoring, the test is asymptotically fully efficient for detecting location differences under an extreme value model. This is equivalent to stating that the test is asymptotically fully efficient for testing equality of lifetime distributions in a proportional hazards, or Lehmann, family, where lifetime T in distribution ℓ has survivor function of the form $S_\ell(t) = S_0(t)^{\delta_\ell}$. Crowley and Thomas (1975) showed that this result still holds under a random censorship model in which the same censoring distribution applies to each of the m samples, but that there is some loss of efficiency when the censoring distributions differ.

Properties for the log rank test are more conveniently studied using the counting process formulation of Section 8.1.4, and a few additional comments are provided there.

8.1.3 The Generalized Wilcoxon Test with Censored Data

The extension of the Wilcoxon test of Example 8.1.2 to the case of censored data has been discussed by several authors. Prentice (1978) suggested the statistic (8.1.11) in conjunction with scores

$$\left.\begin{aligned} \alpha_i &= 1 - 2\prod_{j=1}^{i} \frac{n_j}{n_j + 1} = 1 - 2F_i \\ a_i &= 1 - \prod_{j=1}^{i} \frac{n_j}{n_j + 1} = 1 - F_i \end{aligned}\right\} \qquad i = 1, \ldots, k, \tag{8.1.16}$$

where the n_j are defined as in Section 8.1.2. As motivation for α_i recall that for the logistic distribution that generates the Wilcoxon scores in the uncensored case, $-g'(y)/g(y) = 2G(y) - 1 = 1 - 2\overline{G}(y)$, where $\overline{G}(y)$ is the survivor function corresponding to $g(y)$. In (8.1.16), F_i is roughly equal to the product-limit estimate $\prod_{j=1}^{i}(n_j - 1)/n_j$ at $t_{(i)}+$. Motivation for a_i is provided in Problem 8.1.

The test based on (8.1.16) reduces to the test given in Example 8.1.2 in the case of uncensored data. To see this note that when there is no censoring and no ties, $F_i = n_i/(n+1)$ and $n_i = n - i + 1$, so that $\alpha_i = 1 - 2(n-i+1)/(n+1) = 2i/(n+1) - 1$.

The components of the score statistic (8.1.11) can be written in a simple form

$$U_\ell(\mathbf{0}) = \sum_{i=1}^{k} [d_{\ell i}\alpha_i + (n_{\ell i} - d_{\ell i} - n_{\ell,i+1})a_i]$$

$$= -\sum_{i=1}^{k} F_i d_{\ell i} + \sum_{i=1}^{k} (n_{\ell i} - n_{\ell,i+1})(1 - F_i)$$

$$= -\sum_{i=1}^{k} F_i d_{\ell i} + \sum_{i=1}^{k} (F_{i-1} - F_i) n_{\ell i}.$$

Now $F_{i-1} - F_i = F_i/n_i$, and if $d_i = 1$ (i.e., no ties) then

$$U_\ell(\mathbf{0}) = -\sum_{i=1}^{k} F_i \left(d_{\ell i} - d_i \frac{n_{\ell i}}{n_i}\right) \qquad \ell = 1, \ldots, m-1. \tag{8.1.17}$$

It is interesting to compare this with the score vector (8.1.14) for the log rank test. They differ only in the weight given to the terms $d_{\ell i} - d_i n_{\ell i}/n_i$: whereas in the log rank test the terms are given equal weight, in (8.1.17) they are weighted according to the estimate F_i of the survivor function at $t_{(i)}+$. The Wilcoxon test thus gives relatively more weight to earlier events than later ones.

An estimate of the covariance matrix V for $\mathbf{U}(\mathbf{0})$ has entries

$$V_{\ell s} = \sum_{i=1}^{k} F_i^2 \frac{d_i(n_i - d_i)}{n_i - 1} \frac{n_{\ell i}}{n_i} \left(\delta_{\ell s} - \frac{n_{si}}{n_i}\right) \qquad \ell, s = 1, \ldots, m-1, \tag{8.1.18}$$

as we discuss in the following Section 8.1.4. Equality of the m distributions can be tested with the statistic $X^2 = \mathbf{U}(\mathbf{0})' V^{-1} \mathbf{U}(\mathbf{0})$, which, under suitable conditions, is approximately $\chi^2_{(m-1)}$ under the hypothesis of equality. In addition, although the test has been developed on the assumption that there are no ties in the data, (8.1.17) and (8.1.18) can be employed when there is a small number of ties. In this case one should, however, define F_i as

$$\prod_{j=1}^{i} \frac{n_j - d_j + 1}{n_j + 1}.$$

As an alternative to (8.1.18), a permutation variance for $\mathbf{U}(\mathbf{0})$ can be used, provided that the censoring pattern is essentially the same in the different populations. According to the general formula (8.1.12), when there are no ties, terms in the permutation covariance matrix are

$$V'_{\ell s} = \left(\sum_{i=1}^{n} c_i^2/(n-1)\right)\left(N_\ell \delta_{\ell s} - \frac{N_\ell N_s}{n}\right) \qquad \ell, s = 1, \ldots, m-1, \tag{8.1.19}$$

where c_i represents the score (α_i or a_i) assigned to the ith individual in the combined sample and N_j is the number of individuals from population j ($j = 1, \ldots, m$). Breslow (1970) proved in the two-sample case that (8.1.18) and (8.1.19) are asymptotically equivalent, provided that the same random independent censoring mechanism applies in the two populations. In finite samples the two variance estimates do not usually differ much, assuming that censoring is similar in the different populations.

An illustration of the Wilcoxon test is given in Section 8.1.5.

8.1.4 Counting Process Formulation of m-Sample Tests

Suppose that the hypothesis $H: S_1(t) = S_2(t)$ that two lifetime distributions are identical is to be tested. This can equivalently be expressed as $H: H_1(t) = H_2(t)$, where $H_\ell(t)$ is the cumulative hazard function (c.h.f.) for distribution ℓ ($\ell = 1, 2$). In terms of the counting process notation of Sections 3.2.4 and 7.2.3, a natural nonparametric procedure would be to use a statistic of the form

$$W = \int_0^\infty k(u)[d\hat{H}_1(u) - d\hat{H}_2(u)], \tag{8.1.20}$$

where $\hat{H}_1(t)$ and $\hat{H}_2(t)$ are the Nelson–Aalen estimates based on the independent samples from distributions 1 and 2, and $k(u)$ is a predictable weight function. Interestingly, this approach leads to linear rank tests of the type discussed in the preceding three sections.

Let us set up counting process notation as follows: if there are independent censored random samples from distributions $\ell = 1, 2, \ldots, m$, let $\{(t_{\ell i}, \delta_{\ell i}), i = 1, \ldots, n_\ell\}$ denote the sample from distribution ℓ, and let

$$dN_{\ell i}(t) = I(t_{\ell i} = t, \delta_{\ell i} = 1) \qquad \text{and} \qquad Y_{\ell i}(t) = I(t_{\ell i} \geq t)$$

be the failure and at-risk indicators at time t. As earlier, let $dN_{\ell\cdot}(t) = \sum_i dN_{\ell i}(t)$, $Y_{\ell\cdot}(t) = \sum_i Y_{\ell i}(t)$, and $dN_{\cdot\cdot}(t) = \sum_\ell dN_{\ell\cdot}(t)$, $Y_{\cdot\cdot}(t) = \sum_\ell Y_{\ell\cdot}(t)$. In the two-sample case, it is a sensible requirement that $k(u) = 0$ if $Y_{1\cdot}(u)Y_{2\cdot}(u) = 0$, and in that case (8.1.20) can be written in the alternative form

$$W = \int_0^\infty w(u)[d\hat{H}_1(u) - d\hat{H}(u)], \tag{8.1.21}$$

where $w(u) = k(u)Y_{\cdot\cdot}(u)/Y_{2\cdot}(u)$ and $d\hat{H}(u) = dN_{\cdot\cdot}(u)/Y_{\cdot\cdot}(u)$ is the increment in the Nelson–Aalen estimate based on the combined data from both samples. Note that $\hat{H}(t)$ estimates the common cumulative hazard function under the hypothesis of distributional equality. For the m-sample problem involving a test of equality of m distributions, $H_1(t) = \cdots = H_m(t)$, we consider the obvious extension of (8.1.21), which is a vector $\mathbf{W} = (W_1, \ldots, W_{m-1})'$ with components

$$W_\ell = \int_0^\infty w_\ell(u)[d\hat{H}_\ell(u) - d\hat{H}(u)], \qquad \ell = 1, \ldots, m-1. \tag{8.1.22}$$

In the notation of Sections 8.1.2 and 8.1.3, let $t_{(1)} < \cdots < t_{(k)}$ be the distinct observed failure times in the combined sample from all m distributions, and note that $Y_{\ell\cdot}(t_{(i)}) = n_{\ell i}$, $Y_{\cdot\cdot}(t_{(i)}) = n_i$, $dN_{\ell\cdot}(t_{(i)}) = d_{\ell i}$, and $dN_{\cdot\cdot}(t_{(i)}) = d_i$. Thus (8.1.22) takes the form

$$W_\ell = \sum_{i=1}^{k} w_\ell(t_{(i)}) \left(\frac{d_{\ell i}}{n_{\ell i}} - \frac{d_i}{n_i} \right).$$

Most tests consider weight functions of the form $w_\ell(t) = Y_\ell(t)w(t)$, and in this case we get

$$W_\ell = \sum_{i=1}^{k} w(t_{(i)}) \left(d_{\ell i} - \frac{n_{\ell i}}{n_i} d_i \right), \qquad \ell = 1, \ldots, m-1. \tag{8.1.23}$$

As noted, both the log rank statistics (8.1.14) and the generalized Wilcoxon statistics (8.1.17) are of this form: the former has $w(t_{(i)}) = -1$ and the latter $w(t_{(i)}) = -F_i$. This class of tests is often referred to as the weighted log rank class. Note that for the tests to be rank-based and thus distribution-free when there is no censoring, the function $w(t_{(i)})$ must depend only on the rank (i) and not on the t-value, $t_{(i)}$. When censoring is present the tests are not strictly distribution-free, with even properties under the null hypothesis depending in general on the censoring process and common lifetime distribution. However, the procedures are nonparametric and are valid whatever the underlying distribution happens to be.

Distribution theory for statistics of the form (8.1.22) with $w_\ell(u) = Y_\ell(u)w(u)$ can conveniently be approached through martingale arguments outlined in Appendix F. Note in particular that if $H_1(t) = \cdots = H_m(t)$, then

$$\begin{aligned} W_\ell &= \int_0^\infty w(u) \left\{ dN_{\ell\cdot}(u) - Y_{\ell\cdot}(u) \frac{dN_{\cdot\cdot}(u)}{Y_{\cdot\cdot}(u)} \right\} \\ &= \int_0^\infty w(u) \left\{ dM_{\ell\cdot}(u) - \frac{Y_{\ell\cdot}(u)}{Y_{\cdot\cdot}(u)} dM_{\cdot\cdot}(u) \right\}, \end{aligned}$$

where $dM_{\ell\cdot}(u) = dN_{\ell\cdot}(u) - Y_{\ell\cdot}(u)dH(u)$ and $dM_{\cdot\cdot}(u) = dN_{\cdot\cdot}(u) - Y_{\cdot\cdot}(u)dH(u)$, with $H(u)$ the common cumulative hazard function. The processes $\{M_{\ell\cdot}(u)\}$ and $\{M_{\cdot\cdot}(u)\}$ are martingales, so the behavior of $W_1, \ldots, W_{m-1}$ is easily studied if $w(u)$ is predictable. An application of (F16) in Appendix F gives an estimate of the covariance of W_ℓ and W_s for $\ell, s = 1, \ldots, m-1$ as

$$\begin{aligned} V_{\ell s} &= \int_0^\infty w(u)^2 \frac{Y_{\ell\cdot}(u)}{Y_{\cdot\cdot}(u)} \left(\delta_{\ell s} - \frac{Y_{s\cdot}(u)}{Y_{\cdot\cdot}(u)} \right) \left(\frac{Y_{\cdot\cdot}(u) - dN_{\cdot\cdot}(u)}{Y_{\cdot\cdot}(u) - 1} \right) dN_{\cdot\cdot}(u) \\ &= \sum_{i=1}^{k} w(t_{(i)})^2 \frac{n_{\ell i}}{n_i} \left(\delta_{\ell s} - \frac{n_{si}}{n_i} \right) \frac{(n_i - d_i)d_i}{n_i - 1}. \end{aligned} \tag{8.1.24}$$

The weights $w(t_{(i)}) = -1$ and $w(t_{(i)}) = -F_i$ give the expressions (8.1.15) and (8.1.18) for the log rank and Wilcoxon tests in previous sections.

The statistic $X^2 = \mathbf{W}'V^{-1}\mathbf{W}$, where $\mathbf{W} = (W_1, \ldots, W_{m-1})'$ and $V = (V_{\ell s})$, can be used to test $H_1(t) = \cdots = H_m(t)$. Under the null hypothesis, X^2 is asymptotically $\chi^2_{(m-1)}$.

The generalized Wilcoxon test statistic given earlier is of the form (8.1.23), with $w(t_{(i)}) = -F_i$, where

$$F_i = \prod_{j=1}^{i} \frac{n_j + 1 - d_j}{n_j + 1},$$

as suggested by Prentice (1978). A different generalization of the Wilcoxon test was considered by Gehan (1965), who suggested the weight function $w(t_{(i)}) = n_i$; Prentice and Marek (1979) and others, however, have warned against the use of this test if censoring patterns differ substantially across the m samples. Harrington and Fleming (1982) suggested a family of tests with

$$w(t_{(i)}; \rho) = \hat{S}(t_{(i)})^\rho \qquad \rho \geq 0.$$

The special case $\rho = 0$ gives the log rank test and $\rho = 1$ gives something close to the Prentice–Wilcoxon test. Fleming and Harrington (1991, p. 275) indicate that the test using $w(t_{(i)}; \rho)$ is efficient against alternatives for which any two distributions have hazard functions related by

$$\lambda_2(t) = \lambda_1(t)e^\Delta \left\{ S_1(t)^\rho + [1 - S_1(t)^\rho]e^\Delta \right\}^{-1}.$$

The case $\rho = 0$ gives proportional hazards, $\lambda_2(t) = e^\Delta \lambda_1(t)$. The generalized Wilcoxon test with $\rho = 1$ is efficient against alternatives in which $\lambda_1(t)/\lambda_2(t) \to 1$ as $t \to \infty$ ($S_1(t) \to 0$). This ties in with the rank-based approach in Sections 8.1.1 and 8.1.3, where the test arises from scores based on a log-logistic AFT model for lifetime. This model has hazard ratios with this property.

Two modifications to the preceding tests are sometimes valuable. The first involves stratification, as in Section 7.1.6; in this case, individuals are grouped into strata $j = 1, \ldots, J$ and the hypothesis to be tested is that

$$H_{1j}(t) = H_{2j}(t) = \cdots = H_{mj}(t), \qquad j = 1, \ldots, J, \tag{8.1.25}$$

where $H_{\ell j}(t)$ is the cumulative hazard function for those individuals in stratum j whose lifetimes are from distribution ℓ ($\ell = 1, \ldots, m$). Stratification could be used, for example, to test for the equality of two treatment effects when the lifetime distributions for males and females might be different. A test statistic for (8.1.25) can be obtained by combining stratum-specific statistics of the type (8.1.22) or (8.1.23). In particular, let $W_{\ell(j)}$ be the weighted log rank statistics (8.1.23) based on the individuals in stratum j only, and let $V_{(j)}$ be the estimated covariance matrix for

$\mathbf{W}_{(j)} = (W_{1(j)}, \ldots, W_{m-1(j)})'$, as given by (8.1.24). Then the statistic

$$X^2 = \left(\sum_{j=1}^{J} \mathbf{W}_{(j)}\right)' \left(\sum_{j=1}^{J} V_{(j)}\right)^{-1} \left(\sum_{j=1}^{J} \mathbf{W}_{(j)}\right) \tag{8.1.26}$$

can be used to test (8.1.25). If (8.1.25) is true, then the asymptotic distribution of X^2 in large samples is $\chi^2_{(m-1)}$.

A second modification is useful when $m \geq 3$ and alternative hypotheses of the form $H_1(t) \leq H_2(t) \leq \cdots \leq H_m(t)$ are of interest; note that these alternatives are equivalent to $S_1(t) \geq S_2(t) \geq \cdots \geq S_m(t)$. A test with more power than the test based on $X^2 = \mathbf{W}'V^{-1}\mathbf{W}$ can often be obtained by considering scores $a_1 < \cdots < a_m$ and a statistic $\mathbf{W}(\mathbf{a}) = \mathbf{a}'\mathbf{W}^*$, where $\mathbf{a} = (a_1, \ldots, a_m)'$ and $\mathbf{W}^* = (W_1, \cdots, W_m)$, components given by (8.1.23) with $\ell = m$ also being included. A statistic for testing the hypothesis of equal distributions is

$$X^2(\mathbf{a}) = \mathbf{W}(\mathbf{a})^2[\mathbf{a}'V^*\mathbf{a}]^{-1}, \tag{8.1.27}$$

where V^* is the covariance matrix estimate with entries (8.1.24), with the values $\ell = m$ and $s = m$ also included. The statistics (8.1.27) are invariant to linear transformations $a'_\ell = c_0 + c_1 a_\ell$ of the scores. Quite often the scores $a_\ell = \ell$ $(\ell = 1, \ldots, m)$ are useful, but in cases where the m distributions refer to individuals with certain quantitative characteristics it may be worth considering other choices. A good way to do this is to consider a proportional hazards model for which the quantitative characteristics are represented by a covariate x, so that $H_\ell(t) = H_0(t)\exp(\beta x_\ell)$ for $\ell = 1, \ldots, m$, with $x_1 < \cdots < x_m$ the relevant x-values for distributions $1, \ldots, m$. We then use $a_\ell \propto x_\ell$; see Problem 8.7 for further discussion.

Under the hypothesis $H_1(t) = \cdots = H_m(t)$, the statistics (8.1.27) are asymptotically $\chi^2_{(1)}$ in large samples, and large values of $X^2(\mathbf{a})$ provide evidence against equality.

8.1.5 Discussion and Examples

The log rank and generalized Wilcoxon tests for equality of distributions are effective in settings where alternatives involve stochastic ordering of the distributions in question, that is, either $S_\ell(t) \geq S_j(t)$ or $S_\ell(t) \leq S_j(t)$ for distributions ℓ and j. Each test is efficient in certain settings: the log rank when the m log-lifetime distributions are extreme value distributions differing only in location, and the Wilcoxon when they are logistic distributions differing only in location. There is considerable theoretical and empirical work on the relative efficiencies of weighted log rank tests (e.g., see Lee et al. 1975; Lininger et al. 1979; Leurgans 1983, 1984, and other references in the Bibliographic Notes). If one has specific alternative hypotheses in mind, then this can guide the choice of test via the choice of weight function in (8.1.23). For example, the log rank test will be relatively more effective at detecting differences in the right tails of the distributions, whereas the Wilcoxon will be more sensitive to early differences. Generally speaking, members of the family of tests tend to have good power across a range of stochastically ordered alternative hypotheses, however. A word of caution is that the asymptotic χ^2 approximations to test statistics can be poor if sample sizes are too small (e.g., see Latta 1981; Kellerer and Chmelevsky 1983).

Weighted log rank tests can be ineffective when the hazard or survivor functions for the different distributions cross. Then, statistics like (8.1.20) or (8.1.23) with positive-valued weight functions $k(t_{(i)})$ or $w(t_{(i)})$ tend to have early terms (i.e., for smaller $t_{(i)}$) and later terms of opposite signs, resulting in small values for the test statistic. A good approach in this case is to use time-varying covariates within a regression model, as illustrated in Example 7.1.2 of Section 7.1.8. This gives tests that are nonparametric but not rank-based, since the test statistic depends on the observed times, t_i. A second approach is to consider weighted log rank statistics in which the weights or scores have different signs for large versus small i. Alternatively, one can preselect a time τ and consider separate weighted log rank statistics for the observed data over $(0, \tau]$ and over (τ, ∞). The tests described in Section 8.1.4 apply when lifetimes are subject to independent left truncation as well as right censoring; the data over (τ, ∞) are treated as being left truncated at τ. With this approach the two test statistics (in the $m = 2$ case) or two sets of statistics (when $m > 2$) can either be combined or used separately; an illustration is provided in Example 8.1.3. A third approach is to use tests based on distance measures between distributions, which are robust to crossing survivor functions. For example, one might consider a statistic of Kolmogorov–Smirnov type, $D = \sup|\hat{S}_1(t) - \hat{S}_2(t)|$, for a test of the hypothesis $S_1(t) = S_2(t)$. In D, $\hat{S}_1(t)$ and $\hat{S}_2(t)$ are Kaplan–Meier estimates. It is beyond our scope here to consider such tests, but see Schumacher (1984), Andersen et al. (1993, Sec. V.4) and references therein. Klein and Moeschberger (1997, Sec. 7.6) provide a table of critical values for the two-sample case.

Tests of distributional equality are often associated with measures of the differences between distributions. An advantage of this (assuming that any models upon which the measure is based are satisfactory) is that such a parameter can be estimated. The linear rank tests of Sections 8.1.1–8.1.3 are associated with location shifts in location-scale models for log-lifetime, so one could, for example, supplement a significant test with estimates of the differences in median log-lifetimes for the distributions in question. For the log rank test there is also the close connection with proportional hazards models, and it is customary to estimate the relative risks $H_\ell(t)/H_s(t) = h_\ell(t)/h_s(t)$. Some tests have less appealing measures associated with them, for example, the Kolmogorov–Smirnov statistic D, so it can be a consideration in the selection of a test whether or not one wants associated estimates of distribution differences.

The following examples illustrate procedures in the preceding sections.

Example 8.1.3. (Example 7.1.1 revisited). Some data on remission times for leukemia patients were presented in Example 7.1.1. Patients were given one of two

treatments A and B, and it was desired to test the hypothesis that the remission duration distributions for the two treatment groups was the same.

The log rank statistic $U(0)$ given in (7.1.15) was computed in Example 7.1.1; it is the same as (8.1.14) for the case where $m = 2$, with a change of sign. Using the negative of (8.1.14), we find $U(0) = 3.323$. The variance estimate (8.1.15) for $m = 2$ is the same as (7.1.16), and gives $V = 8.1962$, so the X^2 statistic is $3.223^2/8.1962 = 1.35$. This gives a p-value of about .24 on $\chi^2_{(1)}$ and does not provide any evidence that the remission duration distributions are unequal.

The generalized Wilcoxon statistic (8.1.17) gives $U(0) = -U_1(0) = 2.269$ and (8.1.18) gives $V = 3.0105$, so $X^2 = U(0)^2/V = 1.71$. The p-value on $\chi^2_{(1)}$ is about .19, in agreement with the log rank test and similarly provides no evidence of a difference in distributions.

Plots of the Kaplan–Meier estimates for the two treatment groups indicate that the hazard functions are roughly proportional and suggest that the log rank, generalized Wilcoxon, or other weighted log rank tests should be powerful in this situation. We therefore conclude there is no evidence of a difference.

Example 8.1.4. Data on the survival times of 40 patients with advanced lung cancer were introduced in Example 1.1.9 and considered in Examples 6.3.3 and 6.4.3 using accelerated failure time regression models. The same data were considered as part of a larger data set in Example 7.2.2. One question of interest concerned the existence of a treatment effect: there were two treatments, Standard and Test. Of the various other covariates associated with individuals, only performance status (PS) appears important. Assuming that treatment was assigned randomly to individuals, we could carry out a two-sample test of no treatment effect with additional covariates ignored. A more efficient procedure, however, would be to adjust for the effect of performance status through stratification; we illustrate this procedure here. An approach that adjusts more fully for covariates is via regression modeling, as shown in Examples 6.3.3, 6.3.4, and 7.2.2.

We will carry out a stratified log rank test using three strata based on performance status: individuals with PS 10–30, 40–60, and 70–90, respectively, are in strata 1, 2, and 3. The procedure, described in Section 8.1.4, is to consider the hypothesis (8.1.25), which here is

$$H : H_{1j}(t) = H_{2j}(t), \qquad j = 1, 2, 3,$$

where $H_{1j}(t)$ and $H_{2j}(t)$ are the c.h.f.'s for individuals with treatments Standard and Test. Log rank statistics are computed for each stratum, and the statistic (8.1.26) is used for the test of H. The log rank statistics (8.1.4), or (8.1.23) with $w(t_{(i)}) = 1$ and their associated variance estimates from (8.1.15) or (8.1.24), are shown in Table 8.1; the score statistic for stratum j is denoted as u_j and the variance estimate as v_j. The test based on (8.1.26) returns the value $X^2 = (\sum u_j)^2/\sum v_j = .50$, which is insignificant on $\chi^2_{(1)}$, and agrees quite closely with the result from the proportional hazards regression analysis of Example 7.2.2. The individual hypotheses $H_{1j}(t) = H_{2j}(t)$ for each stratum $j = 1, 2, 3$ can be assessed by considering the log rank

Table 8.1. Log Rank Statistics from Stratified Sample

Stratum (PS)		
10–30	40–60	70–90
$u_1 = 1.02$	$u_2 = -1.95$	$u_3 = 2.64$
$v_1 = .65$	$v_2 = 2.47$	$v_3 = 2.73$

statistics $X^2_j = u^2_j/v_j$; this allows for the possibility that the treatment effect differs according to performance status level. The statistics X^2_j are 1.59, 1.54, and 2.56 for $j = 1, 2, 3$; none of these is significant on $\chi^2_{(1)}$ at the .10 level. Of course, the very small number of individuals in each treatment group within strata means that only very large treatment effects would be likely to be detected.

Example 8.1.5. Problem 7.4 gave data on survival times of patients with bile duct cancer who took part in a clinical trial to compare persons receiving a combination of radiation treatment and drug 5-FU with a control group. A plot of the Kaplan–Meier estimates for the two treatment groups (Treatment and Control) shows them to cross at about 400 days; see Figure 8.1. Weighted log rank tests do not provide evidence of a difference in the two distributions. The log rank value from (8.1.23) with $w(t_{(i)}) = 1$ is $W = -.652$ (with the control group as distribution 1), with associated variance estimate V from (8.1.24) of 9.3615, giving $X^2 = .045$ and a $\chi^2_{(1)}$ p-value of .83. The generalized Wilcoxon test using $w(t_{(i)}) = \hat{S}(t_{(i)})$, where $\hat{S}(t)$ is the

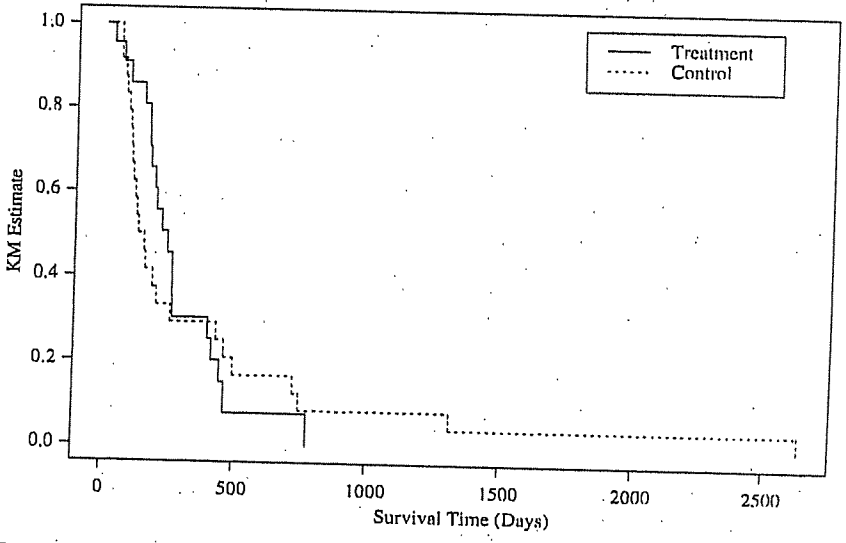

Figure 8.1. Kaplan–Meier estimates of survival with bile duct cancer for treatment and control groups.

Kaplan–Meier estimate from the combined sample, gives $W = -2.24$, $V = 3.750$, $X^2 = 1.34$, and a $\chi^2_{(1)}$ p-value of .25. The Wilcoxon test weights the earlier failures more heavily than later ones, and consequently gives a smaller p-value, but it is still far short of significant.

Other test statistics could be considered. One approach that was discussed for settings where survivor functions cross is to compute separate test statistics for data before and after a certain time. Let us consider this option by computing separate log rank statistics for the data (i.e., death and censoring times) up to $t = 400$ and after $t = 400$. The value 400 has been selected after looking at Figure 8.1, so one should not treat any p-values very seriously; however, the test will give an idea whether any *a priori* choice of test would have been likely to demonstrate a significant difference in distributions. The statistics W' and V' based on the data over $(0, 400]$ are given by (8.1.23) and (8.1.24), with only the $t_{(l)}$ in $(0, 400]$ included, and the statistics W'' and V'' based on the data over $(400, \infty)$ are given by (8.1.23) and (8.1.24), with only $t_{(l)}$ greater than 400 included. This gives $U' = -1.986$, $V' = 7.841$, $U'' = 1.334$, $V' = 1.521$. Neither segment of the data on its own gives a significant result, nor does a test based on the statistic $|U'| + |U''|$.

8.2 ESTIMATION FOR SEMIPARAMETRIC AFT MODELS

Semiparametric inference procedures for the regression coefficients $\boldsymbol{\beta}$ in log-location-scale models of the form (8.0.1) can be based on the ideas used to derive linear rank tests for comparing distributions. This is considered in Section 8.2.1, following which extensions to deal with time-varying covariates are discussed. We will see that this methodology is much harder to implement than that for the semiparametric proportional hazards model in Chapter 7.

8.2.1 Rank-Based Procedures with Fixed Covariates

Linear rank inference procedures for $\boldsymbol{\beta}$ in (8.0.1) can be based on the approach in Section 8.1.1. To do this, consider a hypothesized value $\boldsymbol{\beta}_0$ under which the "errors" $Z_i = Y_i - \boldsymbol{\beta}_0'\mathbf{x}_i$ are independent and identically distributed (i.i.d.). If (y_i, δ_i), $i = 1, \ldots, n$ is a censored random sample of log failure times generated by (8.0.1) and an independent censoring mechanism, the values

$$z_i(\boldsymbol{\beta}_0) = y_i - \boldsymbol{\beta}_0'\mathbf{x}_i, \qquad i = 1, \ldots, n \tag{8.2.1}$$

are a censored random sample from the distribution of Z. Let us write $z_i(\boldsymbol{\beta}_0)$ as z_i for convenience, remembering that it depends on $\boldsymbol{\beta}_0$, and let $z_{(1)} < \cdots < z_{(k)}$ be the distinct values among $z_1, \ldots, z_n$ that correspond to observed failure times (i.e., to units with $\delta_i = 1$). Now consider statistics of the form (8.1.11),

$$\mathbf{U}(\boldsymbol{\beta}_0) = \sum_{l=1}^{k} (a_l \mathbf{x}_{(l)} + a_l s_{(l)}), \tag{8.2.2}$$

where it is assumed that there are no tied failure times, and where $s_{(l)} = \sum \mathbf{x}_\ell$, with the sum being over those individuals ℓ for whom $\delta_l = 0$ and $z_{(l)} \le z_\ell < z_{(l+1)}$. Under (8.0.1) and the hypothesis that $\boldsymbol{\beta}_0$ is the true value for $\boldsymbol{\beta}$, the properties of the statistic $\mathbf{U}(\boldsymbol{\beta}_0)$ are the same as those of the statistic (8.1.11).

As shown in Section 8.1.4, it is convenient to consider linear rank statistics in counting process form, so we will examine estimating functions of the form

$$\mathbf{U}(\boldsymbol{\beta}_0) = \sum_{i=1}^{n} \delta_i w(z_i) \left\{ \mathbf{x}_i - \frac{\sum_{\ell=1}^{n} \mathbf{x}_\ell I(z_\ell \ge z_i)}{\sum_{\ell=1}^{n} I(z_\ell \ge z_i)} \right\}, \tag{8.2.3}$$

where $w(z_i)$ is a weight function that is non-zero only if $\delta_i = 1$ and that depends only on the rank of z_i among values with $\delta = 1$. When there are no tied values, (8.2.3) can be written as

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{k} w_i \left\{ \mathbf{x}_{(i)} - \frac{\sum_{\ell=1}^{n} \mathbf{x}_\ell I(z_\ell \ge z_{(i)})}{\sum_{\ell=1}^{n} I(z_\ell \ge z_{(i)})} \right\}. \tag{8.2.4}$$

It is easily shown that a statistic of the form (8.2.4) can be rewritten in the form (8.2.2), with α_i and a_i given by (8.3.6) in Problem 8.2, and $n_i = \sum I(z_\ell \ge z_{(i)})$. Conversely it can be shown, exactly as in Problem 8.2, that any statistic of the form (8.2.2) for which α_i and a_i satisfy (8.3.7) can be expressed in the form (8.2.4). The form (8.2.3) is also defined when there are tied values among the z_i with $\delta_i = 1$. As with the m-sample test procedures in Section 8.1, we get different estimating functions $\mathbf{U}(\boldsymbol{\beta}_0)$ through different choices of weight functions. The most common choices are $w_i = 1$ and $w_i = \hat{S}(z_{(i)})$, respectively, in (8.2.4); they correspond to the log rank and Wilcoxon weights used for the tests in Section 8.1. Here, $\hat{S}(z)$ is the Kaplan–Meier estimate based on the data (z_i, δ_i), $i = 1, \ldots, n$.

It can be shown using martingale arguments like those mentioned in Section 8.1.4 that $n^{-1/2}\mathbf{U}(\boldsymbol{\beta}_0)$ is asymptotically normal. For notational convenience define for $-\infty < z < \infty$

$$\bar{\mathbf{x}}(z; \boldsymbol{\beta}_0) = \frac{\sum_{\ell=1}^{n} I(z_\ell \ge z)\mathbf{x}_\ell}{\sum_{\ell=1}^{n} I(z_\ell \ge z)}. \tag{8.2.5}$$

Then (8.2.3) and (8.2.4) can be expressed as

$$\mathbf{U}(\boldsymbol{\beta}_0) = \sum_{i=1}^{n} \delta_i w(z_i)\{\mathbf{x}_i - \bar{\mathbf{x}}(z_i; \boldsymbol{\beta}_0)\}, \tag{8.2.6}$$

and $n^{-1/2}\mathbf{U}(\boldsymbol{\beta}_0)$ is asymptotically normal with mean vector $\mathbf{0}$ and a covariance matrix that is estimated consistently by

$$V(\boldsymbol{\beta}_0) = \frac{1}{n} \sum_{i=1}^{n} \delta_i w(z_i)^2 \left\{ \frac{\sum_{\ell=1}^{n} I(z_\ell \ge z_i)[\mathbf{x}_\ell - \bar{\mathbf{x}}(z_i; \boldsymbol{\beta}_0)][\mathbf{x}_\ell - \bar{\mathbf{x}}(z_i; \boldsymbol{\beta}_0)]'}{\sum_{\ell=1}^{n} I(z_\ell \ge z_i)} \right\}. \tag{8.2.7}$$

The estimating function $\mathbf{U}(\boldsymbol{\beta}_0)$ depends on $\boldsymbol{\beta}_0$ only through the ranks of the values $z_i = y_i - \boldsymbol{\beta}_0'\mathbf{x}_i$. Thus, two different values $\boldsymbol{\beta}_0$ will give different $\mathbf{U}(\boldsymbol{\beta}_0)$ values only if the ordering of the z_i they give is different. This means that $\mathbf{U}(\boldsymbol{\beta}_0)$ is a complex step function of $\boldsymbol{\beta}_0$, and it is not necessarily monotone. As a result the preferred way to define a point estimate $\hat{\boldsymbol{\beta}}$ is as the value $\boldsymbol{\beta}_0$ that minimizes some norm $\| \mathbf{U}(\boldsymbol{\beta}_0) \|$, such as $\mathbf{U}(\boldsymbol{\beta}_0)'V(\boldsymbol{\beta}_0)^{-1}\mathbf{U}(\boldsymbol{\beta}_0)$ or $\mathbf{U}(\boldsymbol{\beta}_0)'\mathbf{U}(\boldsymbol{\beta}_0)$. There has been relatively little consideration of point estimation in the literature, in part because it turns out that variance estimation is difficult. In addition, the computation of the vector $\boldsymbol{\beta}_0$ that minimizes some norm can be challenging when $\dim(\boldsymbol{\beta}_0) > 1$. Lin and Geyer (1992) describe an approach based on simulated annealing; Jones (1997) describes another method.

Interval estimation of $\boldsymbol{\beta}$ is more conveniently based on the estimating function $\mathbf{U}(\boldsymbol{\beta}_0)$ than on $\hat{\boldsymbol{\beta}}$. If $\boldsymbol{\beta}_0$ is the true value, then $W_1(\boldsymbol{\beta}_0) = n^{-1}\mathbf{U}(\boldsymbol{\beta}_0)'V(\boldsymbol{\beta}_0)^{-1}\mathbf{U}(\boldsymbol{\beta}_0)$ is asymptotically $\chi^2_{(p)}$, where $p = \dim(\boldsymbol{\beta})$, and an approximate α confidence region for $\boldsymbol{\beta}$ is given by the set

$$\left\{\boldsymbol{\beta}_0 : W_1(\boldsymbol{\beta}_0) \le \chi^2_{(p),\alpha}\right\}. \tag{8.2.8}$$

An asymptotically equivalent alternative is to replace $V(\boldsymbol{\beta}_0)$ in $W_1(\boldsymbol{\beta}_0)$ with $V(\hat{\boldsymbol{\beta}})$. A similar approach can be used to estimate subsets of the parameters. For example, let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ and suppose we want to estimate $\boldsymbol{\beta}_1$. Wei et al. (1990) showed that we can consider

$$W_2(\boldsymbol{\beta}_2; \boldsymbol{\beta}_{10}) = n^{-1}\mathbf{U}(\tilde{\boldsymbol{\beta}})'V(\hat{\boldsymbol{\beta}})^{-1}\mathbf{U}(\tilde{\boldsymbol{\beta}}),$$

where $\tilde{\boldsymbol{\beta}} = (\boldsymbol{\beta}_{10}', \boldsymbol{\beta}_2')'$. Under the hypothesis that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{10}$, the statistic

$$W_3(\boldsymbol{\beta}_{10}) = \min_{\boldsymbol{\beta}_2} W_2(\boldsymbol{\beta}_2; \boldsymbol{\beta}_{10}) \tag{8.2.9}$$

is asymptotically $\chi^2_{(q)}$, where $q = \dim(\boldsymbol{\beta}_1)$. An approximate α confidence region for $\boldsymbol{\beta}_1$ consists of vectors in the set

$$\left\{\boldsymbol{\beta}_{10} : W_3(\boldsymbol{\beta}_{10}) \le \chi^2_{(q),\alpha}\right\}.$$

This procedure can in particular be used to get confidence intervals for individual parameters.

Example 8.2.1. Fleming and Harrington (1991) presented data on 418 patients who took part in a study on primary biliary cirrhosis (PBC) of the liver between 1974 and 1984. This is a disease that eventually leads to destruction of liver function and death. The data, available in Appendix D1 of Fleming and Harrington (1991) and in electronic form (see Appendix G), contain information on survival time T (the number of days between registration in the study and death), assigned treatment,

age, sex, biochemical measurements, and disease conditions. Here we consider only a group of five variables that have been found to be important in predicting survival time. These are

Age—Patient age at time of registration;
Edema—A variable scaled to take values 0, .5, and 1, respectively, denoting three levels of edema, of increasing severity;
Albumin—Serum albumin concentration;
Bilirubin—Serum bilirubin concentration;
Protime—Prothrombin time.

Two patients had missing prothrombin times and are dropped in the analyses discussed here.

Table 8.2 shows the results of fitting several log-location-scale models of the form (8.0.1). Parametric extreme value and logistic models were fitted as described in Sections 6.3.2 and 6.3.3. Table 8.2 shows maximum likelihood estimates (m.l.e.'s) for the regression coefficients, intercept, and scale parameter for each model. For each regression coefficient the Z^2 value $\hat{\beta}_j^2/se(\hat{\beta}_j)^2$ is shown. In addition, a "robust" Z^2 value is shown, obtained by replacing $se(\hat{\beta}_j)$ in Z^2 by a robust standard error calcu-

Table 8.2. Parametric and Semiparametric Log-Location-Scale Models for PBC Data

	Extreme-Value			Semiparametric ($w_i = 1$)	
Term	Estimate	Z^2	Z^2 (robust)	Estimate	W
Intercept	12.15	—	—	—	—
Age	−.026	25.5	19.6	−.027	20.2
Edema	−.64	13.0	8.9	−.69	7.4
Log (albumin)	1.63	14.8	16.5	1.66	9.4
Log (bilirubin)	−.57	106.3	101.6	−.58	68.1
Log (protime)	−1.72	11.0	7.1	−1.88	6.9
Scale	.67	—	—	—	—

	Logistic			Semiparametric ($w_i = \hat{S}(z_{(i)})$)	
Term	Estimate	Z^2	Z^2 (robust)	Estimate	W
Intercept	13.51	—	—	—	—
Age	−.028	22.2	22.9	−.027	21.7
Edema	−.81	13.5	10.3	−.80	10.3
Log (albumin)	1.54	10.8	10.4	1.60	9.9
Log (bilirubin)	−.58	86.1	83.6	−.59	76.6
Log (protime)	−2.33	11.3	7.0	−2.30	9.4
Scale	.51	—	—	—	—

lated using the robust asymptotic covariance matrix (C30) in Appendix C. Finally, semiparametric estimates obtained by Lin and Geyer (1992, Table 1) are shown; they are obtained by minimizing $\| \mathbf{U}(\boldsymbol{\beta}_0) \|$ using the L_1 norm. Also shown are W values, obtained by computing for each β_j the statistic $W_3(\boldsymbol{\beta}_{10})$ given by (8.2.9), with $\beta_j = 0$ replacing $\boldsymbol{\beta}_{10}$ and $\boldsymbol{\beta}_2$ identified with the regression coefficient vector excluding β_j. As for the Z^2 values in the table, the individual W values can be used to test the hypotheses that the regression coefficients equal 0, with approximate p-values obtained from $\chi^2_{(1)}$ probabilities. Two sets of semiparametric values are shown, based on estimating functions (8.2.4) with $w_i = 1$ and $w_i = \hat{S}(z_{(i)})$, respectively. These values provide high efficiency when a model (8.0.1) holds, with errors Z that are approximately extreme value and logistic, respectively.

Table 8.2 displays close agreement between the semiparametric and parametric estimates, with agreement particularly close between the parametric and "efficient" semiparametric estimates. There is relatively little difference between the ordinary and robust parametric standard errors. Residual plots and other model checks discussed in Section 6.2 show both the extreme value and logistic location-scale models (6.1.4) with $u(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$ to be consistent with the data, so this is not surprising. The logistic distribution is slightly better supported within the larger log-Burr family of parametric models discussed in Section 6.4.1, but conclusions based on the two models are for all practical purposes identical.

This example illustrates the fact that, when the location-scale framework (8.0.1) is appropriate, the application of fully parametric models along with model checking has a great deal to recommend it. Parametric estimates of regression coefficients are robust, as discussed in Section 6.3.4, and the use of robust variance estimates maintains the validity of tests and confidence intervals under departures from the assumed error distribution. The agreement between inferences based on suitable parametric models and ones based on semiparametric models is close. The much simpler implementation of the parametric methods makes them the approach of choice in most settings.

8.2.2 Rank-Type Procedures with Time-Varying Covariates

In the accelerated failure time model, covariates effectively alter the rate at which time passes. This suggests that if external covariates $\mathbf{x}(t)$ are time-varying, then we might consider models for which the survivor function is of the form

$$Pr(T \geq t|X) = S_0^*\left(\int_0^t \psi[\mathbf{x}(u); \boldsymbol{\beta}]\, du\right), \tag{8.2.10}$$

where $\psi(\cdot; \boldsymbol{\beta})$ is a family of positive-valued functions specified up to a parameter $\boldsymbol{\beta}$, $S_0^*(\cdot)$ is a survivor function defined on $(0, \infty)$, and $X = \{\mathbf{x}(t),\ t \geq 0\}$ is the covariate process history. In the case where $\mathbf{x}(\mathbf{u}) = \mathbf{x}$ is constant, (8.2.10) gives $Pr(T \geq t|\mathbf{x}) = S_0^*(\psi(\mathbf{x}; \boldsymbol{\beta})t)$, which is the standard AFT model considered in Chapter 6 for fully parametric settings.

A convenient form for (8.2.10) in many applications is given by (6.4.18), which we write here as

$$Pr(T \geq t|X) = S_0^*\left(\int_0^t e^{-\boldsymbol{\beta}'\mathbf{x}(u)}\, du\right). \tag{8.2.11}$$

When $\mathbf{x}(u) = \mathbf{x}$ is constant, this gives

$$\begin{aligned} Pr(T \geq t|\mathbf{x}) &= S_0^*(e^{-\boldsymbol{\beta}'\mathbf{x}}t) \\ &= S_0(y - \boldsymbol{\beta}'\mathbf{x}), \end{aligned}$$

where $y = \log t$ and $S_0(y) = S_0^*(e^y)$. This is the location-scale model (8.0.1) for which semiparametric methods, not requiring specification of $S_0(\cdot)$, were discussed in the preceding section. We will consider the model (8.2.11) in the remainder of this section.

Rank-based estimation procedures can be based on the fact that under (8.2.11) and the hypothesis $H: \boldsymbol{\beta} = \boldsymbol{\beta}_0$, the quantities

$$e_i(\boldsymbol{\beta}_0) = \int_0^{t_i} e^{-\boldsymbol{\beta}_0'\mathbf{x}_i(u)}\, du \tag{8.2.12}$$

act as a set of (possibly censored) residuals. In particular, if the failure time random variable T_i replaces t_i in (8.2.12), then the $e_i(\boldsymbol{\beta}_0)$'s are i.i.d. An estimating function analogous to (8.2.3) with the log rank weights $w(z) = 1$ has been considered by Lin and Ying (1995) and others:

$$\mathbf{U}(\boldsymbol{\beta}_0) = \sum_{i=1}^n \delta_i \left\{\mathbf{x}_i(t_i) - \frac{\sum_{\ell=1}^n I[e_\ell(\boldsymbol{\beta}_0) \geq e_i(\boldsymbol{\beta}_0)]\mathbf{x}_\ell[e_\ell^{-1}(e_i(\boldsymbol{\beta}_0))]}{\sum_{\ell=1}^n I[e_\ell(\boldsymbol{\beta}_0) \geq e_i(\boldsymbol{\beta}_0)]}\right\}, \tag{8.2.13}$$

where $e_\ell^{-1}(e_i(\boldsymbol{\beta}_0))$ denotes the real time, t, which corresponds to the value $e_i(\boldsymbol{\beta}_0)$. That is, t satisfies

$$\int_0^t e^{-\boldsymbol{\beta}_0'\mathbf{x}_\ell(u)}\, du = e_i(\boldsymbol{\beta}_0). \tag{8.2.14}$$

Under appropriate conditions, if $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$ then $n^{-1/2}\mathbf{U}(\boldsymbol{\beta}_0)$ is asymptotically normal with mean vector $\mathbf{0}$ and a covariance matrix that is estimated consistently by

$$V(\boldsymbol{\beta}_0) = \frac{1}{n}\sum_{i=1}^n \delta_i \frac{\sum_{\ell=1}^n I(e_\ell \geq e_i)[\mathbf{x}_\ell(t_i) - \bar{\mathbf{x}}(t_i; \boldsymbol{\beta}_0)][\mathbf{x}_\ell(t_i) - \bar{\mathbf{x}}(t_i; \boldsymbol{\beta}_0)]'}{\sum_{\ell=1}^n I(e_\ell \geq e_i)}, \tag{8.2.15}$$

where for simplicity of notation we have written e_ℓ for $e_\ell(\boldsymbol{\beta}_0)$ and

$$\bar{\mathbf{x}}(t_i; \boldsymbol{\beta}_0) = \frac{\sum_{\ell=1}^n I(e_\ell \geq e_i)\mathbf{x}_\ell[e_\ell^{-1}(e_i)]}{\sum_{\ell=1}^n I(e_\ell \geq e_i)}.$$

The estimating function $\mathbf{U}(\boldsymbol{\beta}_0)$ and estimated covariance matrix $V(\boldsymbol{\beta}_0)$ are formally similar to (8.2.4) and (8.2.7) for the fixed covariate case in the preceding section, and confidence intervals can be found by using the procedures described there.

This and other rank-based procedures require that the values of time-varying covariates for all individuals be known for all t that may appear in one of the equations (8.2.13). This and the need to solve (8.2.14) make application of these methods difficult in many situations. Note also that it may be necessary to know values of $\mathbf{x}_\ell(t)$ for $t > t_\ell$.

8.2.3 Discussion

Regression models with independent continuous responses Y_i and covariates $\mathbf{x}_i$ or $X_i = \{\mathbf{x}_i(t), t \geq 0\}$ have the property that certain functions Z_i of Y_i, $\boldsymbol{\beta}$, and $\mathbf{x}_i$ or X_i are i.i.d. In particular, in the location-scale family (8.0.1), the errors $Z_i = Y_i - \boldsymbol{\beta}'\mathbf{x}_i$ are i.i.d. with some distribution function $F_0(z)$, and in the AFT model (8.2.11) with time-varying covariates the quantities $Z_i = e_i(\boldsymbol{\beta})$ given by (8.2.12) with T_i replacing t_i are i.i.d. This opens the door to the use of distribution-free rank procedures based on likelihood or on estimating functions for $\boldsymbol{\beta}$ that effectively measure association between the ranks of the Z_i and characteristics of the covariate values; the Z_i are i.i.d. and thus independent of the covariate values if $\boldsymbol{\beta}$ is the true value and the model family in question is correct. The presence of censoring makes things more difficult, but rank-type procedures that are at least asymptotically distribution-free can often be devised.

The fact that rank-based estimating functions $\mathbf{U}(\boldsymbol{\beta})$ have discontinuities and are hard to characterize creates difficulties for the estimation of $\boldsymbol{\beta}$, as the preceding sections have shown. With the multiplicative hazards models in Chapter 7 it was seen that other approaches to semiparametric estimation could be developed, in particular, methods based on partial likelihood and on semiparametric maximum likelihood. In fact, when the lifetime data are uncensored, these approaches gave procedures that are also rank-based methods. Alternative approaches are also possible with AFT models. We will briefly discuss one of them, which also turns out to be hard to implement. Additional approaches can be found in references in the Bibliographic Notes.

Let us consider the log-location-scale model (8.0.1) with fixed covariates, and suppose temporarily that the distribution function $F_0(z)$ for the Z_i is known. The maximum likelihood estimating function for $\boldsymbol{\beta}$ from a censored random sample (y_i, δ_i), $i = 1, \ldots, n$ is then

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{x}_i \left\{ \delta_i \frac{f_0'(z_i)}{f_0(z_i)} - (1 - \delta_i) \frac{f_0(z_i)}{S_0(z_i)} \right\}, \tag{8.2.16}$$

where $z_i = y_i - \boldsymbol{\beta}'\mathbf{x}_i$ and $f_0(z)$ and $S_0(z)$ are the p.d.f. and survivor function (s.f.) for Z. Defining $\phi(z) = f_0'(z)/f_0(z)$, we can rewrite (8.2.16) as

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{x}_i \left\{ \delta_i \phi(z_i) - (1 - \delta_i) \frac{\int_{z_i}^{\infty} \phi(u)\, dS_0(u)}{S_0(z_i)} \right\} \tag{8.2.17}$$

$$= \sum_{i=1}^{n} \mathbf{x}_i \left\{ \delta_i \phi(z_i) - (1 - \delta_i) E[\phi(Z_i) | Z_i > z_i] \right\}. \tag{8.2.18}$$

This estimating function is not usable since we do not know $F_0(z)$, but (8.2.18) suggests an approach. Suppose that $\phi(z)$ is an arbitrary function of z and that potential log censoring times y_i^* are independent of the Y_i, given $\mathbf{x}_1, \ldots, \mathbf{x}_n$. It then easily follows for estimating functions of the form (8.2.18) that

$$E\{\mathbf{U}(\boldsymbol{\beta})\} = \sum_{i=1}^{n} \mathbf{x}_i E\{\phi(Z_i)\}, \tag{8.2.19}$$

since $\delta_i = 1$ if and only if $Y_i \leq y_i^*$. Since $E\{\phi(Z_i)\}$ is the same for $i = 1, \ldots, n$, $E\{\mathbf{U}(\boldsymbol{\beta})\} = \mathbf{0}$ if $\sum \mathbf{x}_i = \mathbf{0}$, or if $E\{\phi(Z_i)\} = 0$. One approach would be to include an intercept term in $\boldsymbol{\beta}'\mathbf{x}_i$ and restrict $F_0(z)$ by assuming that $E\{\phi(Z)\} = 0$. We suppose instead that there is no intercept; we can then center the covariates so that $\sum \mathbf{x}_i = \mathbf{0}$, and $\mathbf{U}(\boldsymbol{\beta})$ is unbiased. An obvious approach is now to use the estimating function (8.2.19) with the expectations in the second term estimated nonparametrically. This can be done by replacing $S_0(z)$ with the Kaplan–Meier estimate $\hat{S}_0(z)$ based on the z_i:

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{x}_i \left\{ \delta_i \phi(z_i) - (1 - \delta_i) \frac{\int_{z_i}^{\infty} \phi(u) d\hat{S}_0(u)}{\hat{S}_0(z_i)} \right\} = 0. \tag{8.2.20}$$

This can be solved by a two-stage iterative procedure in which $\hat{S}_0(z)$ is the Kaplan–Meier estimate based on the current estimate $\tilde{\boldsymbol{\beta}}$, and then (8.2.20) with $z_i = y_i - \boldsymbol{\beta}'\mathbf{x}_i$ is solved for $\boldsymbol{\beta}$ to give the new current estimate. The choice $\phi(z) = z$ gives a generalization of least squares to the censored data setting. This was suggested by Buckley and James (1979) and has been studied further by James and Smith (1984) and Ritov (1990). Unfortunately, the estimating equations in (8.2.20) are in general discontinuous, and so not particularly easy to deal with. There is consequently no practical advantage for these methods over the rank-based approaches of Section 8.1.1. Ritov (1990) has established an asymptotic equivalence between estimation based on (8.2.20) and the rank-based methods.

The fact that generalized residuals Z_i are i.i.d. if AFT models such as (8.0.1) and (8.2.11) are correctly specified provides opportunities for model checking based on observed residuals like $\hat{z}_i = y_i - \hat{\boldsymbol{\beta}}'\mathbf{x}_i$, as discussed and illustrated for parametric models in Chapter 6. Although semiparametric estimation methods afford protection against misspecification of $S_0(z)$ in (8.0.1) or (8.2.11), which can occur with fully parametric models, the awkwardness of semiparametric estimation and the ability to check parametric assumptions are strong reasons to use fully parametric methods in most situations. Consequently the methods discussed in Chapter 6 tend to dominate in regression settings.

BIBLIOGRAPHIC NOTES

Linear rank tests have been discussed for uncensored data by many authors; Hajek and Sidak (1967), Lehmann (1975), and Hettmansperger (1984) provide book-length accounts of the general theory and applications. The log rank and Wilcoxon tests were considered by Savage (1956) and Wilcoxon (1945), respectively; Kruskal and Wallis (1952) considered m-sample Wilcoxon tests. These tests were later modified for use with right-censored data: Mantel and Haenszel (1959) and Mantel (1963, 1966) considered what is effectively the log rank procedure and Gehan (1965) and Breslow (1970) considered Wilcoxon procedures. Linear rank tests with censored data were considered by Johnson and Mehrotra (1972) and Mehrotra et al. (1977) for Type 2 censored data, and by Peto and Peto (1972), Kalbfleisch and Prentice (1973) and Prentice (1978) for more general types of censoring. Crowley and Thomas (1975), Leurgans (1983, 1984), Struthers (1984), Cuzick (1985), and others have considered these tests further. Empirical studies by Lee et al. (1975), Peace and Flora (1978), Lininger et al. (1979), Latta (1981) and Kellerer and Chmelevsky (1983) have examined the size and power of the tests in various settings. Peto et al. (1976, 1977) give extensive illustrations of the application of rank tests to the design and analysis of clinical trials. Sample size requirements for log rank tests have been considered by Schoenfeld (1981, 1983) and Lakatos (1988). Collett (1994, Ch. 9) provides additional references.

The counting process formulation of rank tests was taken up by Aalen (1978b) and Gill (1980), where martingale theory was deployed as a powerful tool for devising their properties. Andersen et al. (1982) review this topic and show how statistics of the form (8.1.11) can be expressed in the weighted log rank form (8.1.23); see also Mehrotra et al. (1982). Tarone and Ware (1977) and Morton (1978) had previously considered weighted log rank test statistics, which are very easily handled in the counting process framework. Later work has tended to use this approach, and the books by Gill (1980), Fleming and Harrington (1991) and Andersen et al. (1993) give mathematically detailed presentations of the theory.

Other types of nonparametric two-sample tests were mentioned in Section 8.1.5. For additional examples and discussion see, for example, Fleming et al. (1980, 1987), Tarone (1981), Schumacher (1984), Jones and Crowley (1989, 1990), and Pepe and Fleming (1989, 1991). Sequential versions of tests in this chapter have also been considered. This is outside the scope of the book, but see, for example, Koziol and Petkau (1978), Jones and Whitehead (1979), Tsiatis (1981b), Sellke and Siegmund (1983), and Whitehead (1992). Weighted log rank tests and other m-sample procedures for discrete data are considered by Cook and Lawless (1991), Kalbfleisch and Lawless (1991), and Park (1997).

Klein and Moeschberger (1997, Ch. 7) discuss many types of tests for comparing distributions. Oakes (2001, see Secs. 3.2, 4.4, 5.2, 8.3) gives an interesting commentary on these tests.

The use of linear rank procedures for estimating regression coefficients in the model (8.0.1) was considered by Adichie (1967) and Jurečková (1971). Kalbfleisch and Prentice (1980, Ch. 6) discussed the censored data case informally, and Louis

(1981), Tsiatis (1990), Ritov (1990), Wei et al. (1990), Lai and Ying (1991), and Ying (1993) provided more rigorous treatments and discussions of asymptotic properties. The case of time-varying covariates has been discussed by Robins and Tsiatis (1992) and by Lin and Ying (1995). Bagdonavicius and Nikulin (1997, 2001) consider several types of models with time-varying covariates, one of which is (8.0.1).

Buckley and James (1979) generalized least squares to censored data, improving on an earlier attempt by Miller (1976); Currie (1996) discusses computational aspects of these methods. James and Smith (1984) and James (1986) consider more general estimating function approaches discussed in Section 8.2.3.

COMPUTATIONAL NOTES

Many software packages implement the weighted log rank tests discussed in Sections 8.1.1–8.1.5. A word of caution is appropriate, because different packages may use different version of certain tests (e.g., the generalized Wilcoxon). In addition, some packages employ permutation-based variance estimates that are appropriate only when the censoring processes for the distributions in question are the same.

PROBLEMS AND SUPPLEMENTS

8.1 *Linear rank tests.*

(a) Consider the scores α_i given by (8.1.5) for the case of uncensored data. Show that α_i can be expressed as

$$\alpha_i = E\{\phi(U_{(i)})\}, \qquad i = 1, \ldots, n,$$

where $U_{(1)} < \cdots < U_{(n)}$ are the order statistics from a random sample of size n from the distribution $U(0, 1)$, and

$$\phi(x) = \frac{-g'[G^{-1}(x)]}{G'[G^{-1}(x)]},$$

where $G(y)$ is the cumulative distribution function corresponding to $g(y)$.

(b) It can be shown that an asymptotically equivalent linear rank statistic to (8.1.4) with (8.1.5) is given by (8.1.4) with scores defined by

$$\alpha_i' = \phi\left(\frac{i}{n+1}\right), \qquad i = 1, \ldots, n. \tag{8.3.1}$$

(Note that $E\{U_{(i)}\} = i/(n+1)$.) Compare the values α_i' with the α_i for the log rank and Wilcoxon tests in Examples 8.1.1 and 8.1.2, which correspond to extreme value and logistic models for Y.

(c) For censored samples, let $t_{(1)} < \cdots < t_{(k)}$ be the distinct observed failure times. A common suggestion is to define the score α_i' for $t_{(i)}$ by

$$\alpha_i' = \phi[1 - \tilde{S}(t_{(i)})], \qquad i = 1, \ldots, k, \tag{8.3.2}$$

where $\tilde{S}(t)$ is an estimate of $S(t)$, the common survivor function for $T = \exp Y$ when $\boldsymbol{\theta} = \mathbf{0}$ in (8.1.2). Suppose now that a censored observation t^* lies in the interval $[t_{(i)}, t_{(i+1)})$. A common score assignment for t^* is then

$$\frac{1}{1-u}\int_u^1 \phi(x)\,dx\,\Big|_{u=1-\tilde{S}(t_{(i)})}, \tag{8.3.3}$$

motivated by the fact that we know the lifetime corresponding to t^* exceeds $t_{(i)}$. Obtain (8.3.2) and (8.3.3) for the cases where $G(y)$ has standard extreme value and logistic form, respectively.

(Sections 8.1.1–8.1.3)

8.2 *Equivalence of two forms of linear rank statistics.* Consider the statistics (8.1.23) in the case where there are no tied failure times:

$$W_\ell = \sum_{i=1}^{k} w_i(d_{\ell i} - n_{\ell i}/n_i) \qquad \ell = 1, \ldots, m-1, \tag{8.3.4}$$

where $w_i = w(t_{(i)})$ and $d_{\ell i} = 1$ iff the failure at $t_{(i)}$ is from group ℓ.

(a) Show that (8.3.4) can be written in the form (8.1.11):

$$W_\ell = \sum_{i=1}^{k}[\alpha_i d_{\ell i} + a_i(n_{\ell i} - 1 - n_{\ell,i+1})], \qquad \ell = 1, \ldots, m-1, \tag{8.3.5}$$

where, for $i = 1, \ldots, k$, α_i and a_i are defined as

$$\alpha_i = w_i - \sum_{\ell=1}^{i} w_\ell/n_\ell, \qquad a_i = -\sum_{\ell=1}^{i} w_\ell/n_\ell. \tag{8.3.6}$$

Note that the α_i, a_i values satisfy

$$\alpha_i = n_i a_{i-1} - (n_i - 1)a_i \qquad i = 1, \ldots, k, \tag{8.3.7}$$

with a_0 defined as 0.

(b) Show conversely that if W_ℓ has the form (8.3.5) for scores α_i, a_i that satisfy (8.3.7), then it can be rewritten in the form (8.3.4), with $w_i = \alpha_i - a_i$ $(i = 1, \ldots, k)$.

(Section 8.1.4; Prentice and Marek 1979; Mehrotra et al. 1982)

8.3 Consider m-sample tests based on statistics of the weighted log rank form (8.1.22) or (8.1.23). Relate the ith term of (8.1.24) to the covariance matrix

for the multiple hypergeometric observation $(d_{1i}, \ldots, d_{m-1,i})$ with probability function

$$Pr(d_{1i}, \ldots, d_{m-1,i}) = \prod_{\ell=1}^{m} \binom{n_{\ell i}}{d_{\ell i}} \Big/ \binom{n_i}{d_i}.$$

Link this to the application of the martingale result (F16), from which (8.1.24) can be obtained. Examine whether (8.1.24) is an unbiased estimator of the exact covariance matrix of $(W_1, \ldots, W_{m-1})$ in (8.1.23).

(Section 8.1.4)

8.4 Consider the data on times to a first pulmonary exacerbation for persons with cystic fibrosis, introduced in Example 1.1.8 and subsequently discussed in Examples 3.2.4, 3.2.5, 6.2.3, and 6.3.4.

(a) Carry out tests of no difference in the distributions of time to first exacerbation for the Placebo and Treatment (rhDNase) populations using weighted log rank procedures, with no adjustment for forced expiratory volume (fev).

(b) Carry out stratified versions of the tests in part (a), by dividing subjects into three strata based on their fev values. Form strata of approximately equal sizes by using the .33 and .67 quantiles of the fev measurements as cut points. Compare the p-values for these tests with the corresponding tests in part (a), and with the parametric test for no treatment effect based on the analysis in Example 6.3.4.

(c) How do you feel most comfortable quantifying the treatment effect?

(Section 8.1)

8.5 Consider the two-sample electrical insulation failure time data of Example 5.4.2, for which a log rank test was used in Problem 7.7. Carry out a test of no difference in the two lifetime distributions using the generalized Wilcoxon test, and compare it with the log rank result.

(Section 8.1)

8.6 Consider the data on times to failure of five types of ball bearings, discussed in Problem 5.8. Carry out rank-based tests of equality for the five failure time distributions. Compare results with those of the parametric analysis in Problem 5.8, and discuss the pros and cons of the rank-based and the parametric methods.

(Sections 8.1, 5.4)

8.7 Derive a log rank trend test statistic (8.1.27) from a suitably defined proportional hazards model with $H_\ell(t) = H_0(t)\exp(\beta x_\ell)$, for $\ell = 1, \ldots, m$. Discuss how to develop this test as well as others of linear rank form from the approach described in Section 8.1.

(Sections 8.1.1, 8.1.4)

8.8 The following is another way of looking at the Wilcoxon test. Suppose for simplicity that there is no censoring and let Y_{1i} $(i = 1, \ldots, n_1)$ and Y_{2i} $(i = 1, \ldots, n_2)$ be independent random samples from two continuous distributions with survivor functions $S_1(y)$ and $S_2(y)$. Define

$$U_{ij} = \begin{cases} 1 & \text{if } Y_{2j} \le Y_{1i} \\ -1 & \text{if } Y_{2j} > Y_{1i} \end{cases} \qquad i = 1, \ldots, n_1 \qquad j = 1, \ldots, n_2,$$

and let

$$W = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} U_{ij}.$$

(a) Show that $W = (n+1)U(0)$, where $U(0)$ is the Wilcoxon score statistic given in Example 8.1.2.

(b) Show under H_0: $S_1(y) = S_2(y)$ that $E(W) = 0$ and $\text{Var}(W) = n_1 n_2 (n_1 + n_2 + 1)/12$.

Gehan (1965), Efron (1967), and Breslow (1970) discuss generalized Wilcoxon tests for censored data from this point of view.

(Sections 8.1.1, 8.1.3)

8.9 Specialize (8.2.7) to the m-sample setting discussed in Section 8.1.1, where $\mathbf{x}_i$ is a vector of $m - 1$ indicator covariates and $\boldsymbol{\beta} = \mathbf{0}$ if the m distributions are identical. Show that $V(\mathbf{0})$ given by (8.2.7) is not the same as (8.1.24) when there are tied lifetimes.

(Section 8.2.1)

8.10 *Confidence intervals for a location difference.* Consider the model (8.0.1) with a single indicator covariate x indicating whether an observation comes from distribution 1 $(x = 1)$ or 2 $(x = 0)$. By estimating β we are therefore estimating the difference in location of two distributions with the same shape; this is equal to the difference in the means or in specified quantiles of the distributions.

(a) Use the rank procedure in Section 8.2.1 with a log rank weight function $w(z) = 1$ to estimate β for the two distributions in Example 8.1.1. In particular, obtain a two-sided .95 confidence interval for β based on (8.2.8). Compare this with a confidence interval for the difference in quantiles $\beta = u_1 - u_2$, based on the assumption of two extreme value distributions for log-lifetime, $EV(u_1, b)$ and $EV(u_2, b)$.

(b) Obtain an estimate $\hat{\beta}$ for β using what you consider a reasonable procedure.

(Section 8.2.1; Louis 1981)

CHAPTER 9

Multiple Modes of Failure

9.1 INTRODUCTION

As discussed in Section 1.5, individuals can in some settings fail in different ways, and are then assigned a mode of failure. The modes may refer to the cause of failure, in which case they are often termed competing risks. For example, an individual in a demographic study might be recorded as dying at age t from one of cancer, cardiovascular disease, or other causes. Failure modes can also be defined in other ways, for example, to reflect costs or severity of consequences associated with failure. This section describes the basic characteristics of multiple failure mode problems, following which methods of inference are considered.

9.1.1 Basic Characteristics and Model Specification

A pair (T, C) is defined for each individual, with T the failure time and C the failure mode. It is assumed for now that T is a continuous random variable and that C takes on values in the set $\{1, \ldots, k\}$. The joint distribution of T and C can be specified in a variety of ways, but it is particularly convenient to do this via mode-specific hazard functions given by (1.5.2):

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{Pr(T < t + \Delta t, C = j | T \ge t)}{\Delta t} \qquad j = 1, \ldots, k. \qquad (9.1.1)$$

These functions fully specify the distribution of (T, C). The marginal hazard function for T is

$$\lambda(t) = \sum_{j=1}^{k} \lambda_j(t)$$

and the marginal survivor function for T is therefore

$$S(t) = Pr(T \ge t) = e^{-\Lambda(t)}, \qquad (9.1.2)$$

where $\Lambda(t) = \int_0^t \lambda(u)\,du$ is the cumulative hazard function (c.h.f.) for T. Of course,

$$\Lambda(t) = \sum_{j=1}^{k} \int_0^t \lambda_j(u)\,du = \sum_{j=1}^{k} \Lambda_j(t).$$

Since $Pr(T \in [t, t+\Delta t), C = j) = Pr[T < t + \Delta t, C = j | T \geq t] Pr(T \geq t)$, it follows that

$$F_j(t) = Pr(T \leq t, C = j) = \int_0^t \lambda_j(u) S(u)\,du, \tag{9.1.3}$$

and $f_j(t) = F_j'(t) = \lambda_j(t)S(t)$ are the subdistribution and subdensity functions for mode j failures.

The $F_j(t)$'s or $f_j(t)$'s also specify the distribution of (T, C). Note that

$$\pi_j = Pr(C = j) = F_j(\infty), \qquad j = 1, \ldots, k \tag{9.1.4}$$

$$F(t) = 1 - S(t) = \sum_{j=1}^{k} F_j(t), \tag{9.1.5}$$

and that $\lambda_j(t) = f_j(t)/S(t)$.

The mode-specific hazard functions $\lambda_j(t)$ often have intuitive and scientific appeal. In human mortality studies, for example, they represent the mortality rates from specific causes at age t, conditional on survival up to age t. Various probabilities are also typically of interest; these include the $F_j(t)$'s, which are sometimes referred to as (mode-specific) cumulative incidence functions, and conditional probabilities such as

$$F_j^*(t) = Pr(T \leq t | C = j) = \frac{1}{\pi_j} F_j(t). \tag{9.1.6}$$

Parametric models can be specified in different ways. The most common approach is to specify the $\lambda_j(t)$ parametrically, but one could instead specify the $F_j^*(t)$'s and π_j. The one set of functions can be obtained from the other, but the interpretability of individual parameters is tied to the approach taken.

Example 9.1.1. The common approach is to choose convenient parametric specifications for the mode-specific hazard functions (9.1.1). For example, with no covariates present, Weibull parameterizations

$$\lambda_j(t; \alpha_j, \beta_j) = \frac{\beta_j}{\alpha_j} \left(\frac{t}{\alpha_j}\right)^{\beta_j - 1} \qquad j = 1, \ldots, k \tag{9.1.7}$$

are often useful; to incorporate covariates, multiplicative hazards assumptions such as

$$\lambda_j(t|\mathbf{x}) = \lambda_{0j}(t) e^{\beta_j' \mathbf{x}}$$

are often made. These types of models are easy to fit with standard survival analysis software, as we will see in Section 9.1.2. However, the subdistributions $F_j(t)$ in (9.1.3), and other quantities that might be of interest, are complex functions of all of the parameters.

Another approach would be to specify the distributions $F_j^*(t)$ in (9.1.6) parametrically, and to treat the π_j as additional parameters. For example, we might adopt Weibull distributions with hazard functions $h_j^*(t)$ of the same form as the right side of (9.1.7), in which case the parameters all have rather direct interpretations. This model has more parameters than the one represented by (9.1.7), and it cannot be as easily fitted using standard software. The fact that

$$\lambda_j(t) = \frac{\pi_j f_j^*(t)}{1 - \sum_{\ell=1}^{k} \pi_\ell F_\ell^*(t)} \qquad j = 1, \ldots, k \tag{9.1.8}$$

also makes it clear that the mode-specific hazard functions have complex forms for these types of model.

The functions just introduced fully describe the distribution of (T, C) in multiple failure mode settings. A physical process involving additional structure is sometimes relevant: this is called the series system or latent failure times model. Consider a system with k essential physical components, each of which is liable to fail, and let T_j denote the lifetime or failure time of component j $(j = 1, \ldots, k)$. The system fails when the first component fails, so its lifetime is $T = \min(T_1, \ldots, T_k)$. We can also identify a failure mode that indicates which component failed with each system: $C = j$ such that $T = T_j$. This framework seems interesting, since it appears that we can consider multivariate models $F(t_1, \ldots, t_k)$ for the joint distribution of $T_1, \ldots, T_k$. Within a series system this is entirely notional, however, since all that is ever realized or observed is the pair (T, C). Even if the joint distribution is considered meaningful, it is inestimable solely on the basis of observations (T_i, C_i): it is easily seen that two different distributions $F(t_1, \ldots, t_k)$ can give the same distribution for (T, C). It is also impossible to determine whether $T_1, \ldots, T_k$ are mutually independent or not; for every distribution having nonindependent T_j, there exists a distribution with independent T_j that gives the same distribution of (T, C). Problems 9.1, 9.2, and 9.3 amplify this point.

9.1.2 Likelihood Function Formulation

Suppose that observations are taken on a random sample of n individuals and that right censoring is possible; we ignore covariates for now. If T_i is censored at t_i, the eventual failure mode is unknown, so the observed data for individual i consist of either $(T_i = t_i, C_i)$ or $T_i > t_i$. The likelihood function under the assumption of independent censoring is therefore

$$L = \prod_{i=1}^{n} f_{C_i}(t_i)^{\delta_i} S(t_i)^{1-\delta_i}, \tag{9.1.9}$$

where $\delta_i = 1$ if t_i is a failure time and 0 if it is a censoring time. To explore this further, note from (9.1.2) that

$$S(t) = \exp\left\{-\sum_{j=1}^{k} \Lambda_j(t)\right\}$$

$$= \prod_{j=1}^{k} G_j(t), \tag{9.1.10}$$

where $G_j(t) = \exp\{-\Lambda_j(t)\}$. The functions $G_j(t)$, $j = 1, \ldots, k$ have the mathematical properties of continuous survivor functions, but they are not the survivor functions of any observable random variables. However, the likelihood (9.1.9) can be rewritten, using the notation $\delta_{ij} = I(C_i = j)$, as

$$L = \prod_{i=1}^{n}\prod_{j=1}^{k} f_j(t_i)^{\delta_{ij}} S(t_i)^{1-\delta_i}$$

$$= \prod_{i=1}^{n}\prod_{j=1}^{k} [\lambda_j(t_i) S(t_i)]^{\delta_{ij}} S(t_i)^{1-\delta_i}$$

$$= \prod_{i=1}^{n}\prod_{j=1}^{k} g_j(t_i)^{\delta_{ij}} G_j(t_i)^{1-\delta_{ij}}, \tag{9.1.11}$$

where

$$g_j(t) = \lambda_j(t) G_j(t) = -G_j'(t). \tag{9.1.12}$$

By reversing the order of the products in (9.1.11) we see that L has the same form as a product of k likelihood functions, each of which represents a censored sample associated with a particular failure mode. The jth part in the product,

$$L_j = \prod_{i=1}^{n} g_j(t_i)^{\delta_{ij}} G_j(t_i)^{1-\delta_{ij}}, \tag{9.1.13}$$

is exactly of the standard form (2.2.3) for a lifetime distribution with probability density function (p.d.f.) $g_j(t)$, and survivor function (s.f.) $G_j(t)$, though as mentioned, these functions do not here correspond to any observable random variable.

The form of the likelihood function, L, in (9.1.11) indicates that the $\lambda_j(t)$'s or $\Lambda_j(t)$'s are estimable from data on (T, C). Furthermore, if the $\lambda_j(t)$'s involve separate parameters $\boldsymbol{\theta}_j$ for $j = 1, \ldots, k$ (and so likewise for the $G_j(t)$'s and $g_j(t)$'s), then $L(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k)$ factors into separate pieces $L_j(\boldsymbol{\theta}_j)$. Inference for $\boldsymbol{\theta}_j$ can therefore be based on (9.1.13); for a given j, this amounts to treating a failure of mode j at t_i as a failure, but a failure of any other mode at t_i as censoring. Parametric inference for models of this type is easily implemented using standard survival analysis

methods and software. Similarly, nonparametric and semiparametric methods can be implemented for models based on the $\lambda_j(t)$'s. It will be noted, on the other hand, that unless there is no censoring the likelihood function does not factor for parametric models in which the conditional distributions $F_j^*(t)$ in (9.1.6) involve separate parameters $\boldsymbol{\theta}_j$; this follows from (9.1.8) and (9.1.11).

Let us also consider briefly the form of the likelihood function when data on lifetimes are discrete or grouped. In particular, suppose that intervals $I_\ell = (a_{\ell-1}, a_\ell]$ are defined for $\ell = 1, \ldots, m+1$, with $0 = a_0 < a_1 < \cdots < a_m < a_{m+1} = \infty$, and assume that the numbers of individuals failing by modes $1, \ldots, k$ in each of the intervals $I_1, \ldots, I_m$ is observed. It is assumed further that there is no censoring except possibly in I_{m+1}; the case where censoring may occur in intervals $I_1, \ldots, I_m$ is discussed in Section 9.3 and Problems 9.11 and 9.12.

Define for $\ell = 1, \ldots, m$ and $j = 1, \ldots, k$

$$\pi_{j\ell} = Pr\{\text{individual fails in } I_\ell \text{ by mode } j\} \tag{9.1.14}$$

$$= \int_{a_{\ell-1}}^{a_\ell} f_j(u)\, du$$

$$= \pi_j \int_{a_{\ell-1}}^{a_\ell} f_j^*(u)\, du.$$

The likelihood function based on n independent individuals is then

$$L = \left\{\prod_{\ell=1}^{m}\prod_{j=1}^{k} \pi_{j\ell}^{d_{j\ell}}\right\} S(a_m)^{d_{m+1}}, \tag{9.1.15}$$

where $d_{j\ell}$ is the number of mode j failures in I_ℓ and

$$d_{m+1} = n - \sum_{\ell=1}^{m}\sum_{j=1}^{k} d_{j\ell}$$

is the number of individuals surviving to a_ℓ. The likelihood is of multinomial form, and if the data arose from grouping continuous lifetimes, then models in which the distributions $F_j^*(t)$ are parameterized separately are reasonably convenient. The case of discrete-lifetime distributions is also covered by the formulation in (9.1.15); in this case, the definition (9.1.14) holds, but the two lines following it are irrelevant.

The next three sections consider nonparametric, parametric, and semiparametric methods for multiple failure modes data.

9.2 NONPARAMETRIC METHODS

Consider a censored random sample of (T_i, C_i)'s, as described in Section 9.1.2. Based on the fact that the likelihood function (9.1.1) factors into separate pieces

(9.1.13) for each failure mode, nonparametric estimation is straightforward. In particular, the piece (9.1.13) has the mathematical form of a censored data likelihood for a lifetime distribution with survivor, density, and hazard functions $G_j(t)$, $g_j(t)$, and $\lambda_j(t)$. Consequently, $G_j(t)$ could be estimated using the Kaplan–Meier estimate based on the data (t_i, δ_{ij}), $i = 1, \ldots, n$. Since $G_j(t)$ is not the s.f. for any random variable, however, this is of limited direct interest, and it is more useful to consider the cumulative mode-specific hazard function $\Lambda_j(t)$. This equals $-\log G_j(t)$, and so could be estimated from the Kaplan–Meier estimate of $G_j(t)$, but it is more common to use the Nelson–Aalen estimator for a c.h.f. By (3.2.13), this takes the form

$$\hat{\Lambda}_j(t) = \sum_{i:t_i \le t} \frac{\delta_{ij}}{n_i}, \qquad j = 1, \ldots, k, \tag{9.2.1}$$

where $Y_\ell(t) = I(t_\ell \ge t)$ and $n_i = \sum_{\ell=1}^n Y_\ell(t_i)$ is the number of individuals alive and uncensored just prior to time t_i. The variance estimate from (3.2.15) is typically used with (9.2.1): this is

$$\widehat{\mathrm{Var}}[\hat{\Lambda}_j(t)] = \sum_{i:t_i \le t} \frac{\delta_{ij}}{n_i^2}. \tag{9.2.2}$$

The marginal s.f. $S(t)$ for T is easily estimated by ignoring the associated failure modes and using the Kaplan–Meier estimate based on the data (t_i, δ_i), $i = 1, \ldots, n$. This gives

$$\hat{S}(t) = \prod_{i:t_{(i)} < t} \left(\frac{n_i' - d_i'}{n_i'} \right) \tag{9.2.3}$$

as in (3.2.2), where $t_{(1)} < \cdots < t_{(k)}$ are the distinct times at which failures occur, and d_i' and n_i' are the numbers of failures and individuals at risk at $t_{(i)}$, respectively. The variance estimate (3.2.3) can be used in association with $\hat{S}(t)$. An alternative estimate of $S(t)$ is

$$\hat{S}(t) = \exp[-\hat{\Lambda}(t)] = \exp\left[-\sum_{j=1}^k \hat{\Lambda}_j(t) \right], \tag{9.2.4}$$

with the $\hat{\Lambda}_j(t)$'s given by (9.2.1).

The subdistribution or cumulative incidence functions $F_j(t)$ of (9.1.3) are also of interest. In view of (9.1.3), a natural nonparametric estimate is, in counting process notation,

$$\hat{F}_j(t) = \int_0^t \hat{S}(u)\, d\hat{\Lambda}_j(u).$$

This gives

$$\hat{F}_j(t) = \sum_{i:t_i \le t} \hat{S}(t_i) \cdot \frac{\delta_{ij}}{n_i} \qquad j = 1, \ldots, k. \tag{9.2.5}$$

If there is only a single failure mode, then (9.2.5) reduces to $1 - \hat{S}(t+)$, with $\hat{S}(t)$ as in (9.2.3). If there are $k \ge 2$ failure modes, but no censoring, then $\hat{F}_j(t)$ equals the fraction of individuals with $t_i \le t$ and $C_i = j$, that is, the empirical subdistribution function for mode j. When there is no censoring, $\sum \hat{F}_j(t)$ equals $1 - \hat{S}(t+)$, but this is not the case generally.

Variance estimates for $\hat{F}_j(t)$ can be obtained by using extensions of counting process ideas discussed in Section 3.2.4; Andersen et al. (1993, pp. 298–304) and references therein can be consulted for this approach. An alternative is to consider a model in which the hazard functions $\lambda_j(t)$ are piecewise-constant, as in Problem 3.7, Example 6.5.1, and Section 7.4. This allows parametric maximum likelihood methods to be employed and, as discussed in Section 7.4, it is also possible to obtain consistent nonparametric estimates of parameters, and associated variance estimates, by letting the number of intervals increase while interval lengths approach zero. The development is outlined below.

Define intervals $I_\ell = [a_{\ell-1}, a_\ell)$ for $\ell = 1, \ldots, m$, with $0 = a_0 < a_1 < \cdots < a_m$, where a_m is some suitably large value. Assume that the $\lambda_j(t)$'s are piecewise-constant,

$$\lambda_j(t) = \lambda_{j\ell}, \qquad t \in I_\ell, \tag{9.2.6}$$

for $j = 1, \ldots, k$ and $\ell = 1, \ldots, m$. The c.h.f.'s are then

$$\Lambda_j(t) = \sum_{\ell=1}^m \lambda_{j\ell} \Delta_\ell(t), \tag{9.2.7}$$

where $\Delta_\ell(t)$, as in (1.3.26), is the length of the intersection of I_ℓ and $[0, t)$. From (9.1.11), the likelihood function can be written as

$$L = \prod_{j=1}^k \prod_{i=1}^n \lambda_j(t_i)^{\delta_{ij}} e^{-\Lambda_j(t_i)},$$

and so

$$\frac{\partial \log L}{\partial \lambda_{j\ell}} = \frac{d_{j\ell}}{\lambda_{j\ell}} - \Delta_\ell,$$

where $\Delta_\ell = \sum_{i=1}^n \Delta_\ell(t_i)$ and $d_{j\ell} = \sum_{i=1}^n I(t_i \in I_\ell)\delta_{ij}$. This immediately gives the maximum likelihood estimates (m.l.e.'s)

$$\hat{\lambda}_{j\ell} = d_{j\ell}/\Delta_\ell \qquad j = 1, \ldots, k; \qquad \ell = 1, \ldots, m \tag{9.2.8}$$

and it also follows from taking second derivatives of $\log L$ that the $\hat{\lambda}_{j\ell}$'s are asymptotically independent with asymptotic variances estimated by

$$\widehat{\text{Var}}(\hat{\lambda}_{j\ell}) = \frac{d_{j\ell}}{\Delta_\ell^2}. \tag{9.2.9}$$

Now consider $\hat{F}_j(t)$ based on (9.1.3). An asymptotic variance can be obtained by using Theorem B2 in Appendix B. For this we require the derivatives

$$w_{r\ell}^{(j)} = \partial F_j(t)/\partial \lambda_{r\ell}.$$

Differentiating through the integral sign in (9.1.3), we find

$$w_{r\ell}^{(j)} = \int_0^t \{I(j=r)S(u)I(u \in I_\ell) - \Delta_\ell(u)\lambda_j(u)S(u)\}\,du. \tag{9.2.10}$$

We can obtain a variance estimate for $\hat{F}_j(t)$ by computing the $w_{r\ell}^{(j)}$ using estimates $\hat{\lambda}_{j\ell}$ for $\lambda_{j\ell}$. With the vector $\hat{\mathbf{w}}^{(j)}$ representing the $\hat{w}_{r\ell}^{(j)}$ in some specified order, and the vector $\hat{\mathbf{v}}$ representing the variances (9.2.9) in the same order, we get

$$\widehat{\text{Var}}[\hat{F}_j(t)] = \hat{\mathbf{w}}^{(j)\prime}\,\text{diag}(\hat{\mathbf{v}})\hat{\mathbf{w}}^{(j)} \tag{9.2.11}$$

as an estimate, where diag $(\mathbf{v})$ is an $mk \times mk$ diagonal matrix with the entries of $\mathbf{v}$ on the diagonal.

Let us also take the limit as $m \to \infty$ and interval lengths approach 0, in order to get a variance estimate for the nonparametric estimate (9.2.5). With $\Delta_\ell^* = |a_\ell - a_{\ell-1}|$ small, we have $\Delta_\ell(t) \doteq \Delta_\ell^* I(t \geq a_\ell)$, and from (9.2.10),

$$\begin{aligned} w_{r\ell}^{(j)} &\doteq I(j=r)\Delta_\ell^* S(a_\ell)I(t \geq a_\ell) - \sum_{u:a_u \leq t} S(a_u)\Delta_\ell(a_u)\lambda_j(a_u)\Delta_u^* \\ &\doteq I(t \geq a_\ell)\Delta_\ell^* \left\{ I(j=r)S(a_\ell) - \sum_{a_u=a_\ell}^{t} f_j(a_u)\Delta_u^* \right\} \\ &\doteq I(t \geq a_\ell)\Delta_\ell^* \left\{ I(j=r)S(a_\ell) - [F_j(t) - F_j(a_\ell)] \right\}. \end{aligned}$$

This gives, via (9.2.11),

$$\widehat{\text{Var}}[\hat{F}_j(t)] \doteq \sum_{r=1}^{k}\sum_{\ell=1}^{m} I(t \geq a_\ell)(\Delta_\ell^*)^2\{I(j=r)\hat{S}(a_\ell) - [\hat{F}_j(t) - \hat{F}_j(a_\ell)]\}^2 \frac{d_{j\ell}}{\Delta_\ell^2}.$$

Noting that $\Delta_\ell \doteq \Delta_\ell^* Y_\cdot(a_\ell)$, where $Y_\cdot(u) = \sum I(t_i \geq u)$ is the number of individuals at risk at time u, we obtain, upon taking the limit as the $\Delta_\ell^* \to 0$ and $m \to \infty$,

the variance estimate

$$\widehat{\text{Var}}[\hat{F}_j(t)] = \sum_{r=1}^{k}\int_0^t \hat{S}(u)^2 \left\{ I(j=r) - \frac{[\hat{F}_j(t) - \hat{F}_j(u)]}{\hat{S}(u)} \right\}^2 \frac{dN_j(u)}{Y_\cdot(u)^2}, \tag{9.2.12}$$

where $dN_j(u)$ is the number of failures of mode j at time u.

The parameters π_j can by (9.1.4) be estimated as

$$\hat{\pi}_j = \hat{F}_j(\infty) \qquad j = 1, \ldots, k. \tag{9.2.13}$$

However, by (9.2.5) this is not estimable unless the largest observed time is a failure time τ, in which case $\hat{S}(t) = 0$ for $t > \tau$; if the largest time is a censoring time, $\hat{S}(t)$ is undefined beyond it, as discussed in Section 3.2.1. Even when $\hat{S}(t) = 0$ for $t > \tau$, the estimates $\hat{\pi}_j$ do not satisfy $\sum \hat{\pi}_j = 1$, except when there is no censoring. A common procedure is to renormalize the estimates as $\hat{\pi}_j' = \hat{\pi}_j / \sum_{\ell=1}^{k} \hat{\pi}_\ell$ in that case. In general, we can, of course, estimate $P(C = j | T \leq \tau)$, which for large τ is essentially as useful as estimating π_j.

Finally, nonparametric estimation of cumulative hazards can still be done when lifetimes are left truncated, exactly as discussed for the Nelson–Aalen and Kaplan–Meier estimates in Section 3.5.1. If the lifetimes of all individuals are truncated, then only the functions $\Lambda_j(t) - \Lambda_j(u_{\min})$ are estimable, where $u_{\min}$ is the minimum left-truncation time in the data.

Example 9.2.1. Example 1.1.10 gave data from a life test on a small appliance. Failures were classified into 18 different modes, though among the 33 observed failures only 7 modes were represented, and only modes 6 and 9 appeared more than twice. We will focus here on failure mode 9 for an illustration of nonparametric methods. Therefore, define $C = 1$ if failure occurs by mode 9 and $C = 2$ if it occurs by any other mode; let $C = 0$ denote that the failure time is censored.

Table 9.1 shows the ordered failure times and failure modes, along with the Nelson–Aalen estimates $\hat{\Lambda}_1(t)$ and $\hat{\Lambda}_2(t)$ given by (9.2.1). The Kaplan–Meier estimate (9.2.3) of $S(t)$ and estimated cumulative incidence functions $\hat{F}_1(t)$ and $\hat{F}_2(t)$ from (9.2.5) are also shown. Plots of the $\hat{\Lambda}_j(t)$'s and $\hat{F}_j(t)$'s are given in Figures 9.1 and 9.2, respectively.

The cumulative mode-specific hazard functions $\hat{\Lambda}_j(t)$ show, as is apparent from the raw data, that mode 2 failures predominate early and mode 1 failures predominate later on. The slopes of the plots in Figure 9.1 provide rough estimates of the hazard functions $\lambda_1(t)$ and $\lambda_2(t)$. Figure 9.2 shows the way that failures of the two types accumulate over time. The estimated marginal probabilities π_1 and π_2 of mode 1 and mode 2 failures cannot, strictly speaking, be estimated, since the largest observation is a censoring time. However, the probabilities $P(C = j | T \leq 13{,}403)$ are estimated by $\hat{F}_j(13{,}403)/[\hat{F}_1(13{,}403) + \hat{F}_2(13{,}403)]$ for $j = 1, 2$. The values $\hat{F}_1(13{,}403) = .509$ and $\hat{F}_2(13{,}403) = .451$ then give the values .530 and .470. For all intents and purposes these can be regarded as estimates of π_1 and π_2.

Table 9.1. Multiple Failure Mode Estimates for Appliance Life Test Data

t_i	C_i	$\hat{\Lambda}_1(t_i)$	$\hat{\Lambda}_2(t_i)$	$\hat{S}(t_i)$	$\hat{F}_1(t_i)$	$\hat{F}_2(t_i)$
11	2	.000	.028	1.000	.000	.028
35	2	.000	.056	.972	.000	.056
49	2	.000	.086	.944	.000	.083
170	2	.000	.116	.917	.000	.111
329	2	.000	.147	.889	.000	.139
381	2	.000	.180	.861	.000	.167
708	2	.000	.213	.833	.000	.194
958	2	.000	.247	.806	.000	.222
1,062	2	.000	.283	.778	.000	.250
1,167	1	.037	.283	.750	.028	.250
1,594	2	.037	.322	.722	.028	.278
1,925	1	.077	.322	.694	.056	.278
1,990	1	.119	.322	.667	.083	.278
2,223	1	.162	.322	.639	.111	.278
2,327	2	.162	.367	.611	.111	.306
2,400	1	.210	.367	.583	.139	.306
2,451	2	.210	.417	.556	.139	.333
2,471	1	.262	.417	.528	.167	.333
2,551	1	.318	.417	.500	.194	.333
2,565	0	.318	.417	.472	.194	.333
2,568	1	.380	.417	.472	.224	.333
2,694	1	.447	.417	.443	.253	.333
2,702	2	.447	.488	.413	.253	.363
2,761	2	.447	.565	.384	.253	.392
2,831	2	.447	.649	.354	.253	.422
3,034	1	.538	.649	.325	.283	.422
3,059	2	.538	.749	.295	.283	.451
3,112	1	.649	.749	.265	.313	.451
3,214	1	.774	.749	.236	.342	.451
3,478	1	.917	.749	.207	.372	.451
3,504	1	1.084	.749	.177	.401	.451
4,329	1	1.284	.749	.148	.431	.451
6,367	0	1.284	.749	.118	.431	.451
6,976	1	1.617	.749	.118	.470	.451
7,846	1	2.117	.749	.079	.509	.451
13,403	0	2.117	.749	.039	.509	.451

The life test that gave these data was carried out during the development of the appliance in question, and it was of interest to consider what effect the removal of certain failure modes would have on the overall failure time distribution, $S(t)$. This is often done by assuming that removal of a mode of failure reduces the hazard function for that mode to 0, and leaves the hazard functions for other modes unchanged. This assumption cannot be checked on the basis of the observed data, however, and is usually unrealistic, because removal of a failure mode involves changes that also affect

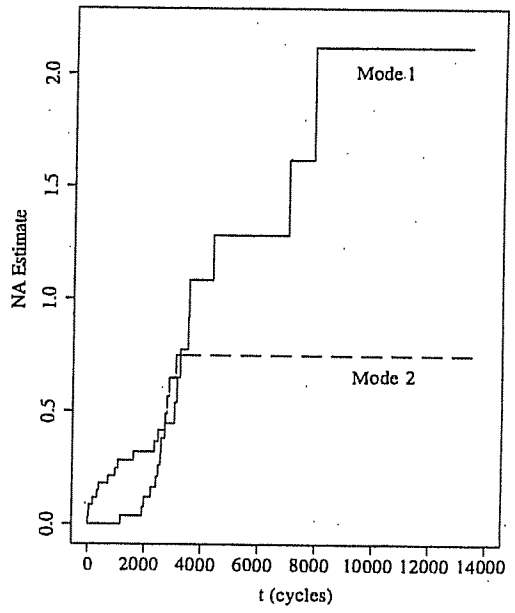

Figure 9.1. NA estimates of cumulative mode-specific hazards, appliance test data.

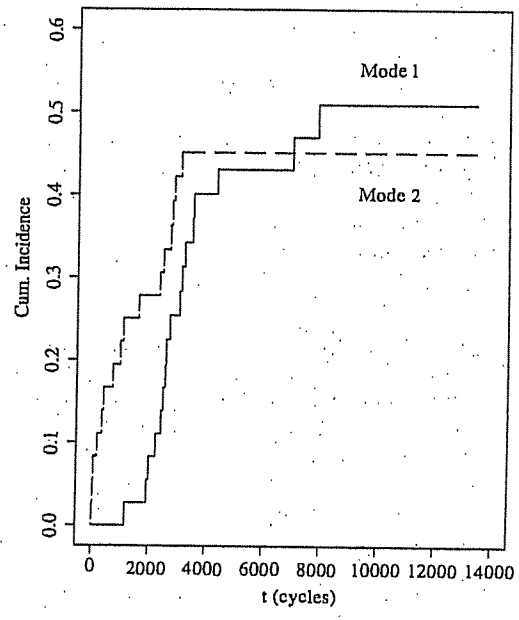

Figure 9.2. Estimates of mode-specific cumulative incidence functions, appliance test data.

other failure mechanisms within an individual. For the present situation a check on this is to some extent possible, since Nelson (1970b) discussed additional data from tests that were run after certain improvements in the appliance had been made. These data show that the two mode-specific hazard functions here were greatly affected by these changes.

Parametric modeling of the $\Lambda_j(t)$'s and other quantities is also possible. No particular model is suggested by the background to this problem, though flexible forms such as (9.1.7) could be considered. There is little point in doing this here, since the situation regarding failure mechanisms and models changes significantly as design modifications to the appliance are made.

Methods similar to those above can also be given for grouped data. This is done in the following section on parametric methods, since grouping the lifetimes into a finite number of intervals makes models finite-dimensional.

9.3 PARAMETRIC METHODS

As discussed in Section 9.1, the most convenient parametric models for continuous time data are ones for which the mode-specific hazard functions (9.1.1) are specified as $\lambda_j(t; \boldsymbol{\theta}_j)$, with $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k$ functionally independent parameters. By (9.1.11) and (9.1.12), the likelihood function then factors as a product,

$$L(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k) = \prod_{j=1}^{k} L_j(\boldsymbol{\theta}_j), \tag{9.3.1}$$

with $L_j(\boldsymbol{\theta}_j)$ given by (9.1.13). Models for which $\lambda_j(t; \boldsymbol{\theta}_j)$ is of Weibull, log-logistic, log-normal, and a few other common forms can be fitted using standard survival analysis software, as discussed in Section 9.1.2. Weibull forms for which

$$\lambda_j(t; \alpha_j, \gamma_j) = \frac{\gamma_j}{\alpha_j}\left(\frac{t}{\alpha_j}\right)^{\gamma_j - 1} \tag{9.3.2}$$

are often useful.

Regression models for the $\lambda_j(t)$'s are also easily handled. Consider a fixed vector of covariates $\mathbf{x}$; then it is only necessary to replace $\lambda_j(t)$, $g_j(t)$, and $G_j(t)$ in (9.1.11) with $\lambda_j(t|\mathbf{x})$, $g_j(t|\mathbf{x})$, and $G_j(t|\mathbf{x})$. Parametric accelerated failure time (AFT) specifications are convenient, since parametric survival analysis software handles many models of this type. The Weibull (AFT) model corresponding to (9.3.2) is often useful; it has $\boldsymbol{\theta}_j = (\boldsymbol{\beta}_j, \gamma_j)$ where

$$\lambda_j(t|\mathbf{x}; \boldsymbol{\theta}_j) = \frac{\gamma_j}{\alpha_j(\mathbf{x})}\left(\frac{t}{\alpha_j(\mathbf{x})}\right)^{\gamma_j - 1}, \tag{9.3.3}$$

and $\alpha_j(\mathbf{x}) = \exp(\boldsymbol{\beta}_j'\mathbf{x})$.

Estimation of quantities such as the distribution functions $F_j(t)$ or $F_j^*(t)$ is straightforward in principle, though they are in general complex functions of all of $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k$. Variance estimates can be obtained by a tedious application of Theorem B2; bootstrap methodology is an alternative approach for obtaining variance estimates or confidence intervals.

Standard diagnostic techniques from Chapters 3–6 can be used to assess parametric model assumptions, because of the formal connection between the individual failure modes and univariate lifetime models. If there are several failure modes or if censoring is heavy, then there may be relatively few observed failures for a specific mode j. As discussed in earlier chapters, this limits one's ability to detect model inadequacy.

Example 9.3.1. The data in Table 9.2 give the survival times for two groups of laboratory mice, all of which were exposed to a fixed dose of radiation at an age of 5 to 6 weeks (Hoel 1972). The first group of mice lived in a conventional lab environment and the second group was kept in a germ-free environment. The cause of death for each mouse was assigned after autopsy to be one of three things: thymic lymphoma (C_1), reticulum cell sarcoma (C_2), or other causes (C_3). The mice all died by the end of the experiment, so there is no censoring.

It was of particular interest to compare the mortality from the different failure modes in the conventional and germ-free environments. Plots of Nelson–Aalen estimates (9.2.1) for the three causes show that the hazard functions for C1 (thymic lymphoma) are similar in the two environments, but that those for modes C2 and C3 differ substantially. Figure 9.3 shows the $\hat{\Lambda}_1(t)$'s for the two environments, and Figure 9.4 shows the $\hat{\Lambda}_2(t)$'s.

Table 9.2. Survival Times and Causes of Death for Laboratory Mice

Control Group	Germ-Free Group
C_1 Deaths	
159,189,191,198,200,207,220,235,245,250, 256,261,265,266,280,343,350,383,403,414, 428,432	158,192,193,194,195,202,212,215,229,230, 237,240,244,247,259,300,301,321,337,415, 434,444,485,496,529,537,624,707,800
C_2 Deaths	
317,318,399,495,525,536,549,552,554,557, 558,571,586,594,596,605,612,621,628,631, 636,643,647,648,649,661,663,666,670,695, 697,700,705,712,713,738,748,753	430,590,606,638,655,679,691,693,696,747, 752,760,778,821,986
C_3 Deaths	
40,42,51,62,163,179,206,222,228,252,259, 282,324,333,341,366,385,407,420,431,441, 461,462,482,517,517,524,564,567,586,619, 620,621,622,647,651,686,761,763	136,246,255,376,421,565,616,617,652,655, 658,660,662,675,681,734,736,737,757,769, 777,800,806,825,855,857,864,868,870,870, 873,882,895,910,934,942,1015,1019

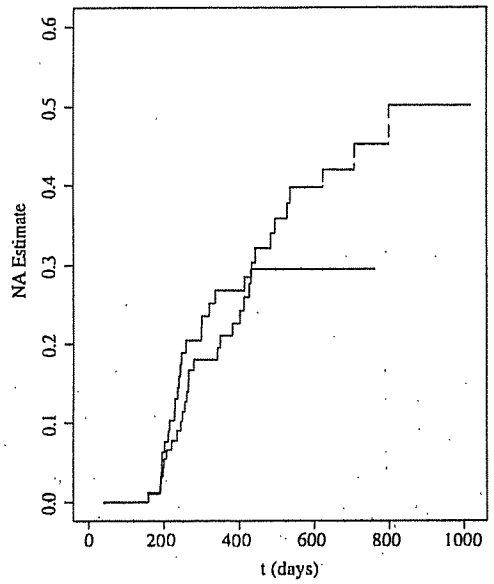

Figure 9.3. NA estimates of cumulative hazards for mode C1 (thymic lymphoma): ——— conventional environment; – – – germ-free environment.

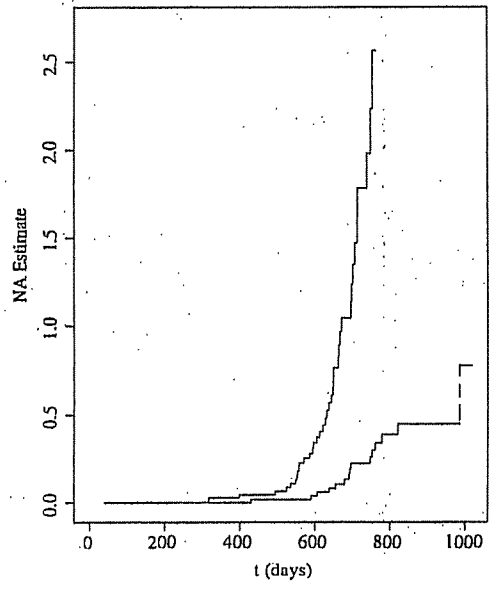

Figure 9.4. NA estimates of cumulative hazards for mode C2 (reticulum cell sarcoma): ——— conventional environment; – – – germ-free environment.

Formal tests for modes C2 and C3 provide strong evidence of a difference. Distribution-free tests are considered in Section 9.4. Plots of $\log \hat{\Lambda}_j(t)$ versus $\log t$ for $j = 2, 3$ are roughly linear and suggest that Weibull forms (9.3.2) might be adopted, so an alternative procedure is to use a parametric comparison. For example, for mode C2 the values of $\hat{u}_j = \log \hat{\alpha}_j$ and $\hat{b}_j = \hat{\gamma}_j^{-1}$ in (9.3.2) for the two environments are (with standard errors in brackets) $\hat{u}_2 = 6.52(.02)$, $\hat{b}_2 = .124(.015)$ for the conventional and $\hat{u}_2 = 6.92(.06)$, $\hat{b}_2 = .193(.039)$ for the germ-free environment. A likelihood ratio test of the hypothesis that (u_2, b_2) are the same in the two environments gives an observed value of 60.9 on two degrees of freedom, so there is very strong evidence against the hypothesis.

Example 9.3.2. PH Model with Piecewise-Constant Hazards. The proportional hazards regression model with piecewise-constant hazard function, discussed in Example 6.5.1 and in Section 7.4, is easily extended to the multiple failure modes setting. For a specified set of cut points $0 = a_0 < a_1 < \cdots < a_m$ as in Section 9.2, suppose that

$$\lambda_j(t|\mathbf{x}) = \lambda_{j0}(t)e^{\beta_j'\mathbf{x}}, \tag{9.3.4}$$

with

$$\lambda_{j0}(t) = \alpha_{j\ell}, \qquad a_{\ell-1} \le t < a_\ell, \tag{9.3.5}$$

for $j = 1, \ldots, k$ and $\ell = 1, \ldots, m$. Because of the factorization of the likelihood (9.1.11) into k pieces, the m.l.e.'s $\hat{\boldsymbol{\alpha}}_j$ and $\hat{\boldsymbol{\beta}}_j$ $(j = 1, \ldots, k)$ and inferences concerning $(\boldsymbol{\alpha}_j, \boldsymbol{\beta}_j)$'s are obtained by k separate applications of the procedures in Example 6.5.1. It is, in particular, straightforward to assess the effects of covariates on hazard functions.

Cumulative incidence or subdistribution functions are more complicated to deal with. The extension of (9.1.3) to include covariates gives

$$F_j(t|\mathbf{x}) = \int_0^t S(u|\mathbf{x})\lambda_j(u|\mathbf{x})\,du, \tag{9.3.6}$$

Since

$$S(u|\mathbf{x}) = \exp\left\{-\sum_{r=1}^{k} \Lambda_{r0}(u)e^{\beta_r'\mathbf{x}}\right\},$$

$F_j(t|\mathbf{x})$ depends on all of the parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k$ and the $\alpha_{j\ell}$ for $j = 1, \ldots, k$. The development of variance estimates for $\hat{F}_j(t|\mathbf{x})$ via Theorem B2 is therefore tedious, and for interval estimation it may be simpler to use bootstrap methodology. Further discussion of this model is given in Section 9.4.

9.3.1 Grouped or Discrete Data

Data with multiple failure modes often come in grouped, or life table, form in areas such as demography and actuarial science. In the absence of covariates, and under the assumption that there is no censoring except in the last interval, the likelihood function is given by (9.1.15),

$$L = \left\{ \prod_{\ell=1}^{m} \prod_{j=1}^{k} \pi_{j\ell}^{d_{j\ell}} \right\} S(a_m)^{d_{m+1}}, \tag{9.3.7}$$

where lifetimes are grouped into intervals $I_\ell = [a_{\ell-1}, a_\ell)$, $\ell = 1, \ldots, m+1$ with $a_0 = 0$ and $a_{m+1} = \infty$. The parameter $\pi_{j\ell}$ is the probability an individual fails in $[a_{\ell-1}, a_\ell)$ by mode j, and $d_{j\ell}$ is the observed number of individuals failing in $[a_{\ell-1}, a_\ell)$ by mode j. The failure modes for the d_{m+1} individuals still alive at time a_m are unknown. Note that $S(a_m) = 1 - \sum_{j,\ell} \pi_{j\ell}$.

An alternative way to write (9.3.7) is in terms of conditional probabilities

$$q_{j\ell} = Pr(\text{individual fails of mode } j \text{ in } I_\ell | \text{ alive at } a_{\ell-1}) = \pi_{j\ell}/S(a_{\ell-1}). \tag{9.3.8}$$

Then (9.3.7) can be expressed as

$$L = \prod_{\ell=1}^{m} \prod_{j=1}^{k} q_{j\ell}^{d_{j\ell}} (1 - q_{\cdot\ell})^{n_\ell - d_{\cdot\ell}}, \tag{9.3.9}$$

where $q_{\cdot\ell} = \sum_j q_{j\ell}$, $d_{\cdot\ell} = \sum_j d_{j\ell}$, and $n_\ell = n - (d_1 + \cdots + d_{\ell-1})$ is the number of individuals alive at $a_{\ell-1}$. This representation also holds when censoring of individuals is allowed at the ends of the intervals: in that case, n_ℓ is the number of individuals alive and uncensored at $a_{\ell-1}$, and therefore at risk of failure in $[a_{\ell-1}, a_\ell)$. The likelihood (9.3.9) is proportional to (3.6.8), used previously for life table data, when there is only one mode of failure. As discussed in Sections 3.6 and 4.3.1, the occurrence of censoring within intervals is more difficult to handle, and requires further assumptions; see Problem 9.12.

If an underlying parametric continuous-time model is assumed, then the $\pi_{j\ell}$ or $q_{j\ell}$ in (9.3.7) or (9.3.8) are as given by (9.1.14). The likelihood (9.3.7) is of multinomial form, and (9.3.9) is of product-multinomial form; either can be maximized with standard optimization software (see Appendix D). In the case where there is no underlying parametric model, there are km parameters $q_{j\ell}$ that satisfy $0 \le q_{j\ell} \le 1$ and $0 \le q_{1\ell} + \cdots + q_{k\ell} \le 1$ for each ℓ. It is easily seen that (9.3.8) is maximized over this parameter space by

$$\hat{q}_{j\ell} = d_{j\ell}/n_\ell \qquad j = 1, \ldots, k; \qquad \ell = 1, \ldots, m. \tag{9.3.10}$$

The m.l.e. of $S(a_\ell)$ is

$$\hat{S}(a_\ell) = \prod_{u=1}^{\ell-1} (1 - \hat{q}_{\cdot u}),$$

and estimates of other quantities are also easily obtained. Confidence intervals and tests can be handled using likelihood ratio statistics or Wald statistics. For the latter, it is seen directly from inversion of the observed information matrix based on (9.3.9) that asymptotic variance estimates of the $\hat{q}_{j\ell}$ are

$$\begin{aligned} &\widehat{\text{Var}}(\hat{q}_{j\ell}) = \hat{q}_{j\ell}(1 - \hat{q}_{j\ell})/n_\ell && j = 1, \ldots, k; && \ell = 1, \ldots, m \\ &\widehat{\text{Cov}}(\hat{q}_{j\ell}, \hat{q}_{r\ell}) = -\hat{q}_{j\ell}\hat{q}_{r\ell}/n_\ell && j \ne r; && \ell = 1, \ldots, m \\ &\widehat{\text{Cov}}(\hat{q}_{j\ell}, \hat{q}_{ru}) = 0 && \ell \ne u. \end{aligned} \tag{9.3.11}$$

An application of these procedures is considered in Problem 9.9.

9.4 SEMIPARAMETRIC METHODS FOR MULTIPLICATIVE HAZARDS MODELS

As noted in Section 9.3, regression modeling for competing failure modes is conveniently undertaken by considering specifications $\lambda_j(t|\mathbf{x})$ for the mode-specific hazard functions, given a vector of covariates $\mathbf{x}$. If multiplicative hazards models of the type considered in Chapter 7 are used, then semiparametric methods presented there can be applied. In particular, PH models of the form

$$\lambda_j(t|\mathbf{x}) = \lambda_{0j}(t) e^{\boldsymbol{\beta}_j' \mathbf{x}} \qquad j = 1, \ldots, k \tag{9.4.1}$$

can be treated exactly as in Chapter 7 by considering each mode of failure separately.

The partial likelihood methods of Section 7.1.1 clearly apply. This can be seen by developing the partial likelihood for $\boldsymbol{\beta}_j$ through consideration of the probability that a particular individual fails by mode j at time t, given that one of the individuals at risk at time t fails by mode j. This gives a partial likelihood

$$L(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k) = \prod_{i=1}^{n} \prod_{j=1}^{k} \left(\frac{e^{\boldsymbol{\beta}_j' \mathbf{x}_i}}{\sum_{\ell=1}^{n} Y_\ell(t_i) e^{\boldsymbol{\beta}_j' \mathbf{x}_\ell}} \right)^{\delta_{ij}}, \tag{9.4.2}$$

where $Y_\ell(t) = I(t_\ell \ge t)$. This factors into a product of terms $L_j(\boldsymbol{\beta}_j)$ for $j = 1, \ldots, k$, with

$$L_j(\boldsymbol{\beta}_j) = \prod_{i=1}^{n} \left(\frac{e^{\boldsymbol{\beta}_j' \mathbf{x}_i}}{\sum_{\ell=1}^{n} Y_\ell(t_i) e^{\boldsymbol{\beta}_j' \mathbf{x}_\ell}} \right)^{\delta_{ij}}, \tag{9.4.3}$$

which is of exactly the form (7.1.5). Similarly, the generalized Nelson–Aalen estimate,

$$\hat{\Lambda}_{0j}(t) = \sum_{i: t_i \le t} \left(\frac{\delta_{ij}}{\sum_{\ell=1}^{n} Y_\ell(t_i) e^{\hat{\boldsymbol{\beta}}_j' \mathbf{x}_\ell}} \right), \tag{9.4.4}$$

can be motivated, just as (7.1.32) was in Section 7.1.5. Procedures for hypothesis tests or interval estimation for $\boldsymbol{\beta}_j$, $\Lambda_{0j}(t)$, or $\Lambda_j(t|\mathbf{x})$ also can be based on the methods of Section 7.1, and the methodology extends to deal with time-varying covariates $\mathbf{x}(t)$ in place of $\mathbf{x}$ in (9.4.1). Finally, note that it is not necessary to include the same covariates in the models for different failure modes. Certain components of $\boldsymbol{\beta}_j$ in (9.4.1) can be restricted to be equal to zero, so the corresponding terms drop out of the model.

Example 9.4.1. (Example 9.3.1 revisited). An examination of the mouse mortality data in Example 9.3.1 indicated that the mortality hazard functions for mode C1 (thymic lymphoma) were similar for the conventional and germ-free laboratory environments, but that the hazards for mode C2 (reticulum cell sarcoma) were quite different. The Nelson–Aalen estimates of the cumulative hazards $\Lambda_j(t)$, $j = 1, 2$, also suggest that proportional hazards models

$$\lambda_j^C(t) = e^{\beta_j}\lambda_j^{GF}(t) \qquad j = 1, 2$$

would be reasonable. In this case, it is sensible to test the equality of the conventional and germ-free hazards using the log rank methodology of Sections 7.1.2 and 8.1.2. An application of the log rank test based on (8.1.14) and (8.1.15) gives X^2 values of 1.16 and 45.0, respectively, for modes C1 and C2. The $\chi^2_{(1)}$ p-values indicate that there is no evidence of a difference in the thymic lymphoma mortality hazard functions, but very strong evidence of a difference for the reticulum cell sarcoma hazards. The latter result agrees closely with the parametric test considered in Example 9.3.1.

Example 9.4.2. Tuli et al. (2000) discussed an observational study on children with hydrocephalus for whom internal shunts were inserted surgically. These shunts drain excess cerebrospinal fluid away from the head, typically to the abdominal area, and have led to a major decline in neurological deficit and death. Shunts are designed to stay in patients over their lifetimes, but "failures" occur, in which a blockage, infection or other condition requires that an existing shunt be partially replaced. In the study reported by Tuli et al. (2000), data were available on the initial shunt, and on any subsequent failures and replacement shunts, for 839 children who had initial shunts inserted during the years 1987–1996. Lawless et al. (2001) describe the data and give additional analyses. We will focus here on the analysis of the time to first shunt failure, from insertion.

Information on failures up to the end of 1997 was available, and of the 839 patients, 453 experienced a failure. Censoring times for children not experiencing a failure range from about 1 year to about 11 years. Three primary modes (causes) of failure were defined: Obstruction, Infection, and Other (other causes). About 70% of the observed failures were due to obstruction, and 15% were due to each of the other two causes. Mortality is also a competing risk or mode of failure; 121 deaths occurred among the patients during the study period, though not necessarily before a first shunt failure.

The primary objectives of analysis are to identify risk factors associated with each cause of failure. For illustration, we consider the first failures of the Obstruction and Infection modes. Preliminary analysis (see Lawless et al. 2001) identified three primary risk factors:

1. The age of the child at the time of shunt insertion. In the analysis here, age was categorized as

   $$\text{age} < 0, \qquad 0 < \text{age} < 1 \text{ year}, \qquad \text{age} > 1 \text{ year}.$$

   The age < 0 category is due to the fact that some children were born prematurely and had shunts inserted before their full-term birth date.
2. The etiology, or cause of the hydrocephalus. This is represented by eight categories: Ivhemm (intraventricular hemorrhage), Men (meningitis), Adsten (acqueductal stenosis), Tumor, Trauma, Mmc (myelominingocele), Other (other causes), and Con (congenital).
3. Shunt type, classified as vp (ventriculoperitoneal) or other.

Table 9.3 shows estimated regression coefficients $\hat{\beta}_j$ and standard errors for proportional hazards models (9.4.1), with binary covariates used to represent age and etiology categories. The baseline category for age was age > 1 and for etiology was

Table 9.3. PH Model Fits for Shunt Failures Due to Infection and Obstruction

Covariate	Infection $\hat{\beta}$	Infection *se*	Obstruction $\hat{\beta}$	Obstruction *se*
Age				
age < 0	1.33	.47	1.14	.26
(age < 0)$*I(1 \le t < 2)$	—	—	.32	.43
(age < 0)$*I(t \ge 2)$	—	—	−1.23	.55
$0 \le$ age < 1	.77	.42	.76	.20
Etiology				
Ivhemm	1.04	.57	.67	.25
Men	.93	.71	.40	.34
Adsten	1.30	.61	.55	.30
Tumor	.93	.66	.69	.28
Trauma	2.14	.81	.95	.44
Mmc	1.21	.54	.63	.22
Other	.93	.58	.42	.25
Shunt type				
Shunt type	—	—	−.87	.21
Shunt type $*I(1 \le t < 2)$	—	—	.79	.76
Shunt type $*I(t \ge 2)$	—	—	1.26	.63

Con. Thus, for example, the value $\exp(\hat{\beta}) = \exp(1.33) = 3.78$ for the covariate "age < 0" for the Infection mode estimates the hazard function for this age category to be 3.78 times that for the baseline "age > 1" category. Shunt type was not significant for Infection failures, so has been omitted from that model. In addition, diagnostic checks described in Sections 7.1 and 7.2 indicated that the effects of the age < 0 and shunt-type covariates on Obstruction failures were time-varying. As a result, three covariates were defined for shunt type: $I(\text{shunt type} = vp)$, $I(\text{shunt type} = vp)I(1 \le t < 2)$, and $I(\text{shunt type} = vp)I(t \ge 2)$. Three similar covariates were defined for the age < 0 indicator covariate. Other diagnostic checks do not contradict the models represented in Table 9.3.

The estimated baseline cumulative hazard functions $\hat{\Lambda}_{01}(t)$ and $\hat{\Lambda}_{02}(t)$, as well as the raw data, show that Infection (and Other) failures virtually all occur within 300 days of shunt insertion, whereas many Obstruction failures occur long after insertion. This, along with the results shown in Table 9.3, provide a rather clear picture. Younger age at insertion is highly associated with an increased risk of failure due to both infection and obstruction, but in the case of obstruction, the excess risk for the age < 0 group disappears by about two years after insertion. There is a suggestion that this may also be true for the $0 \le$ age < 1 group, but the evidence falls short of statistical significance. The vp shunt type (which is used in about 90% of the cases) is associated with a lower risk of obstruction failure over the first year after insertion, but a higher risk after about two years. Finally, the seven etiology categories in Table 9.3 all have estimated risks, relative to the baseline category Congenital, greater than one. The estimated failure hazards associated with trauma cases are especially high, but there are only 19 out of 839 patients whose hydrocephalus is due to trauma, and the standard errors are consequently large.

9.4.1 Estimation of Cumulative Incidence Functions

Subdistribution or cumulative incidence functions $F_j(t|\mathbf{x})$ can also be estimated under the PH model. Note first that the survivor function for T given $\mathbf{x}$ can be estimated as

$$\hat{S}(t|\mathbf{x}) = \exp\left\{-\sum_{j=1}^{k} \hat{\Lambda}_{0j}(t) e^{\hat{\beta}_j'\mathbf{x}}\right\}. \tag{9.4.5}$$

Variance estimates for the quantities

$$\hat{\Lambda}_j(t|\mathbf{x}) = \hat{\Lambda}_{0j}(t) e^{\hat{\beta}_j'\mathbf{x}} \tag{9.4.6}$$

were presented in Section 7.1.5; see (7.1.39) for the variance estimate for $\hat{\Lambda}_j(t|\mathbf{x})$. Because the overall partial likelihood (9.4.2) factors, the estimators $(\hat{\boldsymbol{\beta}}_j, \hat{\Lambda}_{0j}(t))$ are asymptotically independent for $j = 1, \ldots, k$ and a variance estimate for $\log \hat{S}(t|\mathbf{x})$, and thus $\hat{S}(t|\mathbf{x})$, is easily obtained.

The cumulative incidence function $F_j(t|\mathbf{x})$ can be estimated as

$$\begin{aligned}\hat{F}_j(t|\mathbf{x}) &= \int_0^t \hat{S}(u|\mathbf{x})\, d\hat{\Lambda}_j(u|\mathbf{x}) \\ &= \int_0^t \exp\left\{-\sum_{\ell=1}^{k} \hat{\Lambda}_{0\ell}(u) e^{\hat{\beta}_\ell'\mathbf{x}}\right\} e^{\hat{\beta}_j'\mathbf{x}}\, d\hat{\Lambda}_{0j}(u) \\ &= \sum_{i:t_i \le t} \delta_{ij} \exp\left\{-\sum_{\ell=1}^{k} \hat{\Lambda}_{0\ell}(t_i) e^{\hat{\beta}_\ell'\mathbf{x}}\right\} \frac{e^{\hat{\beta}_j'\mathbf{x}}}{\sum_{\ell=1}^{n} Y_\ell(t_i) e^{\hat{\beta}_j'\mathbf{x}_\ell}}.\end{aligned} \tag{9.4.7}$$

It will be noted that $\hat{S}(t|\mathbf{x})$ in (9.4.5) and $\hat{F}_j(t|\mathbf{x})$ in (9.4.7) do not have the same form when $\mathbf{x} = \mathbf{0}$ as the estimates in (9.2.3) and (9.2.5) for the no-covariate case. It is possible to use an alternative estimate for $S(t|\mathbf{x})$, based on the extension of the product representation (1.2.16) to the case of covariates, but (9.4.5) is simpler and does not differ much from other sensible estimates. Andersen et al. (1993, Sec. 7.2.3) consider the alternative approach in a general setting that includes the multiple failure modes model here.

Variance estimation for $\hat{F}_j(t)$ in the no-covariate case was discussed in Section 9.2. The results for the present situation are much more complicated; see, for example, Benichou and Gail (1990) or Cheng et al. (1998). A reasonably straightforward approach that is satisfactory in practical situations is to use a model with piecewise-constant baseline hazard functions for variance estimation. This PH model was discussed in Example 9.3.2; the baseline mode-specific hazard functions $\lambda_{0j}(t)$ are of the form (9.3.5), which we write here in terms of parameters $\lambda_{j\ell}$,

$$\lambda_{0j}(t) = \lambda_{j\ell} \qquad t \in I_\ell, \tag{9.4.8}$$

where $I_\ell = [a_{\ell-1}, a_\ell)$, $\ell = 1, \ldots, m$, with cut points $0 = a_0 < a_\ell < \cdots < a_m$. As in (9.2.7), the corresponding c.h.f.'s are

$$\Lambda_{0j}(t) = \sum_{\ell=1}^{m} \lambda_{j\ell} \Delta_\ell(t), \tag{9.4.9}$$

where $\Delta_\ell(t)$ is the length of the intersection of I_ℓ and $[0, t)$.

Since the likelihood function (9.1.11) (extended to include covariates) factors into separate pieces for each failure mode, inference procedures for $\boldsymbol{\lambda}_j = (\lambda_{j1}, \ldots, \lambda_{jm})$ and $\boldsymbol{\beta}_j$ under the proportional hazards model (9.4.1) follow, exactly as given in Example 6.5.1. In particular, the m.l.e. for $\boldsymbol{\beta}_j$ can be found by maximizing the profile log-likelihood function from (6.5.11),

$$\ell_{pj}(\boldsymbol{\beta}_j) = \sum_{i=1}^{n} \delta_{ij} \left\{\boldsymbol{\beta}_j'\mathbf{x}_i - \log\left[\sum_{\ell=1}^{m} I(t_i \in I_\ell) \sum_{r=1}^{n} \Delta_\ell(t_r) e^{\boldsymbol{\beta}_j'\mathbf{x}_r}\right]\right\},$$

and by (6.5.10) the $\lambda_{j\ell}$ are estimated by

$$\hat{\lambda}_{j\ell} = \frac{d_{j\ell}}{\sum_{i=1}^{n} \Delta_\ell(t_i) e^{\hat{\beta}_j' \mathbf{x}_i}}, \qquad \ell = 1, \ldots, m, \tag{9.4.10}$$

where $d_{j\ell}$ is the number of individuals observed to fail in interval I_ℓ of mode j. As shown in Section 7.4, the m.l.e.'s for $\boldsymbol{\beta}_j$ and for $\Lambda_{0j}(t)$ under this model approach the estimates for $\boldsymbol{\beta}_j$ obtained from the Cox partial likelihood (9.4.3) and the generalized Nelson–Aalen estimates (9.4.4), when m increases and the interval lengths $|a_\ell - a_{\ell-1}|$ approach zero.

To obtain a variance estimate for $\hat{F}_j(t|\mathbf{x})$ in the piecewise-constant hazards model is straightforward, but somewhat tedious. From Example 6.5.1, it follows that the information matrix for $\boldsymbol{\lambda}_j$ and $\boldsymbol{\beta}_j$ is, in partitioned form,

$$I_j(\boldsymbol{\lambda}_j, \boldsymbol{\beta}_j) = \begin{pmatrix} D_j & C_j \\ C_j' & B_j \end{pmatrix}, \tag{9.4.11}$$

where

$$D_j = \text{Diag}(d_{j\ell}/\lambda_{j\ell}^2)$$

is a diagonal $m \times m$ matrix, C_j is an $m \times p$ matrix with (ℓ, q) entry,

$$(C_j)_{\ell q} = \sum_{i=1}^{n} \Delta_\ell(t_i) x_{iq} e^{\beta_j' \mathbf{x}_i},$$

and B_j is a $p \times p$ matrix,

$$B_j = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \Lambda_{0j}(t_i) e^{\beta_j' \mathbf{x}_i}.$$

The m.l.e. $\hat{F}_j(t|\mathbf{x})$ involves $(\hat{\boldsymbol{\lambda}}_r, \hat{\boldsymbol{\beta}}_r)$ for all $r = 1, \ldots, k$. By the asymptotic independence of the $(\hat{\boldsymbol{\lambda}}_r, \hat{\boldsymbol{\beta}}_r)$'s and an application of Theorem B2 (Appendix B), it follows that

$$\widehat{\text{Var}}[\hat{F}_j(t|\mathbf{x})] = \sum_{r=1}^{k} \hat{\mathbf{w}}_r^{(j)\prime} I_r(\hat{\boldsymbol{\lambda}}_r, \hat{\boldsymbol{\beta}}_r)^{-1} \hat{\mathbf{w}}_r^{(j)}, \tag{9.4.12}$$

where $\mathbf{w}_r^{(j)}$ is an $(m + p) \times 1$ vector,

$$\mathbf{w}_r^{(j)} = \left(\frac{\partial F_j(t|\mathbf{x})}{\partial \lambda_{r1}}, \ldots, \frac{\partial F_j(t|\mathbf{x})}{\partial \lambda_{rm}}, \frac{\partial F_j(t|\mathbf{x})}{\partial \beta_{r1}}, \ldots, \frac{\partial F_j(t|\mathbf{x})}{\partial \beta_{rp}} \right)'.$$

The entries of this vector are found from

$$F_j(t|\mathbf{x}) = \int_0^t S(u|\mathbf{x}) \lambda_{0j}(u) e^{\beta_j' \mathbf{x}} \, du,$$

which give

$$\frac{\partial F_j(t|\mathbf{x})}{\partial \lambda_{r\ell}} = e^{\beta_j' \mathbf{x}} \int_0^t S(u|\mathbf{x}) [I(r = j) I(u \in I_\ell) - e^{\beta_r' \mathbf{x}} \Delta_\ell(u) \lambda_{0j}(u)] \, du \tag{9.4.13}$$

$$\frac{\partial F_j(t|\mathbf{x})}{\partial \boldsymbol{\beta}_r} = \mathbf{x} e^{\beta_j' \mathbf{x}} \int_0^t S(u|\mathbf{x}) [I(r = j) - \Lambda_{0r}(u) e^{\beta_r' \mathbf{x}}] \lambda_{0j}(u) \, du. \tag{9.4.14}$$

A suitable practical procedure is to approximate the variance of $\hat{F}_j(t|\mathbf{x})$ based on the semiparametric estimates using (9.4.12) with a moderately large value of m, and the a_ℓ chosen so that each interval $[a_{\ell-1}, a_\ell)$ has at least a few observed failures. Rather than compute the estimates of $\boldsymbol{\beta}_j$ and $\Lambda_{0j}(t)$ under the piecewise constant model, we can use the semiparametric estimates $\hat{\boldsymbol{\beta}}_j$ and $\hat{\Lambda}_{0j}(t)$ and approximate the entries in (9.4.11), (9.4.13), and (9.4.14) by replacing $\lambda_{j\ell}$ with

$$\hat{\lambda}_{j\ell} = \frac{\hat{\Lambda}_{0j}(a_\ell) - \hat{\Lambda}_{0j}(a_{\ell-1})}{a_\ell - a_{\ell-1}},$$

$\lambda_{0j}(u) \, du$ by $d\hat{\Lambda}_{0j}(u)$, and other quantities by their semiparametric estimates.

BIBLIOGRAPHIC NOTES

Multiple failure modes and competing risks problems have a long history in actuarial science and demography; Gail (1975) and Seal (1977) provide historical reviews. Early applications of actuarial techniques in medical contexts were considered by Cornfield (1957), Berkson and Elveback (1960), Kimball (1969), Pike (1970), Hoel (1972), and many others. David and Moeschberger (1978), Elandt-Johnson and Johnson (1980), and Crowder (2001) contain numerous additional references.

The emphasis on mode-specific hazard functions is implicit in the work of Altschuler (1970), Nelson (1969), and Aalen (1976), where the Nelson–Aalen nonparametric estimation techniques of Section 9.2 were introduced. Prentice et al. (1978) discuss mode-specific hazards and the estimation of competing risks characteristics in some detail. The fact that data on failure time, T, and mode of failure, C, do not allow discrimination between independent and nonindependent risks was noted by Cox (1959, 1962) and studied in more detail by Tsiatis (1975) and Peterson (1976). Connections with censoring were considered by Williams and Lagakos (1977), Lagakos and Williams (1978), and Kalbfleisch and MacKay (1979). Crowder (2001) provides a discussion of models and identifiability issues for competing risks.

Some aspects of nonparametric estimation have been developed within the more general context of continuous time Markov processes. For example, Aalen (1978a), Aalen and Johansen (1978) and Fleming (1978a,b) developed estimation of subdistribution probabilities $F_j(t)$, as discussed in Sec. 9.2. Andersen et al. (1993, Sec. 4.4) discuss these ideas in some detail. Matthews (1988) obtains empirical likelihood-based confidence intervals for subdistribution functions.

Parametric models of several types have been discussed in the literature. Specifications based directly on the mode-specific hazard functions, as in (9.3.2) and (9.3.3) are very common. Nelson (1982) and Crowder (2001) provide additional illustrations and references. Models in which the conditional subdistributions $F_j^*(t)$ in (9.1.6) and the marginal probabilities $\pi_j = Pr(C = j)$ are parameterized directly have also been considered (e.g., Mendenhall and Hader 1958; Larson and Dinse 1985). Another line of approach has been the use of parametric multivariate failure time distributions for the series system or latent failure time model described in Section 9.1.1 (e.g., Hoel 1972; Moeschberger 1974; David and Moeschberger 1978; Crowder et al. 1991, Sec. 7.4). As has been noted, such models are physically relevant only occasionally, and in any case, crucial aspects of such models are uncheckable using only data on (T, C).

Actuarial and demographic life table methodology for grouped multiple failure modes data as in Section 9.3 has been extensively developed. In these areas the term multiple decrement life table is often used to refer to the multiple modes of death or failure. Seal (1977) provides a historical overview and many details can be found in books such as Elandt-Johnson and Johnson (1980), Manton and Stallard (1988), and Namboodiri and Suchindran (1987).

Semiparametric methods based on proportional hazards models were considered by Holt (1978) in matched pairs settings and are described in some detail by Kalbfleisch and Prentice (1980, Sec. 7.2). Benichou and Gail (1990) and Cheng et al. (1998) consider interval estimation of subdistribution functions $F_j(t|\mathbf{x})$. Many authors have provided illustrations of this methodology to medical data (e.g., Kay 1986; Gaynor et al. 1993; Lunn and McNeil 1995). Other approaches, mainly based on conditional subdistributional functions, have also been examined (e.g., Fine 1999), but are considerably more awkward to implement.

Problems in which the mode of failure can be identified only up to a set of modes have received a good deal of study; this phenomenon is referred to as failure mode "masking." Flehinger et al. (1998) provide numerous references to this area.

PROBLEMS AND SUPPLEMENTS

9.1 Consider the series system model introduced following Example 9.1.1, and let $f(t_1, \ldots, t_k)$ be the joint p.d.f. for $T_1, \ldots, T_k$.

(a) Show that if $T_1, \ldots, T_k$ are mutually independent with s.f.'s $S_j(t)$, then the mode-specific hazard functions (9.1.1) are the hazard functions $h_j(t) = -d \log S_j(t)/dt$ for $T_1, \ldots, T_k$.

(b) Show also that in the likelihood function (9.1.13) based on data (T, C), $G_j(t) = S_j(t)$.

(c) If $T_1, \ldots, T_k$ are not independent, write down the joint probability distribution of (T, C), based on $f(t_1, \ldots, t_k)$. Show that there exists a joint model with independent $T_1, \ldots, T_k$ that gives exactly the same distribution for (T, C).

(Section 9.1.1)

9.2 *Independent censoring.* A lifetime distribution subject to right censoring can be considered as a model with two failure modes, $C = 1$ (lifetime observed at t) and $C = 2$ (censoring at t before lifetime observed). In this case, it is instructive to consider the series system model for (T_1, T_2), where $T_1 = T$ is the lifetime and T_2 is the censoring time for an individual. Let $\lambda_j(t)$, $j = 1, 2$, be the mode-specific hazard functions (9.1.1) and let $h_j(t)$, $t = 1, 2$, be the hazard functions for the marginal distributions of T_1 and T_2. Assume that (T_{1i}, T_{2i}) are independent for $i = 1, \ldots, n$, with joint p.d.f. $f(t_1, t_2)$.

(a) Show that the independent censoring definition (2.2.11) is in this case equivalent to the condition

$$h_1(t) = \lambda_1(t). \tag{9.5.1}$$

(b) Show that if T_1 and T_2 are independent, then (9.5.1) holds. Show also that although (9.5.1) does not hold in general, it is slightly weaker than independence of T_1 and T_2, and can hold when T_1 and T_2 are not independent.

(Section 9.1.1; Williams and Lagakos 1977; Kalbfleisch and MacKay 1979)

9.3 Consider the bivariate lifetime distribution with survivor function

$$S(t_1, t_2) = \exp[-(\lambda_1 t_1 + \lambda_2 t_2 + \lambda_1 \lambda_2 \theta t_1 t_2)],$$

where $\lambda_1 > 0$, $\lambda_2 > 0$, and $0 \le \theta \le 1$. Suppose that only series system data $T = \min(T_1, T_2)$ and $C = j : T = T_j$ is available.

(a) Obtain the mode-specific hazard functions (9.1.1) and also the hazard functions for the marginal distributions of T_1 and T_2.

(b) Discuss by direct illustration why it would not be possible to assess the adequacy of the joint model for (T_1, T_2) on the basis of data on (T, C).

(Section 9.1.1)

9.4 If the mode-specific hazard functions $\lambda_j(t)$ are proportional, significant simplifications occur. Suppose that $\lambda_j(t) = w_j \lambda(t)$ for $j = 1, \ldots, k$, where $0 < w_j < 1$ and $\sum w_j = 1$, so that $\lambda(t)$ is the marginal hazard function for T.

(a) Show that T and C are statistically independent, and that $\pi_j = w_j$. Determine the subdistribution functions $F_j(t)$.

(b) Consider nonparametric estimation of the distribution of (T, C) in this case.

(Sections 9.1, 9.2)

9.5 *Cause removal.* As discussed in Example 9.2.1, the question of how failures or mortality would change if some specific cause could be removed is often of interest, but difficult to study. In some situations, however, it might be reasonable to assume that if failures of mode j could be prevented, then the marginal hazard function for lifetime T would decrease from $\lambda(t)$ to

$$\lambda_{(-j)}(t) = \lambda(t) - \lambda_j(t). \tag{9.5.2}$$

(a) Suppose a system's main cause of failure is deterioration over time, but that failures are also occasionally caused by accidental external factors. Suppose that these two modes of failure have respective hazard functions $\lambda_1(t) = .3t^2$ and $\lambda_2(t) = .1$, with time measured in years. Determine the survival function $S(t)$ for T both with and without the second failure mode, assuming that (9.5.2) with $j = 2$ applies.

(b) Compute the change in expected lifetime that would result from removal of mode 2.

(Section 9.1)

9.6 For the mouse data of Example 9.3.1, fit Weibull forms (9.3.2) for each of the three mode-specific hazard functions, $\lambda_j(t)$. Assess the fit of the parametric models by plotting the corresponding estimates of $\Lambda_j(t)$ and the Nelson–Aalen estimates (9.2.1) on the same graph. Comment on the fit in each case. In the same way, compare the parametric and nonparametric estimates of the subdistribution functions $F_j(t)$.

(Sections 9.2, 9.3)

9.7 (*Continuation of Problem 9.6*). Fit an eight-parameter model to the mouse data, in which the conditional subdistribution functions $F_j^*(t)$ of (9.1.6) are of Weibull form,

$$F_j^*(t) = 1 - \exp[-(t/\alpha_j)^{\gamma_j}] \qquad j = 1, 2, 3$$

and with π_1, π_2, π_3 satisfying $0 < \pi_j < 1$, $\pi_1 + \pi_2 + \pi_3 = 1$. Assess the fit of the model by comparing the estimates of $F_j(t) = \pi_j F_j^*(t)$ with the nonparametric estimates from Problem 9.6.

(Section 9.3)

9.8 Davis and Lawrance (1989) considered data from a laboratory test on pneumatic tires. The test involved rotating the tires against a steel drum until some type of failure occurred. Failures were classified into six modes or categories: 1—open joint on the inner liner; 2—rubber chunking on the shoulder; 3—loose casing low on the sidewall; 4—cracking of the tread rubber; 5—cracking on the sidewall; 6—all other causes. The data are shown in Table 9.4, with $C = 0$ denoting that a tire did not fail under test, so that its failure time is censored. Times are in hours.

Table 9.4. Times and Modes of Failure for Pneumatic Tires

T	C	T	C	T	C	T	C	T	C	T	C
6	0	135	3	204	6	222	4	244	1	300	0
30	0	136	6	205	4	222	4	244	3	300	0
47	5	137	3	205	6	224	4	246	6	300	4
72	1	142	5	206	4	224	6	246	4	300	4
74	3	144	3	207	4	225	4	249	4	300	0
81	1	148	3	207	4	225	5	250	4	300	0
84	3	153	3	207	1	226	4	252	4	300	0
84	3	155	1	207	4	227	4	253	4	300	0
84	3	157	6	208	1	227	2	255	4	300	0
90	3	158	5	208	6	228	4	258	4	300	0
96	1	159	0	208	4	229	4	259	0	300	0
101	6	162	1	208	3	229	6	262	5	300	0
105	5	165	4	209	1	230	5	265	4	300	0
105	3	172	5	209	4	230	1	266	1	300	0
106	4	177	2	210	6	230	2	268	2	300	0
107	1	179	3	210	6	231	4	269	4	300	0
111	6	181	4	210	4	232	1	270	5	300	0
111	4	188	1	210	6	232	2	270	2	306	4
111	4	188	6	211	4	233	1	271	4	306	4
118	3	191	6	212	4	233	4	271	4	314	4
118	4	193	3	213	4	233	4	281	4	318	6
119	3	195	4	214	4	234	4	281	3	320	4
120	4	197	5	215	4	234	4	285	1	332	4
126	5	198	6	215	4	236	4	285	4	335	0
131	1	200	0	215	4	237	4	286	4	342	6
132	1	200	2	215	3	239	6	286	4	347	4
133	6	201	4	216	4	241	4	295	1		
133	3	203	4	217	4	241	4	297	2		
135	4	204	3	220	5	243	4	299	3		

(a) Obtain and plot nonparametric estimates of the mode-specific c.h.f.'s $\Lambda_j(t)$.

(b) At $t = 200$ hours the inflation pressure in unfailed tires was reduced; it was believed that this would accelerate failure. Discuss whether this appears to be the case, paying attention to the different failure modes. Discuss how parametric models could be used to carry out formal hypothesis tests.

(Sections 9.2, 9.3; Davis and Lawrance 1989)

9.9 Mendenhall and Hader (1958) presented data on the failure times of radio transmitter receivers. Failures were classified as one of two types: those confirmed on arrival at the maintenance center (Type 1) and those unconfirmed (Type 2). The data consist of a failure time and type for each receiver, except that when

Table 9.5. Frequency Distribution of Failure Time and Type for Radio Receivers

Time Interval (hours)	Type 1 Failures	Type 2 Failures	Total Failures
[0, 50)	26	15	41
[50, 100)	29	15	44
[100, 150)	28	22	50
[150, 200)	35	13	48
[200, 250)	17	11	28
[250, 300)	21	8	29
[300, 350)	11	7	18
[350, 400)	11	5	16
[400, 450)	12	3	15
[450, 500)	7	4	11
[500, 550)	6	1	7
[550, 600)	9	2	11
[600, 630)	6	1	7
[630, ∞)	—	—	44
Total	218	107	369

observation ceased after 630 hours, 44 of 369 receivers had still not failed and so have censored failure times. The data are shown in grouped form in Table 9.5, as given by Cox (1959).

(a) Estimate the conditional probabilities, $q_{j\ell}$, of (9.3.8).

(b) It is of interest whether the hazard functions for failure types 1 and 2 are proportional. Fit a model for which the $q_{j\ell}$ take the discrete proportional hazards form of Section 7.3.1,

$$\log[-\log(1-q_{2\ell})] = \log[-\log(1-q_{1\ell})]+\beta \qquad \ell = 1, \ldots, m, \quad (9.5.3)$$

Carry out a likelihood ratio test of H: $\beta = 0$.

(Section 9.3; Cox 1959)

9.10 (*Continuation of Problem 9.9*). For the radio transmitter receiver data in the preceding problem, consider an underlying continuous-time failure model with constant mode-specific hazard functions

$$\lambda_j(t) = \lambda_j, \qquad j = 1, 2. \quad (9.5.4)$$

(a) Fit this model to the data in Table 9.5. Carry out a likelihood ratio goodness-of-fit test of the model (9.5.4) against the unrestricted model based on the conditional failure probabilities $q_{j\ell}$, as in part (a) of Problem 9.9.

(b) Fit a model in which the underlying mode-specific hazard functions $\lambda_j(t)$ are of the Weibull form (9.3.2). Carry out a test of the model (9.5.4) through a likelihood ratio test of the hypothesis H: $\gamma_1 = 1$, $\gamma_2 = 1$ in (9.3.2). Comment on the agreement between this test of (9.5.4) and that in part (a).

(Section 9.3)

9.11 (*Continuation of Problem 9.10*). Mendenhall and Hader (1958) fitted a three-parameter model in which the conditional subdistribution functions, $F_j^*(t)$, of (9.1.6) are of exponential form:

$$F_j^*(t) = 1 - \exp(-\lambda_j t) \qquad j = 1, 2, \quad (9.5.5)$$

and $\pi_1 = p$, $\pi_2 = 1 - p$.

(a) Obtain the m.l.e.'s for λ_1, λ_2, and p. Carry out a likelihood ratio test of this model against the unrestricted model based on the $q_{j\ell}$ in Problem 9.9, part (a).

(b) Fit a five-parameter model in which the $F_j^*(t)$'s have the Weibull form of Problem 9.7. Test the model (9.5.5) through a likelihood ratio test of H: $\gamma_1 = 1$, $\gamma_2 = 1$. Comment on the results of this test and that in part (a).

(c) The raw (ungrouped) failure time data upon which Table 9.5 is based are given by Mendenhall and Hader (1958) and Crowder et al. (1991, p. 152). Discuss how tests of the models (9.5.4) or (9.5.5) can be based on the ungrouped data.

(Section 3.3; Mendenhall and Hader 1958)

9.12 *Censoring with grouped lifetime data.* Consider the grouped data, or life table, setting discussed in Sections 9.2, 9.3, and earlier in Sections 3.6 and 4.3. Suppose that censoring or withdrawal of individuals from risk of failure can occur within time intervals $I_\ell = [a_{\ell-1}, a_\ell)$, $\ell = 1, \ldots, m$. Censoring can be thought of as a competing failure mode, so define

$q_{1\ell} = Pr$ (an individual is observed to fail in I_ℓ| at risk at $a_{\ell-1}$)

$q_{2\ell} = Pr$ (an individual is observed to withdraw in I_ℓ| at risk at $a_{\ell-1}$).

(a) Let $S(t)$ be the survivor function for failure time t. In general $q_{1\ell}$ does not equal

$$q_\ell^{(S)} = 1 - \frac{S(a_\ell)}{S(a_{\ell-1})} \qquad \ell = 1, \ldots, m.$$

Give an expression for $q_{1\ell}$ and $q_{2\ell}$ in the case in which random withdrawal times C and failures T are independent, with C having continuous s.f., $G(t)$, and T continuous s.f., $S(t)$.

(b) Discuss why it is an uncheckable assumption whether censoring time is independent of failure time, using only information on the number of fail-

ures and withdrawals in each interval. Discuss the ramifications for testing the fit of a parametric model $S(t; \theta)$ for T.

(Sections 3.6, 4.3, 9.2, 9.3)

9.13 (*Continuation of Problem 9.12*). Consider the extension of the preceding problem to the case of k failure modes $1, 2, \ldots, k$, and let $q_{j\ell}^{(S)}$ be the probability an individual fails in I_ℓ of mode j, given that $T \geq a_{\ell-1}$. If w_ℓ is the number of withdrawals in I_ℓ, then an extension of the standard life table procedure of Section 3.6 is to estimate $q_{j\ell}^{(S)}$ by

$$\hat{q}_{j\ell}^{(S)} = \frac{d_{j\ell}}{n_\ell - .5w_\ell},$$

where $d_{j\ell}$ is the observed number of mode j failures in I_ℓ.

(a) Use this procedure to estimate the $q_{j\ell}^{(S)}$ and thus probabilities $F_j(a_\ell) = Pr(T \leq a_\ell, C = j)$. Use the theory of Section 3.6 to develop variance estimates for $\hat{F}_j(a_\ell)$.

(b) Apply these ideas to the data in Table 9.6 on survival times of 5982 women diagnosed with cervical cancer over the years 1942–1954 (Chiang 1961; Elandt-Johnson and Johnson 1980, Example 12.1). The survival times are grouped by years since diagnosis, and deaths are divided into those due to the cancer (C_1) and those due to other causes (C_2). Because of variable lengths of follow-up, many women have censored survival times, which are also grouped.

(Sections 9.2, 9.3)

9.14 Byar and Green (1980) discussed data from a randomized clinical trial on 483 patients with stage 3 and 4 prostate cancer. Patients were assigned to four treatment groups (Placebo, .2 mg of drug DES per day, 1.0 mg of DES per day, and 5 mg of DES per day). During the study, 125 patients died from prostate

Table 9.6. Survival Data for Women Diagnosed With Cervical Cancer

Interval (years)	At Risk n_l	Cervical Cancer d_{1l}	Other Causes d_{2l}	Censored w_l
0–1	5,982	1,175	201	576
1–2	4,030	588	96	501
2–3	2,845	221	48	459
3–4	2,117	121	44	379
4–5	1,573	63	28	306
5–6	1,176	27	34	254
6–7	861	18	16	167
7–8	660	13	12	161
8–9	474	6	8	116
9–10	344	9	5	85

cancer, 139 died from cardiovascular-related diseases, and 80 died from other causes. The data are reproduced in Andrews and Herzberg (1985), and include covariate information, such as age, history of cardiovascular disease, and disease stage.

The comparison of survival times for patients in the different treatment groups is of particular interest, but DES can have adverse cardiovascular effects as well as possibly decreasing the risk of mortality due to the cancer. Use the semiparametric proportional hazards methodology of Section 9.4 to carry out a thorough investigation of the multiple modes of death, with specific attention paid to treatment and covariate effects.

(Section 9.4; Cheng et al. 1998)

9.15 *Incomplete information on failure modes.* Sometimes the time of failure, T, is observed but the mode of failure is uncertain; in reliability this is often referred to as failure masking. Consider the case of three modes, and suppose that for some failures it is possible to say only that the mode was in the set $\{1, 2\}$. For the remainder, the failure mode is known, except for units whose lifetimes are censored. Assume that the probability a unit's failure mode is "masked" does not depend on the mode or time of failure.

(a) Examine maximum likelihood estimation for a model in which the mode-specific hazard functions are constant: $\lambda_j(t) = \lambda_j$ $(j = 1, 2, 3)$.

(b) Consider nonparametric estimation of the $\Lambda_j(t)$'s or $F_j(t)$'s of Section 9.1.1.

(Miyakawa 1984; Dinse 1986)

CHAPTER 10

Goodness-of-Fit Tests

10.1 INTRODUCTION

It is important to check the adequacy of models upon which inferences or actions are based. Models vary in the complexity and strength of their assumptions, and model-checking needs vary correspondingly. In the simplest case a model may involve a single random variable, Y, with distribution function, $F(y)$. The main problems are often to check whether some specific form, $F_0(y)$, is consistent with observed data, and whether assumptions about observations Y_i $(i = 1, \ldots, n)$ being independent and identically distributed (i.i.d.) are satisfactory. Problems involving covariates or dependence among responses use models for joint distributions of responses $Y_1, \ldots, Y_n$ given covariate vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Model checking is necessarily more complicated in these cases.

Informal methods of model checking emphasize graphical procedures such as probability and residual plots. These diagnostic tools have been discussed and used in preceding chapters. In most cases the variation inherent in graphical summaries is substantial, even when the data are generated by the assumed model, and the eye cannot always determine whether features in a plot are within the bounds of natural random variation. In addition, some types of assumptions are difficult to represent graphically. Consequently, formal hypothesis tests are an important part of model checking.

The term goodness-of-fit test is used to denote hypothesis tests concerning the distribution of some response variable, Y, in a population or process. The objective is to test a null hypothesis H_0 that specifies that the distribution of Y, or of Y given covariates $\mathbf{x}$, is of some specified form. For example, well-known statistics such as Pearson's chi-squared and Kolmogorov–Smirnov statistics can be used to test the hypothesis that Y has some specified distribution function,

$$H_0: F(y) = F_0(y). \tag{10.1.1}$$

For models involving covariates, tests directed at different specific assumptions may be used. For example, a location-scale regression model (6.1.3) may assume that (i) only the location of Y's distribution depends on $\mathbf{x}$, (ii) the location parameter $u(\mathbf{x})$

has a specific form, and (iii) the distribution of Y given $\mathbf{x}$ has a specific form. Each of these assumptions can be addressed by hypothesis tests.

Three main approaches to model testing can be identified:

(i) Tests based on embedding a proposed model M_0 in a larger or "expanded" family of models M_1 (often termed model expansion).

(ii) Tests based on a comparison of a proposed parametric model M_0 with a nonparametric counterpart.

(iii) Tests based on the properties of a proposed model M_0.

These categories are not distinct; for example, in some cases a test of type (ii) can also be formulated as type (i).

If the models M_0 and M_1 in approach (i) are both parametric, then standard hypothesis-testing methods apply, as illustrated throughout the book. For example, tests of Weibull and log-logistic lifetime distributions were obtained in Section 5.5.2 by embedding them in the log-Burr family of models. Tests of the proportional hazards (PH) assumption in the Cox model (7.1.1) were based on the expanded model (7.1.62) in Section 7.1.9.

Tests of type (ii) and (iii) are often related to graphical model checks. For example, probability plots and other graphical checks in Section 3.3 involve comparisons of parametric estimates $F(y; \hat{\boldsymbol{\theta}})$ and empirical or nonparametric estimates $\hat{F}(y)$. Regarding approach (iii), hypothesis tests for regression models are often based on functions of residuals $\hat{e}_i = g(y_i, \mathbf{x}_i; \hat{\boldsymbol{\theta}})$ as in (6.2.6). These test statistics are chosen to reflect the fact that the corresponding values $e_i = g(Y_i, \mathbf{x}_i; \boldsymbol{\theta})$ are i.i.d. and independent of the $\mathbf{x}_i$ if the null model is correct, just as plots of the $\hat{e}_i$ are suggested by these properties.

Tests of type (i) involve alternative models M_1 to the proposed model M_0, and will have good power for detecting departures from M_0 that are in, or close to, M_1. Certain procedures of types (ii) or (iii) are "omnibus" tests without specified alternative parametric models. They are designed to provide checks for the model against broad classes of alternatives and may not have very high power for certain departures from the model M_0.

Graphical methods and model expansion satisfy the majority of practical model-checking needs, but other procedures are sometimes useful. There is an extensive literature on goodness-of-fit tests, both generally and for lifetime models specifically. It is possible to provide only a short overview here, with references to the broader literature. Before discussing general methods of testing fit (Section 10.2), some tests for specific lifetime distributions (Section 10.3), and tests for regression models (Section 10.4), we note a few points.

1. Models only approximate reality, and with sufficient amounts of data formal tests will tend to indicate statistically significant lack of fit. The important question is usually whether a model approximates reality well enough for practical purposes.

2. Small amounts of data may not provide sufficient information to detect departures from a model. Likewise, censoring and other forms of incompleteness can make it impossible to detect certain types of departure. It is generally a good idea to assess the sensitivity of conclusions or decisions to model variation that is consistent with the observed data.
3. When covariates are present, checks on distributional shape for Y given $\mathbf{x}$ and on the form of dependence on $\mathbf{x}$ are often confounded to some extent.
4. Many goodness-of-fit tests are not good at detecting outliers or certain types of mixtures. Influence diagnostics and model expansion are important tools when outliers or mixtures are plausible.

10.2 SOME GENERAL METHODS OF TESTING FIT

This section describes some general methods of testing hypotheses H_0: $F(y) = F_0(y)$ as in (10.1.1). Tests based on comparing empirical distribution, survivor, or cumulative hazard functions with hypothesized models are discussed in Section 10.2.1. Tests based on expanded parametric models and on grouped lifetimes are considered in Sections 10.2.2 and 10.2.3, respectively.

10.2.1 Tests Based on Comparing Empirical and Hypothesized Distributions

The comparison of parametric estimates with nonparametric counterparts is a standard way of assessing parametric models. For example, the Kaplan–Meier estimate $\hat{F}(t)$ of a distribution function may be compared graphically with parametric estimates $F(t; \hat{\boldsymbol{\theta}})$; this has been used as an informal model checking tool in various parts of the book. For most practical purposes such a plot, perhaps enhanced with nonparametric confidence limits or bands around $\hat{F}(t)$, is sufficient. Nevertheless, formal goodness-of-fit tests based on measures of the distance between, say, $\hat{F}(t)$ and $F(t; \hat{\boldsymbol{\theta}})$ have been extensively studied. This area will be described briefly.

For simplicity we first consider the hypothesis H_0: $F(y) = F_0(y)$ in the case where $F_0(y)$ is a fully specified distribution and there is a complete random sample $y_1, \ldots, y_n$ from F. The empirical cumulative distribution function (c.d.f.) is then

$$\hat{F}_n(y) = \frac{1}{n} \sum_{i=1}^{n} I(y_i \leq y).$$

Many statistics for testing H_0: $F(y) = F_0(y)$ have been based on measures of the distance between $\hat{F}_n(y)$ and $F_0(y)$; see, for example, D'Agostino and Stephens (1986, Ch. 4) or Shorack and Wellner (1986). Two well-known examples are the Kolmogorov–Smirnov statistic

$$D_n = \sup_y |\hat{F}_n(y) - F_0(y)| \qquad (10.2.1)$$

and the Cramer–von Mises statistic

$$W_n^2 = n\int_{-\infty}^{\infty} [\hat{F}_n(y) - F_0(y)]^2 \, dF_0(y). \tag{10.2.2}$$

Sufficiently large values of D_n or W_n^2 provide evidence against the hypothesized model $F_0(y)$.

The distributions of D_n and W_n^2 do not depend on $F_0(y)$, assuming that H_0 is true. This is obvious from the alternative expressions (see Problem 10.1)

$$D_n = \max_{1\le i\le n}\left[\frac{i}{n} - F_0(y_{(i)}),\, F_0(y_{(i)}) - \frac{(i-1)}{n}\right] \tag{10.2.3}$$

$$W_n^2 = \sum_{i=1}^{n}\left[F_0(y_{(i)}) - \frac{(i-.5)}{n}\right]^2 + \frac{1}{12n}, \tag{10.2.4}$$

since under H_0 the $F_0(Y_{(i)})$'s are the order statistics in a random sample from the uniform distribution on (0, 1). Finite sample and asymptotic distributions for D_n and W_n^2 are available; Stephens (1974) and D'Agostino and Stephens (1986, Ch. 4) provide discussion and references.

When data are Type 2 or singly Type 1 censored, the statistics (10.2.1) and (10.2.2) are modified so that the supremum and integral are over the range $y \le C$, where C is the common censoring value. For W_n^2 the expression (10.2.4) is now replaced with

$$W_{n,r}^2 = \sum_{i=1}^{r}\left[F_0(y_{(i)}) - \frac{(i-.5)}{n}\right]^2 + \frac{r}{12n^2} - \frac{n}{3}\left[\frac{r}{n} - F_0(C)\right]^3, \tag{10.2.5}$$

where r is the number of uncensored observations and $C \ge y_{(r)}$ is the upper limit on observation. Finite sample and asymptotic distributions are available for these cases as well (D'Agostino and Stephens, Ch. 4).

The tests and distributional results just mentioned are of limited use; rarely does one wish to test a fully specified model $F_0(y)$. When $F_0(y)$ contains unknown parameters, $\boldsymbol{\theta}$, statistics such as (10.2.1) or (10.2.2) can be modified by replacing $F_0(y)$ with the maximum likelihood estimate (m.l.e.) $F_0(y; \hat{\boldsymbol{\theta}})$. The distribution theory for such statistics is then much more difficult, and in general depends on F_0 and on $\boldsymbol{\theta}$. However, asymptotic theory is available (e.g., see Problem 10.11), and simulation can often be used to approximate p-values for specified test statistics. The case of location-scale models is more tractable, since test statistics exist whose distribution under H_0 does not depend on $\boldsymbol{\theta}$. This is discussed in Section 10.3, along with the use of simulation.

Generalizations of the preceding approach to the case of arbitrarily right-censored data (and also some other incomplete data settings) can be based on comparisons of nonparametric and parametric estimates of the c.d.f., $F(t)$. With lifetime data, we

can consider, for example,

$$Z_n(t) = \sqrt{n}[\hat{F}_n(t) - F_0(t; \hat{\boldsymbol{\theta}})], \tag{10.2.6}$$

where $\hat{F}_n(t)$ is the Kaplan–Meier estimate of $F(t)$ and $F_0(t; \boldsymbol{\theta})$ is the hypothesized parametric family. An alternative approach (Hjort 1990a) is to consider

$$Z_n(t) = \sqrt{n}[\hat{H}_n(t) - H_0(t; \hat{\boldsymbol{\theta}})], \tag{10.2.7}$$

where $\hat{H}_n(t)$ is the Nelson–Aalen estimate and $H_0(t; \boldsymbol{\theta})$ is the parametrically specified cumulative hazard function (c.d.f.). The asymptotic theory for the processes $\{Z_n(t), t \ge 0\}$ is in either case rather complicated and does not lead to easily useable tests based on statistics analogous to (10.2.1) or (10.2.2). However, the use of (10.2.6) or (10.2.7) in connection with the approach of Section 10.2.3 is more promising, and in some cases p-values can be approximated by simulation.

10.2.2 Model Expansion and Smooth Tests

Suppose we wish to test the hypothesis that the probability density function (p.d.f.) $f(y)$ of a random variable Y has a specified parametric form,

$$H_0: f(y) = f_0(y; \boldsymbol{\theta}) \tag{10.2.8}$$

where $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters. The most widely used approach for testing models is to embed the model in a larger family $f(y; \boldsymbol{\theta}, \boldsymbol{\beta})$ such that $f(y; \boldsymbol{\theta}, \boldsymbol{\beta}_0) = f_0(y; \boldsymbol{\theta})$ for some $\boldsymbol{\beta}_0$. The hypothesis (10.2.8) is equivalent to $H_0: \boldsymbol{\beta} = \boldsymbol{\beta}_0$ within the extended model, and may be tested using standard parametric methods.

This approach has been used extensively in previous chapters. In general, H_0 can be tested by using a likelihood ratio test within the model $f(y; \boldsymbol{\theta}, \boldsymbol{\beta})$. In some cases it is convenient to use partial score tests of $H_0: \boldsymbol{\beta} = \boldsymbol{\beta}_0$ instead (see Appendix C). They require the m.l.e. of $\boldsymbol{\theta}$ under the model (10.2.8) only; it is not necessary to obtain the m.l.e.'s of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ in the extended model. To test hypotheses of the form (10.2.8), Neyman (1937) suggested a general method of extending a null model, combined with a partial score test. These are sometimes referred to as Neyman smooth tests and have received a good deal of subsequent study (see Rayner and Best 1989, 1990). The general approach is potentially useful, so let us consider it briefly.

Neyman suggested extended models with densities of the form

$$g(y; \boldsymbol{\theta}, \boldsymbol{\beta}) = C(\boldsymbol{\theta}, \boldsymbol{\beta}) f_0(y; \boldsymbol{\theta}) \exp\left[\sum_{j=1}^{k} \beta_j h_j(y, \boldsymbol{\theta})\right], \tag{10.2.9}$$

where the $h_j(y, \boldsymbol{\theta})$'s are specified orthonormal functions; the model (10.2.8) is given by $\boldsymbol{\beta} = \mathbf{0}$. Kopecky and Pierce (1979) and Thomas and Pierce (1979) proposed a

different extended family,

$$g(y; \boldsymbol{\theta}, \boldsymbol{\beta}) = C(\boldsymbol{\beta}) f_0(y; \boldsymbol{\theta}) \exp\left[\sum_{j=1}^{k} \beta_j F_0^j(y; \boldsymbol{\theta})\right], \tag{10.2.10}$$

where $F_0(y; \boldsymbol{\theta})$ is the c.d.f. corresponding to $f_0(y; \boldsymbol{\theta})$. Once again, $\boldsymbol{\beta} = \mathbf{0}$ gives the model (10.2.8). In either (10.2.9) or (10.2.10), one would typically use a small value of k (often 1 or 2).

Let us consider the model (10.2.10) a little further; for discussion of (10.2.9), see Rayner and Best (1989, Ch. 6). The partial score test of $H_0: \boldsymbol{\beta} = \mathbf{0}$ arising from uncensored data and the log-likelihood function

$$\ell(\boldsymbol{\theta}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \log g(y_i; \boldsymbol{\theta}, \boldsymbol{\beta})$$

is based on the statistic (C10), which here becomes

$$W_k = \mathbf{U}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_0, \mathbf{0})' \mathcal{I}_0^{\boldsymbol{\beta}\boldsymbol{\beta}} \mathbf{U}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}}_0, \mathbf{0}). \tag{10.2.11}$$

The m.l.e. $\hat{\boldsymbol{\theta}}_0$ is that obtained under the model (10.2.8), and the vector $\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}, \mathbf{0})$ has elements

$$\frac{\partial \ell}{\partial \beta_j}\Big|_{\boldsymbol{\beta}=\mathbf{0}} = \sum_{i=1}^{n} \left[F_0^j(y_i; \boldsymbol{\theta}) - (1+j)^{-1}\right] \qquad j = 1, \ldots, k. \tag{10.2.12}$$

Note that in (10.2.12), $(1+j)^{-1}$ is the expectation of $F_0^j(Y_i; \boldsymbol{\theta})$ under (10.2.8). The matrix $\mathcal{I}_0^{\boldsymbol{\beta}\boldsymbol{\beta}}$ is the $k \times k$ block of the partitioned inverse Fisher information matrix $\mathcal{I}(\boldsymbol{\theta}, \boldsymbol{\beta})^{-1}$ corresponding to $\boldsymbol{\beta}$. Thomas and Pierce (1979, Sec. 3) show that the only terms in $\mathcal{I}(\boldsymbol{\theta}, \boldsymbol{\beta})$ that do not already arise for the model (10.2.8) are

$$E\left[\sum_{i=1}^{n} \frac{\partial F_0^j(y_i; \theta)}{\partial \theta_\ell}\right] \qquad j = 1, \ldots, k; \qquad \ell = 1, \ldots, p. \tag{10.2.13}$$

They and Kopecky and Pierce (1979) compute the necessary expectations for $j \leq 4$ in the case of exponential, Weibull, and normal models (10.2.8).

Extending these tests to handle censored data has been problematic, since the expectations required to evaluate $\mathcal{I}(\boldsymbol{\theta}, \boldsymbol{\beta})$ are not available without a model for censoring. One might expect that the Fisher information matrix could just be replaced with the observed information $I(\boldsymbol{\theta}, \boldsymbol{\beta})$ evaluated at $(\hat{\boldsymbol{\theta}}_0, \mathbf{0})$. However, Gray and Pierce (1985) found that this matrix is not necessarily positive definite, so a general extension to deal with censored data is lacking.

Alternatives to models (10.2.9) and (10.2.10), which have advantages for censored data, can be based on the hazard function; Pena (1998) considers such models. In the current setting, the extended models would be of the form

$$h(t; \boldsymbol{\theta}, \boldsymbol{\beta}) = h_0(t; \boldsymbol{\theta}) \exp\left[\sum_{j=1}^{k} \beta_j g_j(t; \boldsymbol{\theta})\right], \tag{10.2.14}$$

where $h(t; \boldsymbol{\theta}, \boldsymbol{\beta})$ and $h_0(t; \boldsymbol{\theta})$ are the hazard functions for a lifetime variable T in the expanded and null models, respectively, and the $g_j(t; \boldsymbol{\theta})$'s are bounded predictable functions. In particular, the $g_j(t; \boldsymbol{\theta})$'s can be random in this formulation. The log-likelihood function from a censored random sample (t_i, δ_i), $i = 1, \ldots, n$, when written in the form (2.2.17), gives the likelihood score vector in the form (2.2.18). For the model (10.2.14) the derivatives $\partial \ell / \partial \beta_j$ $(j = 1, \ldots, k)$ evaluated at $\boldsymbol{\beta} = \mathbf{0}$ and $\hat{\boldsymbol{\theta}}_0$, the m.l.e. of $\boldsymbol{\theta}$ in the null model, are

$$U_j(\hat{\boldsymbol{\theta}}_0, \mathbf{0}) = \sum_{i=1}^{n} \int_0^\infty g_j(t; \hat{\boldsymbol{\theta}}_0)[dN_i(t) - Y_i(t) h(t; \hat{\boldsymbol{\theta}}_0)\, dt]. \tag{10.2.15}$$

These statistics can be used for a partial score test of the hypothesis $\boldsymbol{\beta} = \mathbf{0}$ and have been investigated by Hjort (1990a), Pena (1998), and others. Note that when $k = 1$ and we use the data-dependent form $g_1(t; \boldsymbol{\theta}) = Y_\cdot(t)^{-1}$, (10.2.15) reduces to $n^{-1/2}$ times the test statistic (10.2.7) involving the Nelson–Aalen and parametric estimates of the c.h.f.

As discussed in Section 10.2.1, asymptotic distributions for statistics of the form (10.2.15) can be complicated in cases where the $g_j(t; \hat{\boldsymbol{\theta}}_0)$'s are data-dependent. For fully parametric models (10.2.14), if the functions $g_j(t; \boldsymbol{\theta})$ are fixed, then standard maximum likelihood theory produces tests that are in principle easily implemented. Care is needed in the specification of (10.2.14) to avoid identifiability or estimability problems, and to provide test statistics with good power against departures from $h_0(t; \boldsymbol{\theta})$. Limited simulation studies (Pena 1998) suggest that standard χ^2 large-sample approximations for the score statistics are sufficiently accurate for practical purposes.

10.2.3 Tests Based on Grouped Data

With grouped uncensored data, tests of fit can be based on the multinomial distribution; this yields the well-known Pearson and likelihood ratio tests. These are reviewed briefly below, following which the case of censored data is considered.

Take the usual grouped data setup where lifetimes T or other observations can fall into $k+1$ intervals $I_j = [a_{j-1}, a_j)$, $j = 1, \ldots, k+1$, with $a_0 = 0$, $a_{k+1} = \infty$ and a_k as an upper limit of observation. Let d_j represent the number of observations in a random sample of size n that lie in I_j and $p_j = Pr(T \in I_j)$ be the probability an observation lies in I_j, where $\sum p_j = 1$. The objective is to test some hypothesis

$$H_0: p_j = p_j(\boldsymbol{\theta}) \qquad j = 1, \ldots, k+1, \tag{10.2.16}$$

where the parameter $\boldsymbol{\theta}$ is of dimension $p < k$. In the case where T has a continuous lifetime distribution with c.d.f. $F(t; \boldsymbol{\theta})$,

$$p_j(\boldsymbol{\theta}) = F(a_j; \boldsymbol{\theta}) - F(a_{j-1}; \boldsymbol{\theta}). \tag{10.2.17}$$

Let $\hat{p}_{j0} = p_j(\hat{\boldsymbol{\theta}})$ denote the m.l.e.'s for p_j $(j = 1, \ldots, k+1)$ obtained from the multinomial likelihood function based on $d_1, \ldots, d_{k+1}$,

$$L(p_1, \ldots, p_k) = \prod_{j=1}^{k+1} p_j^{d_j}, \tag{10.2.18}$$

under the null hypothesis assumption (10.2.16). The Pearson statistic for testing H_0 is then

$$X^2 = \sum_{j=1}^{k+1} \frac{(d_j - e_j)^2}{e_j}, \tag{10.2.19}$$

where $e_j = np_j(\hat{\boldsymbol{\theta}})$. Under H_0, the limiting distribution of X^2 as $n \to \infty$ is $\chi^2_{(k-p)}$.

The likelihood ratio test of (10.2.16) is based on the statistic $\Lambda = 2 \log L(\hat{p}_1, \ldots, \hat{p}_k) - 2 \log L(\hat{p}_{10}, \ldots, \hat{p}_{k0})$, where $\hat{p}_j = d_j/n$ is the unrestricted m.l.e. for p_j from (10.2.18). It is easily seen that

$$\Lambda = 2 \sum_{j=1}^{k+1} d_j \log(d_j/e_j), \tag{10.2.20}$$

with e_j defined as for X^2. The limiting distribution of Λ under H_0 is $\chi^2_{(k-p)}$, and the tests based on (10.2.19) and (10.2.20), are asymptotically equivalent. A nice feature of these tests is that a comparison of observed class frequencies d_j and expected frequencies e_j under H_0 provides insight concerning departures from the hypothesized model (10.2.16). The tests are less powerful than specialized parametric tests at detecting specific types of departure from H_0, but are easily used and effective in many situations. More power to detect departures is provided by larger values of k, though we want k small enough that the χ^2 approximations for X^2 or Λ are accurate. D'Agostino and Stephens (1986, Ch. 3) discuss the choice of intervals.

When the grouped data arise from a continuous distribution, we sometimes know the exact values $t_1, \ldots, t_n$ of the observations. If the m.l.e. of $\boldsymbol{\theta}$ based on the ungrouped data is used in (10.2.19) or (10.2.20), the limiting distributions are no longer $\chi^2_{(k-p)}$, but are given by a linear combination of $\chi^2_{(k-p)}$ and $p\ \chi^2_{(1)}$ random variables (e.g., Chernoff and Lehmann 1954). This limiting distribution is bounded by $\chi^2_{(k-p)}$ and $\chi^2_{(k)}$, and a range for the exact significance level can be obtained from this.

10.2.3.1 Censored Data: Censoring at Interval Endpoints

If data are right censored at a common value C, then the Pearson or likelihood ratio tests can be used with $a_k = C$. More general censoring patterns create problems that can be overcome in various ways.

First consider the special case where all censoring occurs at the ends of intervals; this model was used for life table data in Section 3.6. The numbers of failures and withdrawals (censored lifetimes) in I_j are denoted by d_j and w_j, and the number of individuals alive and uncensored at a_{j-1} is denoted by n_j. As in Section 3.6, let $q_j = Pr(T \in I_j | T \geq a_{j-1})$ and note that the likelihood function (3.6.8),

$$L(q_1, \ldots, q_k) = \prod_{j=1}^{k} q_j^{d_j} (1 - q_j)^{n_j - d_j} \tag{10.2.21}$$

is valid under the assumption that censoring depends only on prior events.

Consider the null hypothesis

$$H_0: q_j = q_j(\boldsymbol{\theta}) \qquad j = 1, \ldots, k, \tag{10.2.22}$$

where $\boldsymbol{\theta}$ is of dimension $p < k$ and the alternative is that the q_j satisfy only $0 \leq q_j \leq 1$. The unrestricted m.l.e.'s from $L(q_1, \ldots, q_k)$ are $\hat{q}_j = d_j/n_j$ and the m.l.e.'s under H_0 are $\hat{q}_{j0} = q_j(\hat{\boldsymbol{\theta}})$. The likelihood ratio statistic for testing H_0 is

$$\Lambda = 2 \log L(\hat{q}_1, \ldots, \hat{q}_k) - 2 \log L(\hat{q}_{10}, \ldots, \hat{q}_{k0}).$$

This can be written as

$$\Lambda = 2 \sum_{j=1}^{k} d_j \log(d_j/\hat{d}_{j0}) + 2 \sum_{j=1}^{k} (n_j - d_j) \log[(n_j - d_j)/(n_j - \hat{d}_{j0})], \tag{10.2.23}$$

where $\hat{d}_{j0} = n_j \hat{q}_{j0}$ estimates the expected number of failures in I_j, given n_j. Under assumptions about censoring discussed in Section 3.6, the limiting distribution of Λ under H_0 is $\chi^2_{(k-p)}$. As for the earlier tests based on (10.2.19) and (10.2.20), this result applies when the m.l.e. $\hat{\boldsymbol{\theta}}$ is obtained from the grouped data likelihood (10.2.21); if $\hat{\boldsymbol{\theta}}$ were obtained from data before grouping, the limiting distribution would be that of a linear combination of $\chi^2_{(k-p)}$ and $p\ \chi^2_{(1)}$ random variables.

A Pearson-type test of (10.2.22) can also be developed for this case (Cook 1988, Ch. 3); the test statistic is

$$X^2 = \sum_{j=1}^{k} \frac{n_j(\hat{q}_j - \hat{q}_{j0})^2}{\hat{q}_{j0}(1 - \hat{q}_{j0})}. \tag{10.2.24}$$

In the case where there is no censoring, this statistic and the standard Pearson statistic (10.2.19) are different but asymptotically equivalent. Likelihood ratio procedures are, on the other hand, invariant under reparameterization, and (10.2.23) is the same as (10.2.20) when there is no censoring.

Table 10.1. Life Test Data for 89 Wood Specimens

Interval (year) j	At Risk (n_j)	Failures (d_j)	Withdrawals (w_j)
1	89	19	1
2	69	6	1
3	62	6	1
4	55	3	1
5	51	4	1
6	46	1	1
7	44	3	1
8	40	4	1
9	35	3	1

Example 10.2.1. Cook (1988, Ch. 3) presented data from a test of preservative treatments for wood, in which 89 pieces of wood were treated and partially embedded in soil. The test covered a period of many years. At the end of each year the pieces were examined and any with more than a specified amount of decay were designated as failures. In addition, one unfailed piece was randomly removed at the end of each year so it could be tested for strength. The data are shown in Table 10.1.

The Weibull distribution is widely used in this area, so let us carry out a test of fit. With the survivor function (s.f.) written as $S(t; \boldsymbol{\theta}) = \exp[-(\theta_1 t)^{\theta_2}]$, with t in years, the m.l.e.'s obtained by maximizing (10.2.21) are $\hat{\theta}_1 = .083$, $\hat{\theta}_2 = .594$. The values $\hat{q}_{j0}$ are then computed as

$$\hat{q}_{j0} = [S(j-1; \hat{\boldsymbol{\theta}}) - S(j; \hat{\boldsymbol{\theta}})]/S(j-1; \hat{\boldsymbol{\theta}}) \qquad j = 1, \ldots, 9.$$

The observed values of the likelihood ratio and Pearson test statistics (10.2.23) and (10.2.24) are $\Lambda = 4.56$ and $X^2 = 4.32$, respectively. Comparison with $\chi^2_{(7)}$ quantiles shows there is no evidence against the Weibull model.

10.2.3.2 Censored Data: General Case

If censoring times are distributed across intervals, then the approach of the preceding subsection no longer applies. If the intervals $j = 1, \ldots, k$ are not too long and censoring not too heavy, then an ad hoc approach in which the likelihood (10.2.21) is replaced with

$$L(q_1, \ldots, q_k) = \prod_{j=1}^{k} q_j^{d_j} (1 - q_j)^{n_j - d_j - .5w_j} \tag{10.2.25}$$

is often satisfactory (see Problem 3.18). Although the limiting distribution for the likelihood ratio statistic Λ under H_0: $q_j = q_j(\boldsymbol{\theta})$ is not $\chi^2_{(k-p)}$ and $\hat{\boldsymbol{\theta}}$ is not consistent for $\boldsymbol{\theta}$ under H_0, the test and associated χ^2 approximation are reasonable in many situations.

Tests can also be developed by considering pseudo interval frequencies $\tilde{d}_j$ defined through the Kaplan–Meier estimate of the distribution or survival function, or through the Nelson–Aalen estimate of the c.h.f. This approach assumes that, although goodness-of-fit is to be checked using grouping intervals, the exact failure and censoring times are available for individuals $i = 1, \ldots, n$.

An approach based on the Kaplan–Meier estimate is as follows. Let $\hat{S}(t)$ denote the Kaplan–Meier estimate (3.2.2) and for the interval $I_j = [a_{j-1}, a_j)$ define the pseudo relative frequencies

$$\tilde{p}_j = \hat{S}(a_{j-1}) - \hat{S}(a_j) \qquad j = 1, \ldots, k. \tag{10.2.26}$$

Consider the parametric model $S(t; \boldsymbol{\theta})$ and the corresponding hypothesis H_0 given by (10.2.16), with

$$p_j(\boldsymbol{\theta}) = S(a_{j-1}; \boldsymbol{\theta}) - S(a_j; \boldsymbol{\theta}).$$

Let $D(\boldsymbol{\theta})$ be a $k \times k$ symmetric matrix and consider estimators $\hat{\boldsymbol{\theta}}$ that minimize the quadratic form

$$[\tilde{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta})]' D(\boldsymbol{\theta})^2 [\tilde{\mathbf{p}} - \mathbf{p}(\boldsymbol{\theta})],$$

where $\tilde{\mathbf{p}} = (\tilde{p}_1, \ldots, \tilde{p}_k)'$ and $\mathbf{p}(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \ldots, p_k(\boldsymbol{\theta}))'$. Under mild conditions, Li and Doss (1993) show that, under H_0, $D(\hat{\boldsymbol{\theta}})[\tilde{\mathbf{p}} - \mathbf{p}(\hat{\boldsymbol{\theta}})]$ is asymptotically normal with mean $\mathbf{0}$ as $n \to \infty$. They also give a quadratic form in $[\tilde{\mathbf{p}} - \mathbf{p}(\hat{\boldsymbol{\theta}})]$ that has a limiting $\chi^2_{(k-p)}$ distribution, and can be used to test H_0.

Akritas (1988), Hjort (1990a), and others have considered tests based on the Nelson–Aalen estimate (3.2.13). The idea here is to compare the Nelson–Aalen estimate $\hat{H}(t)$ of the c.h.f. with the parametric estimate $H(t; \hat{\boldsymbol{\theta}})$ at time points $a_1, \ldots, a_k$. Conditional expected frequencies for the interval I_j are defined as

$$e_j = \int_{a_{j-1}}^{a_j} Y.(s) h(s; \hat{\boldsymbol{\theta}})\, ds \qquad j = 1, \ldots, k,$$

where $\hat{\boldsymbol{\theta}}$ is the m.l.e. based on the ungrouped data and where $Y.(s)$ is the total number of individuals alive and uncensored at time s. Asymptotically, χ^2 test statistics for H_0 that are the sum of

$$X^2 = \sum_{j=1}^{k} (d_j - e_j)^2 / e_j \tag{10.2.27}$$

and a second term that depends on the model $H(t; \boldsymbol{\theta})$ can be obtained. Hjort (1990a) indicates that a slightly conservative procedure is to use just X^2 and treat it as asymptotically $\chi^2_{(k)}$ under H_0.

Simulation might also be used to approximate p-values in many cases. For arbitrarily censored data, it is often satisfactory to assume the independent random-

censorship model (Section 2.2.16) and to estimate the censoring time distribution nonparametrically through the Kaplan–Meier estimate $\hat{F}(c)$. By simulating lifetimes T_i from $F_0(t; \hat{\theta})$ and censoring times C_i from $\hat{F}(c)$, bootstrap samples $\{t_i = \min(T_i, C_i), \delta_i = I(t_i = T_i), i = 1, \ldots, n\}$ and realizations of the goodness-of-fit test statistic can be obtained. This and simulation procedures when only grouped data are available do not appear to have been examined closely.

10.3 TESTS OF FIT FOR SPECIFIC DISTRIBUTIONS

Tests of fit for several important lifetime distributions are considered briefly in this section; all of the models are of the log-location-scale type. The development and use of tests with complete or Type 2 censored data is relatively easy. In addition to existing distribution theory, tables, and special-purpose software, one can use simulation to obtain p-values in most cases. With arbitrarily censored data the main approaches to model assessment are the graphical procedures introduced in Chapter 2 and hypothesis tests based on model expansion. Tests based on approaches such as those in Sections 10.2.1 and 10.2.3 tend to have complicated asymptotic distribution theory, although simulation can often be used for evaluation of p-values. These methods have not yet been carefully assessed.

10.3.1 Tests of Fit for the Exponential Distribution

We consider three approaches that deal with decreasing degrees of specificity concerning departures from the exponential model.

First are tests based on model expansion. By embedding the exponential distribution with survivor function

$$H_0: \quad S(t) = e^{-t/\theta} \qquad t > 0 \tag{10.3.1}$$

in a larger family, simple parametric methods may be used. Two models that include the exponential as a special case are the Weibull distribution (1.3.5) and the gamma distribution (1.3.15). The former, for example, has s.f.

$$S(t) = \exp[-(t/\theta)^{\beta}] \qquad t > 0 \tag{10.3.2}$$

and testing H_0 is equivalent to testing $\beta = 1$. Such parametric procedures have been considered in Chapters 4 and 5. They have the advantage of handling arbitrary censoring or other forms of incomplete data, but may lack power against departures from (10.3.1) that are not approximated by (10.3.2).

We next mention a test that has good power against alternative distributions with monotone hazard functions, but deals only with Type 2 censoring. It is constructed as follows: let $t_{(1)} < t_{(2)} < \cdots < t_{(r)}$ be the r smallest lifetimes in a random sample of n, and define the scaled spacings

$$W_i = (n - i + 1)(t_{(i)} - t_{(i-1)}) \qquad i = 1, \ldots, r,$$

where $t_{(0)} = 0$. The test is based on the so-called Gini statistic (Gail and Gastwirth 1978),

$$G_r = \left(\sum_{i=1}^{r-1} i W_{i+1}\right) \Big/ (r-1) \sum_{i=1}^{r} W_i. \tag{10.3.3}$$

Values of G_r close to 0 or 1 provide evidence against the exponential model; values close to zero suggest increasing hazard function alternatives, and values close to one suggest decreasing hazard function alternatives. Under H_0, the W_i/θ's are independent standard exponential random variables (see Theorem 4.1.1), and it therefore follows that the distribution of G_r depends on r but not on θ or n. The mean and variance of G_r under H_0 are .5 and $[12(r-1)]^{-1}$, respectively (see Problem 10.7), and for r larger than 20 the approximation $[12(r-1)]^{1/2}(G_r - .5) \sim N(0, 1)$ is sufficiently accurate for all practical purposes. Gail and Gastwirth (1978) provide tables for $n = 3, \ldots, 20$; alternatively, p-values for G_r are readily obtained by simulation.

Omnibus tests that aim to detect arbitrary departures from the exponential distribution can be based on the approach of Section 10.2.1. In this case it is necessary to replace values $F_0(y)$ in statistics such as (10.2.3) or (10.2.4) with $\hat{F}_0(y) = 1 - \exp(-y/\hat{\theta})$, where $\hat{\theta}$ is the m.l.e. of θ. Percentage points for certain statistics have been given for complete samples and for samples with a single censoring point; D'Agostino and Stephens (1986, pp. 133–145) provide details. For arbitrarily censored data the approaches based on (10.2.6) or (10.2.7) can be considered; Hjort (1990a) considers the latter possibility in some detail. However, distribution theory for such tests is often quite intractable, so unless the evaluation of p-values by simulation is an option, this approach is not very feasible.

Tests can be based on grouped data, as described in Section 10.2.3. This approach can be thought of as a comparison of empirical and parametric distribution functions at grouping interval endpoints.

Tests for the one-parameter distribution (10.3.1) can also be used for the two-parameter exponential model (4.5.6) in the case of complete or Type 2 censored data. This is because if $Y_{(i)}$ is the ith-order statistic from the two-parameter model, then $Y'_{(i)} = Y_{(i)} - Y_{(1)}$, $i = 2, \ldots, n$, can be treated as the order statistics in a sample of size $n - 1$ from (10.3.1).

Example 10.3.1. Proschan (1963) gave data on the time, in hours of operation, between successive failures of air-conditioning equipment in 13 aircraft. The data for plane number 3 are as follows:

90, 10, 60, 186, 61, 49, 14, 24, 56, 20, 79, 84, 44, 59, 29, 118, 25, 156, 310, 76, 26, 44, 23, 62, 130, 208, 70, 101, 208.

An important question is whether the times T_i between successive failures are i.i.d. This may be assessed informally, for example, by plotting t_i versus t_{i+1} $(i = 1, \ldots, 28)$, or formally by methods described in Chapter 11. We assume

here that the times are i.i.d. and test whether they are consistent with an exponential distribution (10.3.1).

A probability plot of the data as in Section 3.3.1 is reasonably close to linear, but let us consider formal tests of fit. A test of the hypothesis $\beta = 1$ within the Weibull model can be carried out using either large-sample methods or exact methods of Section 5.2.2. We consider the former with the test statistic $Z = (\log \hat{\beta} - 0)/se(\log \hat{\beta})$; this gives $\hat{\beta} = 1.294$, $Z = 1.85$, and a (two sided) p-value obtained from treating Z as $N(0, 1)$ of .065. The G test statistic (10.3.3) gives $G_{29} = .441$, and the $N(0, 1)$ approximation to $[12(28)]^{1/2}(G_{29} - .5)$ gives a p-value of .279. The empirical cumulative distribution function (e.c.d.f.) statistic (10.2.4) with $F_0(y)$ replaced by $1 - \exp(-y/\hat{\theta}) = 1 - \exp(-y/83.52)$ gives an observed value $W^2_{29} = .1216$. Using Table 4.11 in D'Agostino and Stephens (1986), we find the associated p-value to be about .23. Finally, a test could be based on a grouping of the data, as in Section 10.2.3. With the intervals [0,50), [50,100), [100,200), [200,∞) the grouped data m.l.e. for θ in (10.3.1) is found to be 79.11, and the Pearson and likelihood ratio statistics (10.2.19) and (10.2.20) are 1.89 and 1.80, respectively. On $\chi^2_{(1)}$ these give p-values of .17 and .18.

None of the tests indicates significant evidence against the exponential model. The test based on the Weibull model gives a smaller p-value than the others; it is sensitive at detecting monotone hazard function departures from (10.3.1). An examination of an exponential probability plot of the data or a comparison of the e.c.d.f. and $F(t; \hat{\theta})$ suggests that the three largest failure times are largely responsible for the Weibull test's mild indication that β in (10.3.2) could be bigger than 1.

10.3.2 Tests of Fit for the Weibull and Extreme Value Distributions

Consider the extreme value model with survivor function (1.3.9),

$$S(y; u, b) = \exp[-e^{(y-u)/b}] \qquad -\infty < y < \infty, \tag{10.3.4}$$

where $-\infty < u < \infty$ and $b > 0$ are unknown parameters. Because $T = \exp(Y)$ has a Weibull distribution (1.3.6), tests of (10.3.4) also provide tests of fit for the Weibull model. As for the exponential model, we mention tests based on model expansion, on order statistics, and on e.c.d.f. statistics.

As discussed and illustrated in Sections 5.5.2 and 5.5.3, the extreme value model can be tested as a submodel of either the generalized log-Burr or generalized log-gamma families of distributions. We therefore turn to the second approach, which involves spacings of order statistics and is similar in spirit to the test of exponentiality based on (10.3.3). These tests apply to complete or Type 2 censored data only.

The spacings tests are based on the fact that if $Y_{(i)}$, $i = 1, \ldots, r$ are the r smallest observations in a random sample of size n from a location-scale distribution, then the normalized spacings

$$L_i = \frac{Y_{(i+1)} - Y_{(i)}}{E[Y_{(i+1)} - Y_{(i)}]} \qquad i = 1, \ldots, r-1 \tag{10.3.5}$$

have distributions that do not depend on u or b. In addition they are asymptotically independent standard exponential random variables (Pyke 1965). Mann et al. (1973) and Mann and Fertig (1975) proposed goodness-of-fit tests for the extreme value distribution based on statistics of the form

$$M = \sum_{i=m+1}^{r-1} L_i \Big/ \sum_{i=1}^{m} L_i, \tag{10.3.6}$$

where $1 \le m \le r-2$. Under the extreme value model (10.3.4), the distribution of $mM/(r-1-m)$ is well approximated by the F distribution $F_{(2(r-1-m),2m)}$ for r bigger than 20 or so. Large values of M provide evidence against the extreme value model.

Tiku and Singh (1981) suggest the statistic

$$Z^* = 2\sum_{i=1}^{r-2}(r-i+1)L_i/(r-2)\sum_{i=1}^{r-1} L_i. \tag{10.3.7}$$

Large or small values of Z^* provide evidence against the extreme value model, and they provide a normal approximation for calculation of p-values.

The statistics (10.3.6) and (10.3.7) require the expected values $E[Y_{(i+1)} - Y_{(i)}]$. Mann et al. (1973) provide tables for small n, and for larger n the approximation (Blom 1958, p. 73)

$$E[Y_{(i)}] = b \log\left[-\log\left(1 - \frac{i-.5}{n+.25}\right)\right] \tag{10.3.8}$$

can be used. Note that the scale parameter b disappears in (10.3.5) and (10.3.6), and that approximate p-values for either test can be calculated by simulating data from the standard extreme value distribution.

Tests based on e.c.d.f. statistics, such as (10.2.3) and (10.2.4), have also been studied. D'Agostino and Stephens (1986, Sec. 4.2) provide results for complete data. For either complete or Type 2 censored data, the distributions of these types of statistics under the null (extreme value) model do not depend on u or b, so the simplest approach is again to approximate p-values by simulating data from the standard extreme value distribution. The following example provides an illustration.

Example 10.3.2. Example 3.3.2 gave a Weibull probability plot of data on the failure times of 23 ball bearings in an endurance test. The plot looked roughly linear, with a slight amount of curvature at one end. Here we consider formal goodness-of-fit tests of the Weibull model; we do this by testing that the log failure times follow an extreme value distribution.

The M statistics (10.3.6) can be computed using values of $E[Y_{(i+1)} - Y_{(i)}]$ given by Mann et al. (1973) or else they can be approximated using (10.3.8). The M statistic with $m = 11$ ($r = n = 23$ here) is considered by Mann et al., so we use it, finding that $M = 1.302$. The approximation (10.3.8) gives $M = 1.323$. The p-value is

given by tables in Mann et al. as just over .25, indicating there is no evidence against the extreme value model. The approximation $M \sim F_{(22,22)}$ mentioned following (10.3.6) gives the p-value for $M = 1.302$ as .27 and for $M = 1.323$ as .26. If one does not have access to the tables or wants to consider other test statistics of the form (10.3.6), then exact p-values can be estimated to any desired degree of accuracy by simulation. To do this we generate B random samples of size 23 from the standard extreme value model ($u = 0$, $b = 1$) and for each sample obtain the statistic M. The estimated p-value is the fraction of the B values of M that exceed 1.302 or 1.323, depending on whether we used (10.3.8) or not in the calculation of M. For example, using $B = 1000$, we estimated the p-value as .28 for the M-value 1.323 based on (10.3.8).

Tests based on the e.c.d.f. can also be used. The m.l.e.'s of the extreme value location and scale parameters were obtained in Example 5.5.2 as $\hat{u} = 4.405$, $\hat{b} = .476$. The Cramer–von Mises statistic (10.2.4) with $F_0(y) = 1 - \exp[-\exp((y - \hat{u})/\hat{b})]$ gives $W_{23}^2 = .05793$. Table 4.17 in D'Agostino and Stephens (1986) shows the p-value to be over .25. The exact p-value can be approximated by simulation by generating B standard extreme value samples as earlier, obtaining the m.l.e.'s $\hat{u}$, $\hat{b}$, and the value of W_{23}^2 for each, and then determining the fraction of these B values that exceed .05793. Using the same $B = 1000$ samples as for M, we estimated the p-value for W_{23}^2 as .41.

Finally, consider the parametric test of fit obtained by embedding the extreme value distribution within the generalized log-gamma family. This was considered in Example 5.5.2, where the likelihood ratio statistic for testing the extreme value model was found to be 1.45. The p-value, obtained from the approximation $P(\chi_{(1)}^2 \geq 1.45)$, is .23. In this case, all of the goodness-of-fit tests are in close agreement, and there is no evidence against the extreme value model.

10.3.3 Tests of Fit for the Normal and Log-Normal Distributions

Many tests of the normal distribution have been proposed, but relatively few handle censored data. As for the extreme value model, we mention three types of approach.

The normal distribution is a special case within several larger families, so parametric tests of normality are always available. Two useful models that include the normal are the generalized log-gamma model (5.5.10) and the family of Student's-t distributions with location and scale parameters. The former is useful for dealing with asymmetric alternatives, and the latter deals with symmetric longer tailed alternatives to the normal distribution.

Among omnibus tests, one that has been found effective with uncensored data is the Shapiro–Wilk (Shapiro and Wilk 1965) test. It is based on the statistic

$$W = \left(\sum_{i=1}^{n} a_i y_{(i)}\right)^2 \Big/ \sum_{i=1}^{n} (y_i - \bar{y})^2, \tag{10.3.9}$$

where $y_1, \ldots, y_n$ is a random sample and the a_i are functions of means, variances, and covariances of standard normal-order statistics. Small values of W indicate

departures from normality. Tables (e.g., Pearson and Hartley 1972, Table 15) and software are available for this test.

Tiku (1981) gave a test based on spacings that can handle Type 2 censored data. It uses the same form of statistic (10.3.7) as for the extreme value distribution, except that the values $E[Y_{(i+1)} - Y_{(i)}]$ in (10.3.5) are now based on a normal sample. Tiku gives a large-sample approximation for the null distribution of the statistic; p-values can also be obtained by simulation. This test is effective against asymmetric alternatives to normality, but not against symmetric alternatives.

Finally, tests based on the e.c.d.f. can be used with either complete or Type 2 censored samples, in the same way as for the extreme value model. D'Agostino and Stephens (1986, Sec. 4.8) survey this area and provide tables. Simulation is a recommended way to obtain p-values with these types of tests when tables or software are not available.

Example 10.3.3. (Example 10.3.2 continued) Consider once again the ball bearing failure time data for which a Weibull model was found consistent in Example 10.3.2. Here we examine the adequacy of a log-normal model by testing a normal model for log failure time.

The Shapiro–Wilk test based on (10.3.9) can be carried out with the aid of Tables 15 and 16 in Pearson and Hartley (1972). Table 15 gives the observed value $W = .984$ and Table 16 gives a p-value of approximately .93. The Cramer–von Mises test based on (10.2.4) with $F_0(y)$ given by $\Phi[(y - \bar{y})/s] = \Phi[(y - 4.1504)/.5333]$ gives $W_{23}^2 = .0289$. Tables in D'Agostino and Stephens (1986, Sec. 4.8) only show the p-value to be over .25. Finally, the parametric likelihood ratio test based on the generalized log-gamma family in Example 5.5.2 gave an observed value of .35 and a p-value on $\chi_{(1)}^2$ of .55. There is no evidence against the normal distribution of log failure times.

10.3.4 Additional Remarks

A few special tests for other frequently used lifetime distributions exist, mainly for the complete-sample case. For example, D'Agostino and Stephens (1986, p. 156) consider the logistic distribution. Since it is a location-scale model we can use any of the e.c.d.f. test statistics with parameters estimated by maximum likelihood in the case of complete or Type 2 censored samples, obtaining p-values by simulation.

Models for which log failure time distributions are not of location-scale form are more difficult to handle, since even for complete samples the distributions of test statistics typically depend on unknown parameter values. See, for example, O'Reilly and Rueda (1992) for a discussion of the inverse Gaussian model. Problem 10.11 gives the asymptotic distribution of $\sqrt{n}[\hat{F}_n(t) - F_0(t; \hat{\boldsymbol{\theta}})]$, where $\hat{F}_n(t)$ is the e.c.d.f. of $F(t)$ from a complete sample. When there is a censoring process that is specified fully enough, simulation can be used to compute approximate p-values for test statistics, as described at the end of Section 10.2.3 for statistics based on grouped data.

Finally, it is hard to discriminate between certain models when sample sizes are small. For example, Tiku and Singh (1981) show the power of some complete-sample goodness-of-fit tests for the extreme value model against normal and logistic alternatives. For $n = 25$, the power of a size .05 test based on the statistic (10.3.7) is .39 against a normal alternative; for $n = 40$ the power is .62. In settings where it is of interest to compare alternative parametric models, guidance on power or choice of sample size can be obtained by simulation.

10.4 TESTS OF FIT WITH REGRESSION MODELS

10.4.1 General Remarks

For the case of fixed covariates $\mathbf{x}$ a regression model is a specification of the distribution $F(y|\mathbf{x})$ of a response variable Y, given $\mathbf{x}$. In many applications the most crucial issue is the nature of the relationship between Y and $\mathbf{x}$, the precise form of the distribution being secondary. For example, it may be more important to know that a location-scale model

$$F(y|\mathbf{x}) = F_0\left(\frac{y - u(\mathbf{x})}{b}\right), \qquad -\infty < y < \infty \tag{10.4.1}$$

provides a reasonable description than to know the precise form of the c.d.f. $F_0(z)$. A practical difficulty is that the assessment of different components of a model usually cannot be separated fully. For example, checks on the form of F_0 in a model (10.4.1) could be affected by departures from an assumed form for $u(\mathbf{x})$ or by nonconstancy of the scale parameter. Model specification involves the iterative application of model fitting and model checking, and it is easiest if at the start some robust assessment of the form of $F(y|\mathbf{x})$ is possible. This can be difficult when there are many covariates present.

Sections 6.1 and 6.2 discussed regression models and methods for assessing their suitability in some detail. Other parts of Chapter 6 considered location-scale models further, and Section 7.1.9 presented methods for checking PH models. The emphasis in those sections was on plotting techniques and on tests based on model expansion. Sometimes omnibus tests of the full specification of $F(y|\mathbf{x})$ are also helpful. Such tests can be associated with the methods of Sections 10.2 and 10.3, and we consider them briefly here.

Many omnibus tests are based on the fact that if a specific continuous model $F(y|\mathbf{x})$ is "true," then there exist functions of the data and parameters that have known distributions. For continuous responses, for example, the quantities $U_i = F(Y_i|\mathbf{x}_i)$ are independent Uniform(0, 1) random variables, given $\mathbf{x}_1, \ldots, \mathbf{x}_n$. This suggests that one might consider "uniform residuals" $\hat{u}_i = F(y_i|x_i; \hat{\boldsymbol{\theta}})$ and test statistics that "compare" the $\hat{u}_i$ to Uniform(0, 1) variables. The $\hat{u}_i$ are only approximately Uniform(0, 1), independent, or identically distributed, so consideration must be given to the distribution of any such statistic. In addition, censoring or other forms

of incompleteness in the data may make it difficult to find test statistics, and in that case the reliance on model expansion is more pronounced.

The next two sections consider the location-scale and multiplicative hazards models that were studied in Chapters 6 and 7.

10.4.2 Location-Scale Regression Models

Chapter 6 presented many examples where location-scale models (10.4.1) were checked using plots or tests based on model expansion. Assessments of the location-scale form, the constancy of b, and the specification of $u(\mathbf{x})$ were considered. In addition, tests of the form of the "error" distribution $F_0(z)$ were based on expanded parametric models $F_0(z; k)$ in Section 6.4. Another approach in this direction would be to use smooth tests discussed in Section 10.2.3; this has not yet been fully explored for censored data settings.

Tests of fit like those in Section 10.3 can be based on residuals from location-scale models. Consider the hypothesis H_0 that $F_0(z)$ in (10.4.1) is of some specified form, such as standard normal or standard extreme value. Assume further that the model is linear with $u(\mathbf{x}) = \boldsymbol{\beta}'\mathbf{x}$, which includes an intercept term. Let $\hat{\boldsymbol{\beta}}$ and $\hat{b}$ be the m.l.e.'s of $\boldsymbol{\beta}$ and the scale parameter b obtained from an uncensored random sample $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$, and define residuals

$$\hat{z}_i = \frac{y_i - \hat{\boldsymbol{\beta}}'\mathbf{x}_i}{\hat{b}} = \frac{y_i - \hat{u}_i}{\hat{b}} \qquad i = 1, \ldots, n. \tag{10.4.2}$$

Tests of H_0 can be based on straightforward generalizations of statistics used to test the hypothesized distribution $F_0(z)$ in the ordinary location-scale model where

$$Y_i \sim F_0[(y - u)/b] \qquad i = 1, \ldots, n \tag{10.4.3}$$

are i.i.d. random variables.

For example, consider a test based on the Cramer–von Mises form (10.2.4) of e.c.d.f. statistics. The statistic to be used in the regression setting is

$$W_n^2 = \sum_{i=1}^{n}\left[F_0(\hat{z}_{(i)}) - \frac{(i - .5)}{n}\right]^2 + \frac{1}{12n}, \tag{10.4.4}$$

where $\hat{z}_{(i)}$ is the ith smallest value among $\hat{z}_1, \ldots, \hat{z}_n$. Pierce and Kopecky (1979) and Loynes (1980) show that (10.4.4) and, more generally, other test statistics that are permutationally invariant functions of the $\hat{z}_i$ have the same limiting distribution in the regression case as in the i.i.d. case (10.4.3). This means that asymptotic approximations for the i.i.d. case can be used in the regression setting for sufficiently large n, although the accuracy for finite n might need investigation. A more important fact is that the distributions of (10.4.4) and other test statistics based on the residuals $\hat{z}_i$ is parameter-free (independent of $\boldsymbol{\beta}$ and b) for all n, although they depend on $\mathbf{x}_1, \ldots, \mathbf{x}_n$. This follows directly from Theorem E5 in Appendix E, which shows the $\hat{z}_i$ are ancillary. Therefore, p-values for any such test statistic can be obtained by simulation, as follows.

1. Generate response values $y_1, \ldots, y_n$ from the standard distribution $F_0(z)$, that is, (10.4.1) with $\boldsymbol{\beta} = \mathbf{0}$ and $b = 1$.
2. Obtain the m.l.e.'s $\hat{\boldsymbol{\beta}}$, $\hat{b}$ under (10.4.1), using the data $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$.
3. Obtain the residuals $\hat{z}_i$ in (10.4.2) and the test statistic, say W_n^2.

By repeating this process B times, we generate a random sample from the distribution of W_n^2, given $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and can use this to estimate the probability that W_n^2 exceeds any observed value.

The simulation procedure can be applied quite generally to test statistics based on the $\hat{z}_i$ and, in particular, to statistics that are not symmetric functions of $\hat{z}_1, \ldots, \hat{z}_n$. However, it applies to uncensored samples only; test statistics defined for arbitrarily censored samples do not have distributions that are parameter-free. In the spirit of the parametric bootstrap, one could obtain approximate p-values with censored data by simulating data from the hypothesized family with parameters $\hat{\boldsymbol{\beta}}$, $\hat{b}$. The censoring process must be known or estimated in order to generate relevant censored data; this can be problematic when the censoring times are related to covariate values. Some authors (e.g., Akritas and Torbeyns 1997) have developed test statistics that have asymptotic χ^2 distributions under the null model, but the limiting distributions are obtained under assumptions about censoring that may not be reasonable in certain settings. Further development of this area would be useful.

10.4.3 Multiplicative Hazards Models

The PH model, for which the hazard function of lifetime T given covariate vector $\mathbf{x}$ is of the form

$$h(t|\mathbf{x}) = h_0(t)g(\mathbf{x}; \boldsymbol{\beta}), \tag{10.4.5}$$

was discussed at length in Chapter 7. The multiplicative hazards extension with $\mathbf{x}$ depending on t was also considered. Model checking, based on plots and on model expansion, was examined in Section 7.1.9.

There are two main assumptions in any model of the form (10.4.5): the multiplicative assumption that $h(t|\mathbf{x})$ can be represented as a function of t times a function of $\mathbf{x}$, and the assumption regarding the form of $g(\mathbf{x}; \boldsymbol{\beta})$. Most applications of (10.4.5) leave $h_0(t)$ unspecified, so there is no parametric form to be checked or tested. The procedures described in Section 7.1.9 are satisfactory for most practical purposes, but there have been some attempts to develop omnibus tests of fit. These are effectively based on comparisons of c.h.f.'s $H(t|\mathbf{x}) = H_0(t)g(\mathbf{x}; \boldsymbol{\beta})$ or s.f.'s $S(t|\mathbf{x}) = \exp[-H(t|\mathbf{x})]$ estimated under (10.4.5) with nonparametric estimates of $H(t|\mathbf{x})$ or $S(t|\mathbf{x})$; for examples, see McKeague and Utikal (1991), Marzec and Marzec (1996), or Burke and Yuen (1995). Some approaches use grouping intervals for lifetime and covariates, and compare expected and observed frequencies (e.g., Schoenfeld 1980). This approach, however, can often be viewed as arising through a parametric test of (10.4.5) against an expanded model in which covariate-time interactions are introduced. Other goodness-of-fit tests, based on expanded models, have also been proposed. Quantin et al. (1996), for example, consider the family

$$H(t|\mathbf{x}) = \exp(\boldsymbol{\beta}'\mathbf{x})H_0(t)^{\exp(\boldsymbol{\gamma}'\mathbf{x})}, \tag{10.4.6}$$

where the case $\boldsymbol{\gamma} = \mathbf{0}$ gives a PH model.

Parametric PH models are sometimes used, in which $h_0(t)$ is specified as $h_0(t; \boldsymbol{\alpha})$. In this case, one can consider omnibus test statistics based on residuals such as

$$\hat{e}_i = \hat{H}(t_i|x_i) = H_0(t; \hat{\boldsymbol{\alpha}})g(\mathbf{x}_i; \hat{\boldsymbol{\beta}}), \tag{10.4.7}$$

where $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are m.l.e.'s. The functions $e_i = H(T_i|\mathbf{x}_i)$ are independent standard exponential random variables when (10.4.5) with $h_0(t) = h_0(t; \boldsymbol{\alpha})$ is true, so one might, for example, consider statistics that compare the $\hat{e}_i$ to censored exponential samples, either conditionally on $\mathbf{x}$ values or unconditionally. Approximate p-values could be obtained by simulation, assuming sufficient knowledge of the censoring process. Lin and Spiekerman (1996) consider tests of $H_0(t; \boldsymbol{\alpha})$ based on a comparison of $H_0(t; \hat{\boldsymbol{\alpha}})$ and the semiparametric Breslow estimator (7.1.32). They also proposed omnibus tests based on martingale residuals for both proportional hazards and accelerated failure time models.

The need to identify the nature of departures from a fitted model generally points to directed tests, so the appeal of omnibus tests of fit seems less in the regression setting than in the case of identically distributed observations. Nevertheless, empirical study of their effectiveness for proportional hazards (and other) regression models would be useful.

BIBLIOGRAPHIC NOTES

Goodness-of-fit tests associated with empirical processes, as in Section 10.2.1, have an extensive literature. D'Agostino and Stephens (1986, Ch. 4) give many references; see also Durbin (1973, 1975). Shorack and Wellner (1986) discuss empirical processes generally. The literature dealing with censored data is more limited. Hjort (1990a), Andersen et al. (1993, Sections 6.3, 6.4) and Y. Sun (1997) give some results and references; see also references below for grouped data. Fleming and Harrington (1991) and Andersen et al. (1993) consider empirical processes associated with the Kaplan–Meier and Nelson–Aalen estimates.

Smooth tests of fit are discussed by Rayner and Best (1989), but they do not consider censored data or regression models in any detail. Gray and Pierce (1985) consider tests based on an extension of (10.2.10). Pena (1998) provides many results and references for hazard-based models.

Tests based on grouped data have a long history. D'Agostino and Stephens (1986, Ch. 3) has many references and extensions to the material in Section 10.2.3 for uncensored data, including the choice of grouping intervals and the use of data-defined intervals. Turnbull and Weiss (1978) considered censored data tests based on (10.2.23); Cook (1988) examined these and related tests and carried out some simulation studies. Tests for which the Kaplan–Meier estimate from ungrouped data

is used to define expected frequencies are considered by Habib and Thomas (1986), Li and Doss (1993), and others. Akritas (1988) and Hjort (1990a) consider similar tests based on the Nelson–Aalen estimate.

Tests for the exponential distribution have been studied extensively, but rather few methods deal with arbitrarily censored data. D'Agostino and Stephens (1986, Ch. 10) provide an overview, and also (1986, Ch. 9) survey tests of normality.

Tests for regression models based on functions of residuals were considered by Pierce and Kopecky (1979) and Loynes (1980), building on work of Durbin (1975). Gray and Pierce (1985) consider smooth tests designed to assess the uniformity of residuals $F(t_i|\mathbf{x}_i; \hat{\boldsymbol{\theta}})$. In principle similar tests could be used with time-varying covariates in parametric models (Duchesne and Lawless 2000). Pena (1998) studies hazard-based smooth tests. Lin and Spiekerman (1996) consider statistics based on martingale residuals.

Omnibus tests for the Cox PH model have been proposed by various authors. In addition to those already mentioned, the test of Lin and Wei (1991) is interesting; it uses the "IM" approach of White (1982), and is based on a comparison of robust and model-based variance estimates for $\hat{\boldsymbol{\beta}}$. Many goodness-of-fit tests are based on functions of martingale residuals; Therneau and Grambsch (2000) and Verweij et al. (1998) provide discussion and references. It should be remarked that the Cox model is semiparametric and the idea of omnibus tests for it is a little different than for fully parametric models. Many of the so-called omnibus tests in the literature can be obtained as tests of parametric hypotheses within an expanded model (e.g., see Moreau et al, 1986, May and Hosmer 1998).

PROBLEMS AND SUPPLEMENTS

10.1 Show that the statistics (10.2.1) and (10.2.2) can be written in the forms (10.2.3) and (10.2.4) by noting that for an uncensored sample $y_1, \ldots, y_n$ the e.c.d.f. satisfies $\hat{F}_n(y_{(i)}) = i/n$ and $\hat{F}_n(y_{(i)}-) = (i-1)/n$ for $i = 1, \ldots, n$. Generalize both expressions to the case of Type 2 censored data $y_{(1)} < \cdots < y_{(r)}$.

(Section 10.2.1)

10.2 Consider a lifetime distribution with parametric hazard function $h_0(t; \boldsymbol{\alpha})$. Derive smooth tests of fit for the model by considering the expanded family

$$h(t; \boldsymbol{\alpha}, \beta) = h_0(t; \boldsymbol{\alpha}) e^{\beta g(t)},$$

where $g(t)$ is a specified function. In particular, consider the partial score test of H_0: $\beta = 0$. Specialize this to obtain tests for the exponential distribution.

(Section 10.2.2)

10.3 Consider the grouped data goodness-of-fit tests based on (10.2.23) and (10.2.24) in the case of a simple null hypothesis H_0: $q_j = q_{j0}$ $(j = 1, \ldots, k)$,

where the q_{j0} are specified probabilities. One concern about these tests is that they may lack power against certain types of departures from H_0. We can develop parametric tests of H_0 by considering the expanded family of models

$$\psi(q_j) = \psi(q_{j0}) + \beta, \tag{10.5.1}$$

where $\psi(x)$ is a function taking values in $(-\infty, \infty)$ for x in $(0, 1)$.

(a) Derive the partial score test statistic for the hypothesis $\beta = 0$, assuming that the observed data give a likelihood function of the form (10.2.21).

(b) Show that in the case where $\psi(x) = \log(x/(1-x))$, the statistic derived in part (a) specializes to give

$$X^2 = \sum_{j=1}^{k} (d_j - n_j q_{j0})^2 \Big/ \sum_{j=1}^{k} n_j q_{j0}(1 - q_{j0}) \tag{10.5.2}$$

in the notation of Section 10.2.3. The asymptotic distribution of X^2 is $\chi^2_{(1)}$ under H_0, assuming that censoring occurs only at interval endpoints.

(Section 10.2.3; Oleinick and Mantel 1970)

10.4 Table 10.2 is an adaptation of data given by Gail and Ware (1979) on the ages at death of a cohort of males who worked in a particular manufacturing plant during the period 1943–1947. The table shows the number of deaths and withdrawals (due to loss of follow-up) for various age intervals. Conditional probabilities of death q_j for the male population at large are also shown.

Compare the lifetime distribution for the cohort of workers with that for the population at large. You may want to combine the last two or three age intervals because of the large numbers of withdrawals there. Alternatively, you might consider a modification of (10.2.23) to handle withdrawals within intervals.

(Section 10.2.3)

Table 10.2. Survival of a Cohort of Manufacturing Plant Workers

Age Interval (years)	n_i	d_i	w_i	q_i
[40, 45)	191	8	0	.02529
[45, 50)	183	6	0	.03672
[50, 55)	177	11	0	.05709
[55, 60)	166	8	0	.08600
[60, 62)	158	5	14	.04640
[62, 64)	139	8	51	.05470
[64, ∞)	80	80	—	1.0

Table 10.3. Failure Time Frequencies for Centrifuge Cloths

Interval (weeks)	Number of Failures	Interval (weeks)	Number of Failures
[0, 2)	24	[30, 32)	4
[2, 4)	36	[32, 34)	4
[4, 6)	27	[34, 36)	5
[6, 8)	23	[36, 38)	2
[8, 10)	15	[38, 40)	2
[10, 12)	9	[40, 42)	2
[12, 14)	12	[42, 44)	2
[14, 16)	11	[44, 46)	2
[16, 18)	13	[46, 50)	0
[18, 20)	4	[50, 52)	4
[20, 22)	12	[52, 56)	0
[22, 24)	5	[56, 58)	1
[24, 26)	4	[58, 76)	0
[26, 28)	4	[76, 78)	1
[28, 30)	1		

10.5 Jardine (1979) presented the data in Table 10.3 concerning the time to failure for 229 sugar centrifuge cloths.

Assess the fit of (1) a two-parameter Weibull distribution, and (2) an exponential distribution to these data.

(Section 10.2.3)

10.6 Consider the current-status data on the times to initiation of cracks in metal turbine wheels in Example 3.5.5. Carry out a test of fit of the Weibull model considered in Example 4.3.2 by using a likelihood ratio statistic. Discuss whether a χ^2 approximation is liable to be adequate for computation of the p-value, and the alternative use of parametric bootstrap simulation.

(Section 10.2.3)

10.7 Consider the exponential goodness-of-fit statistic G given by (10.3.3). Noting that under an exponential distribution the quantities $D_i = W_i / \sum_{j=1}^{n} W_j$ $(i = 1, \ldots, n)$ have a Dirichlet distribution (see Wilks, 1962, p. 177), express G_r as

$$G_r = \frac{1}{r-1} \sum_{i=1}^{r-1} i D_{i+1}$$

and deduce that $E(G_r) = .5$ and $\mathrm{Var}(G_r) = [12(r-1)]^{-1}$.

(Section 10.3.1; Gail and Gastwirth 1978)

10.8 Let $t_{(1)} < \cdots < t_{(r)}$ be the r smallest observations in a random sample of size n that is hypothesized to come from an exponential distribution with unknown mean θ. Let $W_1, \ldots, W_r$ be defined as in Section 10.3.1 and Problem 10.6, and define

$$T_j = W_1 + \cdots + W_j.$$

(a) Show that the joint p.d.f. of $T_1, \ldots, T_{r-1}$, given $T_r = t_r$, is

$$\frac{(r-1)!}{t_r^{r-1}} \qquad 0 < t_1 < \cdots < t_{r-1} < t_r.$$

This is the distribution of the order statistics from a random sample of size $r-1$ from the uniform distribution on $(0, t_r)$; equivalently, the quantities $Z_i = T_i / T_r$ $(i = 1, \ldots, r-1)$ are distributed like the order statistics in a random sample of size $r-1$ from $U(0, 1)$.

(b) The result of part (a) removes the nuisance parameter θ and allows one to use the goodness-of-fit tests of Section 10.2.1, which are for completely specified distributions. Test the hypothesis that the data in Example 10.3.1 come from an exponential model by using the statistics D_n and W_n^2 (that is, test that the unordered Z_i are independent $U(0, 1)$ random variables). Compare the results of these tests with those carried out in the example.

(c) Suppose one plots Z_i against i/r, the expected value of the ith-order statistic in a random sample of size $r-1$ from the $U(0, 1)$ distribution, for $i = 1, \ldots, r-1$. Relate this plot to the tests in part (b). To what types of departures would you expect the tests in part (b) to be sensitive?

(Sections 10.2.1, 10.3.1)

10.9 Consider the data of Problem 5.8 on failure times for five samples of ball bearing specimens.

(a) Test that the data for each of samples 1 to 5 is a random sample from a Weibull distribution, using tests based on (10.3.7) or on the e.c.d.f. statistic (10.2.4). Use either tables mentioned in Section 10.3.2 or simulation to evaluate the p-value in each case.

(b) Discuss how to combine the results of the five tests. Two possibilities are to (1) combine the test statistics by addition, and (2) to calculate a combined p-value by Fisher's argument that if the null Weibull model is true for each sample, then the p-values α_j $(j = 1, \ldots, 5)$ should be realizations of independent Uniform(0, 1) random variables. Thus $W = -2 \sum \log \alpha_j$ should be $\chi^2_{(10)}$, and large values of W will provide evidence against the null model.

(Section 10.3.2)

10.10 *Confidence intervals for threshold parameters.* Suppose that T has a three-parameter Weibull distribution (4.5.1) with threshold parameter γ. Then $Y = \log(T - \gamma)$ has an extreme value distribution and if $T_{(1)} < \cdots < T_{(r)}$ are the r smallest observations in a random sample of size n, test statistics based on

spacings (10.3.5) for $Y_{(i)} = \log(T_{(i)} - \gamma)$ depend on γ but not on the other two model parameters. Use this fact to show how you could use either of the statistics (10.3.6) or (10.3.7) to obtain confidence intervals for γ, assuming the Weibull model to be correct. Discuss how to apply this approach with other log-location-scale models.

(Section 10.3.2; Mann and Fertig 1975)

10.11 Suppose that $y_1, \ldots, y_n$ are a random sample from a distribution with c.d.f. $F(y; \boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}}$ be the m.l.e. of $\boldsymbol{\theta}$ and $\hat{u}_i = F(y_i; \hat{\boldsymbol{\theta}})$ the uniform residuals $(i = 1, \ldots, n)$. Let $\hat{G}_n(u)$ be the e.c.d.f. based on $\hat{u}_1, \ldots, \hat{u}_n$; then Durbin (1973) showed that under standard regularity conditions the stochastic process $n^{1/2}[\hat{G}_n(u) - u]$ on $(0, 1)$ approaches a Gaussian process limit as $n \to \infty$. The covariance function for the limiting process is

$$\min(u, v) - uv - \mathbf{g}(u; \boldsymbol{\theta})'\mathcal{I}(\boldsymbol{\theta})^{-1}\mathbf{g}(v; \boldsymbol{\theta}), \qquad 0 < u < 1, \qquad 0 < v < 1,$$

where $\mathcal{I}(\boldsymbol{\theta}) = E[-\partial^2 \log f(y; \boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}']$ is the Fisher information matrix for a single observation and $\mathbf{g}(u; \boldsymbol{\theta}) = \partial F(y; \boldsymbol{\theta})/\partial\boldsymbol{\theta}$, evaluated at $y = F^{-1}(u; \boldsymbol{\theta})$

(a) If $F(y; \boldsymbol{\theta})$ is of location-scale form $F_0[(y - u)/b]$, show that the covariance function just given does not depend on u or b. Why must this be the case?

(b) Discuss how the limiting processes could be used to develop goodness-of-fit tests for an arbitrary model $F(y; \boldsymbol{\theta})$.

(Section 10.2; Durbin 1973)

10.12 Use simulation combined with the statistic (10.4.4) to test the adequacy of the extreme value regression model M_3 for the log failure voltages in Example 6.3.2. Compare the results with those based on model expansion in Examples 6.3.2 and 6.4.1.

(Section 10.4.2)

CHAPTER 11

Beyond Univariate Survival Analysis

11.1 INTRODUCTION

This book deals with the analysis of single lifetime variables, while taking into account explanatory variables and other factors. As noted in Section 1.5, many studies involve several types of events, leading to processes with multiple response, duration, or survival times for a single individual. The multiple failure modes setting of Chapter 9 is an example, since failures of different types can occur for an individual. However, it is a very simple example, since each individual can experience only one failure time, T, along with the failure mode, C.

More generally, several events may occur for each individual. Some examples will illustrate this, and how lifetime variables are associated with such events.

Example 11.1.1. Sometimes the units in a study consist of a group or "cluster" of two more individuals whose lifetimes are nonindependent. For example, in studying the effects of heredity on life length, investigators have considered the lifetimes (T_{1l}, T_{2l}) of identical twins (e.g., Hougaard et al. 1992).

Example 11.1.2. Stone (1978) gave data from experiments to investigate the failure of epoxy electrical-cable-insulation specimens under conditions of constant voltage stress. Failures occur because of a phenomenon known as electrical treeing. In this process there is an inception, or initiation, phase in which nothing appears to be happening (when viewed under a microscope), but at some point a defect becomes visible in the material and it then grows, eventually causing failure of the insulation. The data on each specimen consist of the time, T_1, to defect inception (or detection) and the subsequent additional time, T_2, from inception to failure. Table 11.1 shows data from an experiment in which 20 specimens were subjected to a 55-kV stress; times are in minutes. For three of the specimens inception still had not occurred when observation ceased. This resulted in T_1 being censored, and T_2 being unobservable.

Example 11.1.3. Data on patients treated for colon cancer (Moertel et al. 1990; Lin et al. 1999) were discussed in Example 3.3.3 and subsequently. Some of the

Table 11.1. Cable-Insulation Failure Data

Specimen	T_1	T_2	Specimen	T_1	T_2
1	228	30	11	1,227	39
2	106	8	12	254	46
3	246	66	13	> 2,440	—
4	700	72	14	435	85
5	473	25	15	1,155	85
6	> 1,740	—	16	> 2,600	—
7	155	7	17	195	27
8	414	30	18	117	27
9	1,374	90	19	724	21
10	128	4	20	300	96

patients have a recurrence of the disease and later die from it. Example 3.3.3 considered the time, T_1, from treatment to recurrence of the cancer, but two other lifetime variables are also of interest for an individual: T_2 = time from treatment to death, and T_3 = time from cancer recurrence to death from the cancer. The variables satisfy $T_2 \geq T_1$, so are not independent. Moreover, not all individuals may experience the event "cancer recurrence"; for those who do not, the variable T_3 does not exist. It is also possible that a person may die from other causes, either before or after colon cancer has occurred.

Example 11.1.4. Proschan (1963) gave the time intervals, in hours of operation, between successive failures of air conditioning equipment in a number of Boeing 720 aircraft; these data have been discussed by many authors (see, e.g., Lawless 2000). The times between failures for the first four planes are given below. Times appear in the order in which they occurred; that is, for plane 1 the first failure was after 194 hours of operation, the second was after 209 (= 194 + 15) hours, and so on.

Plane 1 194, 15, 41, 29, 33, 181 ($n = 6$)

Plane 2 413, 14, 58, 37, 100, 65, 9, 169, 447, 184, 36, 201, 118, 34, 31, 18, 18, 67, 57, 62, 7, 22, 34 ($n = 23$)

Plane 3 90, 10, 60, 186, 61, 49, 14, 24, 56, 20, 79, 84, 44, 59, 29, 118, 25, 156, 310, 76, 26, 44, 23, 62, 130, 208, 70, 101, 208 ($n = 29$)

Plane 4 74, 57, 48, 29, 502, 12, 70, 21, 29, 386, 59, 27, 153, 26, 326 ($n = 15$)

An important question is whether the times $T_1, T_2, \ldots$ between successive failures for individual aircraft can be considered to be independent and identically distributed (i.i.d.) or whether time trends exist.

Example 11.1.5. A study on the occurrence of pulmonary exacerbations for persons with cystic fibrosis (Fuchs et al. 1994) was described in Example 1.1.8 and

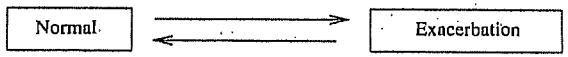

Figure 11.1. A process with recurring states.

considered in Example 3.2.4 and later. The individuals in the study were liable to experience exacerbation spells alternating with periods of normal (for them) pulmonary function. All individuals were exacerbation-free at the start of the study, and analysis of the data earlier in the book focused on the time, T_1, to the first exacerbation. However, the process in question involves possibly repeated events, which it is convenient to represent as follows. At any given time an individual is in either of two states: (1) normal pulmonary function, or (2) affected by an exacerbation. The process for an individual consists of a series of sojourns in states 1 and 2, as represented by Figure 11.1.

Events can be associated with a transition from one state to the other, and various lifetime variables can be defined within the process. The times T_j = length of the jth Normal period and S_j = length of the jth Exacerbation period ($j = 1, 2, \ldots$) are of interest.

The fields of multivariate failure time analysis and event history analysis deal with settings like those cited in the examples. A thorough treatment of these fields is beyond the scope of the book. However, many problems can be addressed using simple extensions of the univariate lifetime methodology from earlier chapters. The present chapter will show how this is done for problems involving clustered lifetimes (Section 11.2), sequences of lifetimes (Section 11.3), and more general processes (Section 11.4). References to more comprehensive treatments of multivariate lifetimes and event histories are given in the Bibliographic Notes at the end of the chapter.

Another topic that arises in lifetime settings concerns processes that are internal to individuals and may be associated with the lifetime in question. Such processes often describe time-varying physical characteristics of the individual, in which case they are often called marker or condition processes. For example, in studying survival time for persons infected with the human immunodeficiency virus (HIV), it is common to consider measures $Z(t)$ such as blood-cell counts or viral load. In studies of the lifetimes of electronic microdevices one may consider measures $Z(t)$ that reflect degradation or deterioration of a unit over time. These types of problems often require joint models for lifetime and the related processes. This topic is considered in Section 11.5.

11.2 MULTIVARIATE LIFETIME DISTRIBUTIONS AND CLUSTERED LIFETIMES

The preceding section gave examples of multiple lifetime variables $T_{1i}, T_{2i}, \ldots, T_{ki}$ associated with the ith individual in some population. Sometimes the individuals are

separate, physical units that can experience two or more events (Examples 11.1.2, 11.1.3, 11.1.4, 11.1.5), and sometimes they are clusters of two or more units on which lifetimes are defined (Example 11.1.1). In either case, the lifetimes $T_{1i}, \ldots, T_{ki}$ are not in general independent, so it is necessary to consider their joint distribution.

Approaches used with multiple lifetimes vary according to the setting. This section considers cases where the lifetimes $T_{1i}, \ldots, T_{ki}$ associated with an individual or cluster are unconstrained relative to one another. That is, there is no prior ordering associated with the times, and they can typically be observed in parallel, or simultaneously in real time. This would apply, for example, to the lifetimes T_{1i}, T_{2i} of a pair of twins, or to the times T_{1i}, T_{2i} until a certain condition appears in the left and right eyes of persons with diabetes (e.g., Huster et al. 1989). Situations where multiple lifetimes for an individual are associated with events that are time-ordered are considered in Sections 11.3 and 11.4.

In this section the term unit will be used to denote individuals or clusters of individuals within which the multiple lifetimes are defined.

11.2.1 Multivariate Lifetime Distributions

Suppose there are lifetime variables $T_1, \ldots, T_k$ associated with a typical unit. Denote their joint probability density function (p.d.f.) as $f(t_1, \ldots, t_k)$, and the survivor and distribution functions, respectively, as

$$S(t_1, \ldots, t_k) = Pr(T_1 \geq t_1, \ldots, T_k \geq t_k) \tag{11.2.1}$$

$$F(t_1, \ldots, t_k) = Pr(T_1 \leq t_1, \ldots, T_k \leq t_k) \tag{11.2.2}$$

for $t_1 \geq 0, \ldots, t_k \geq 0$. Continuous-time models will be emphasized here, in which case

$$f(t_1, \ldots, t_k) = \frac{(-1)^k \partial^k S(t_1, \ldots, t_k)}{\partial t_1 \ldots \partial t_k}. \tag{11.2.3}$$

The marginal survivor and distribution functions will be denoted by

$$S_j(t_j) = Pr(T \geq t_j), \qquad F_j(t_j) = Pr(T \leq t_j).$$

Hazard functions can be defined in various ways for multivariate models. For simplicity, consider the bivariate case. The hazard functions

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{Pr(T_j < t + \Delta t | T_1 \geq t, T_2 \geq t)}{\Delta t} \qquad j = 1, 2$$

$$\lambda_{12}(t_1|t_2) = \lim_{\Delta t \to 0} \frac{Pr(T_1 < t_1 + \Delta t | T_1 \geq t_1, T_2 = t_2)}{\Delta t} \qquad t_1 > t_2 \tag{11.2.4}$$

$$\lambda_{21}(t_2|t_1) = \lim_{\Delta t \to 0} \frac{Pr(T_2 < t_2 + \Delta t | T_1 = t_1, T_2 \geq t_2)}{\Delta t} \qquad t_1 < t_2$$

are often useful, and it is easily seen that they specify the joint distribution of T_1 and T_2. In terms of the joint survivor function $S(t_1, t_2)$, it is seen that

$$\lambda_1(t) = \frac{-\partial S(t_1, t_2)/\partial t_1}{S(t_1, t_2)} \Big|_{t_1 = t_2 = t}$$

$$\lambda_{12}(t_1|t_2) = \frac{-\partial^2 S(t_1, t_2)/\partial t_1 \partial t_2}{\partial S(t_1, t_2)/\partial t_2}, \qquad t_1 > t_2$$

with similar expressions for $\lambda_2(t)$ and $\lambda_{21}(t_2|t_1)$.

In applications, interest often centers on marginal and conditional distributions associated with (11.2.1), on the degree of association among $T_1, \ldots, T_k$, and on the relationship of covariates to these features. Parametric families that have been proposed as multivariate lifetime distribution models include the multivariate lognormal model, for which $(\log T_1, \ldots, \log T_k)$ is multivariate normal, and multivariate generalizations of Weibull and exponential distributions. A thorough discussion of this area is beyond the scope of this book, so only a few models will be considered. References to comprehensive sources are provided in the Bibliographic Notes, and some supplementary material is included in the Problems and Supplements section.

We will focus on problems in which the marginal distributions of the T_j are of primary interest; association among the T_j must be considered, but is secondary. Section 11.2.2 shows how univariate models and methods discussed earlier in the book can be applied in such settings. First, some multivariate models in which the marginal distributions have specific forms are considered.

11.2.1.1 Models with Specified Marginal Distributions

The bivariate case will be considered for simplicity. When the marginal distributions are continuous, it is possible to represent the joint survivor function $S(t_1, t_2)$ in the form

$$S(t_1, t_2) = C[S_1(t_1), S_2(t_2)], \tag{11.2.5}$$

where $S_j(t_j)$ is the marginal survivor function for T_j $(j = 1, 2)$ and C is a bivariate function called a copula. Since $U_1 = S_1(T_1)$ and $U_2 = S_2(T_2)$ are Uniform(0, 1) random variables, the function $C(u_1, u_2)$ must be a bivariate distribution function with uniform marginals defined on $0 \leq u_1 \leq 1, 0 \leq u_2 \leq 1$. In particular, $C(u_1, u_2)$ must satisfy $C(u_1, 1) = u_1$ and $C(1, u_2) = u_2$.

A useful way to develop bivariate lifetime models is through the specification of a parametric family of copulas $C(u_1, u_2; \phi)$ along with a specification of the marginal distributions. This gives a family of the form (11.2.5), in which the parameter ϕ determines the association or dependence structure. Fully parametric or semiparametric models are obtained by choosing parametric or semiparametric specifications for $S_1(t_1)$ and $S_2(t_2)$.

A family of models of this type was considered by Clayton (1978). With $\phi \geq 0$ it has the form (11.2.5) with

$$S(t_1, t_2) = [S_1(t_1)^{-\phi} + S_2(t_2)^{-\phi} - 1]^{-1/\phi}. \tag{11.2.6}$$

This model represents only positive association between T_1 and T_2 (see Problem 11.3). It can be extended to include negative association by allowing $\phi < 0$; in this case, the right side of (11.2.6) is negative for some pairs (t_1, t_2), but a proper joint survivor function is obtained by replacing it with 0 whenever it is negative. In the limit as $\phi \to 0$, (11.2.6) approaches the independence model $S(t_1, t_2) = S_1(t_1)S_2(t_2)$.

Fully parametric models in the family (11.2.6) are easily handled, as shown in the next section. In some applications the marginal survivor functions for T_1 and T_2 are constrained to be identical. Covariates $\mathbf{x}$ can be incorporated by using one of the approaches considered in Chapter 6 to model $S_j(t_j|\mathbf{x})$. The association parameter ϕ is usually assumed not to depend on $\mathbf{x}$. Semiparametric specifications for the $S_j(t_j)$'s in (11.2.6) are also of interest. This is a little more difficult to handle, but some approaches are discussed in Section 11.2.2.

An extension of (11.2.6) to the case of $k \geq 3$ variables is to take the joint survivor function as

$$S(t_1, \ldots, t_k) = \left[\sum_{j=1}^{k} S_j(t_j)^{-\phi} - (k-1)\right]^{-1/\phi} \tag{11.2.7}$$

for $\phi \geq 0$. This is easily seen to be a legitimate survivor function with marginals $S_j(t_j)$. A drawback of this model is that association among the variables $T_1, \ldots, T_k$ is governed by a single parameter, ϕ. This is adequate in cases where the T_j are exchangeable and the $S_j(t)$'s are identical, but is an undesirable restriction in many settings. Multivariate models in which $S(t_1, \ldots, t_k)$ takes the form $C[S_1(t_1), \ldots, S_k(t_k); \boldsymbol{\phi}]$ with $\boldsymbol{\phi}$ a vector of parameters can be considered; this provides more flexibility.

11.2.1.2 Some Other Models

Many parametric families of multivariate distributions have been studied in the literature, and can be applied to lifetime data. The multivariate log-normal family, in which $(\log T_1, \ldots, \log T_k)$ has a k-variate normal distribution, is often useful and allows flexibility in association structure. Some models, on the other hand, make strong assumptions about association. For example, an often discussed model is the multivariate Burr family (e.g., Crowder et al. 1991, p. 140; Hougaard 2000, p. 235), for which

$$S(t_1, \ldots, t_k) = \left[1 + \sum_{j=1}^{k} \lambda_j t_j^{\delta_j}\right]^{-\nu}. \tag{11.2.8}$$

Association is governed by the single parameter, ν (see Problem 11.4). In addition, the marginal survivor functions are

$$S_j(t_j) = [1 + \lambda_j t_j^{\delta_j}]^{-\nu},$$

so ν also affects the marginal distributions and is not purely an association parameter.

Another common approach to the specification of multivariate models is through random effects; Hougaard (2000) provides an extensive treatment of this topic. The standard approach is to define a random vector $\boldsymbol{\alpha}_i$ associated with the ith unit, and to assume that (1) given $\boldsymbol{\alpha}_i$, the lifetimes $T_{1i}, \ldots, T_{ki}$ are independent, and (2) the $\boldsymbol{\alpha}_i$ are i.i.d. across units $i = 1, \ldots, n$. A widely used model is the shared frailty family, in which $\boldsymbol{\alpha}_i$ is a positive-valued scalar and the hazard function for T_{ji} given α_i is of the form $\alpha_i \lambda_j(t)$ for $j = 1, \ldots, k$. If α_i has distribution function $G(\alpha)$, then the joint survivor function for $T_{1i}, \ldots, T_{ki}$ is

$$S(t_1, \ldots, t_k) = \int_0^\infty \exp\left[-\alpha \sum_{j=1}^{k} \Lambda_j(t_j)\right] dG(\alpha), \tag{11.2.9}$$

where $\Lambda_j(t)$ is the cumulative hazard function (c.h.f.) for T_{ji}, given α_i. An interesting special case is where α_i has a gamma distribution (1.3.15) with mean 1 and variance ϕ. It is easily seen that (11.2.9) then gives

$$S(t_1, \ldots, t_k) = \left[1 + \phi \sum_{j=1}^{k} \Lambda_j(t_j)\right]^{-1/\phi}. \tag{11.2.10}$$

This is of the same form as the Clayton model (11.2.7); to see this, note from (11.2.10) that the marginal survivor function for T_j is

$$S_j(t_j) = [1 + \phi \Lambda_j(t_j)]^{-1/\phi} \tag{11.2.11}$$

and rewrite $S(t_1, \ldots, t_k)$ in terms of the $S_j(t_j)$'s.

For $k > 2$ the fact that the model involves only a scalar random effect and a single parameter ϕ affecting association is often a drawback, and models with vector random effects are called for. Note also that although the families (11.2.7) and (11.2.10) have the same form, the marginal survivor functions (11.2.11) arising from (11.2.10) depend on ϕ, whereas in copula models of the form (11.2.7) it is typically assumed that the $S_j(t_j)$'s do not depend functionally on ϕ. As a specific illustration, note that by defining $\xi_j = \lambda_j \nu$, we can write the multivariate Burr model (11.2.8) as

$$S(t_1, \ldots, t_k) = \left[1 + \frac{1}{\nu} \sum_{j=1}^{k} \xi_j t_j^{\delta_j}\right]^{-\nu},$$

which is of the form (11.2.10) with $\phi = \nu^{-1}$ and $\Lambda_j(t) = \xi_j t^{\delta_j}$. Although this appears to be a rather general family, the marginal distributions and association struc-

ture are linked. For example, as $\nu \to \infty$, the case where $T_1, \ldots, T_k$ are independent is obtained, and the marginal distributions are Weibull:

$$S_j(t_j) = \exp(-\xi_j t_j^{\delta_j}) \qquad j = 1, \ldots, k. \tag{11.2.12}$$

For any other given value of ν, the marginal distributions are of another log-Burr form (1.3.21), determined by ν:

$$S_j(t_j) = \left[1 + \frac{1}{\nu}\xi_j t_j^{\delta_j}\right]^{-\nu}.$$

This may be contrasted with a model (11.2.7) in which the $S_j(t_j)$'s are of Weibull form (11.2.12) no matter what $\phi = \nu^{-1}$ equals.

Random-effects models like (11.2.9) and (11.2.10) with covariates added, and analogous models with fixed unit effects α_i, specify the relative effects of covariates on lifetimes $T_{1i}, \ldots, T_{ki}$ within the same unit. Thus, they are valuable in designed studies where different treatments or covariate factor levels are assigned to individuals within a unit or cluster. However, when marginal effects of covariates are of interest, models like (11.2.7) with covariate effects for the $S_j(t)$'s separated from association parameters are preferable. The following example illustrates these issues in the important context of paired data.

Example 11.2.1. Paired Data. Suppose that the effect of a single scalar covariate is of interest and that it is possible to run studies involving pairs of individuals, with different covariate values x_{1i}, x_{2i} assigned to the two individuals in pair i. A commonly used model is one where the hazard functions for T_{1i} and T_{2i} are of the form

$$\lambda_{ji}(t|x_{ji}) = \lambda_{0i}(t)e^{\beta x_{ji}}, \qquad j = 1, 2,$$

and where T_{1i} and T_{2i} are independent given x_{1i} and x_{2i}. In Problem 7.2, a stratified proportional hazards (PH) analysis with the pairs defining the strata was considered for inference about β. An alternative model is one for which

$$\lambda_{ji}(t|x_{ji}, \alpha_i) = \alpha_i \lambda_0(t)e^{\beta x_{ji}}, \qquad j = 1, 2,$$

where α_i is a cluster or pair effect. This makes the stronger assumption that the hazard functions $\lambda_{0i}(t)$ are proportional to one another. If the α_i are regarded as fixed effects, the stratified PH analysis is still the approach of choice, although if there is no censoring, some parametric alternatives are available (Holt and Prentice 1974). A more widely used approach (e.g., Wild 1983) is to assume that the α_i are i.i.d. random effects, independent of the x_{ji}; this offers the possibility of more efficient inferences for β.

The conditional (on α_i or on individuals being in the same pair) hazard ratio for T_{1i} and T_{2i} under these models is

$$\frac{\lambda_{1i}(t|x_{1i}, \alpha_i)}{\lambda_{2i}(t|x_{2i}, \alpha_i)} = \exp[\beta(x_{1i} - x_{2i})],$$

so $\exp(\beta)$ is the hazard ratio when $x_{1i} - x_{2i} = 1$. This is not the same as the hazard ratio for two randomly selected individuals not in the same pair. For example, if the α_i follow a gamma distribution with mean 1 and variance ϕ, then the joint distribution of T_{1i} and T_{2i} is of the form (11.2.10), and the marginal distribution for either, conditional on the covariate value x, is given by (11.2.11) as

$$S(t|x) = [1 + \phi\Lambda_0(t)e^{\beta x}]^{-1/\phi}.$$

The hazard function corresponding to this is $h(t|x) = -S'(t|x)/S(t|x)$, and so the ratio of the hazard functions for two random individuals from different pairs and with covariate values x_1 and x_2 is

$$\frac{h(t|x_1)}{h(t|x_2)} = e^{\beta(x_1 - x_2)}\left[\frac{1 + \phi\Lambda_0(t)e^{\beta x_2}}{1 + \phi\Lambda_0(t)e^{\beta x_1}}\right]. \tag{11.2.13}$$

This ratio does not equal $\exp(\beta)$ when $x_1 - x_2 = 1$, and in fact (11.2.13) is closer to one than $\exp(\beta)$ for all t if $\phi > 0$.

The hazard ratio (11.2.13) is not constant, unlike the pair-specific ratio. An alternative model for which the marginal hazard functions were of PH form might be desirable in some contexts. For example, one could use the bivariate model (11.2.7), with

$$S_j(t|x_{ji}) = \exp[-H_0(t)e^{\beta x_{ji}}], \qquad j = 1, 2.$$

Whether this model or the pair-specific random-effects PH model was preferable would depend on their relative abilities to fit the data and on the desire for a pair-specific or marginal representation of the covariate effect. It is often noted that if a positive stable law is taken as the distribution for the random-effects α_i (Hougaard 1986), then both the pair-specific and marginal hazard ratios are of PH form. However, the ratios take on different (constant) values, reflecting the fact that they are still measuring two different things.

The distinction between pair-specific and marginal effects disappears for log-location-scale models. For example, suppose that given i.i.d. random effects α_i, the joint distribution of $(Y_{1i}, Y_{2i}) = (\log T_{1i}, \log T_{2i})$ is specified by

$$Y_{ji} = \alpha_i + \beta_0 + \beta_1 x_{ji} + e_{ji}, \qquad j = 1, 2,$$

where e_{1i} and e_{2i} are i.i.d. and independent of α_i, and where $E(\alpha_i) = E(e_{ji}) = 0$. Then,

$$E(Y_{1i} - Y_{2i}|x_{1i}, x_{2i}, \alpha_i) = E(Y_{1i} - Y_{2i}|x_{1i}, x_{2i}) = \beta(x_{1i} - x_{2i}),$$

so the differences of conditional and marginal means coincide.

On a final point, it should be noted that the models discussed here all make the assumption that the conditional p.d.f.'s for T_{1i} satisfy

$$f(t_{1i}|x_{1i}, x_{2i}, \alpha_i) = f(t_{1i}|x_{1i}, \alpha_i),$$

and similarly for T_{2i}. Neuhaus and Kalbfleisch (1998) show that this is sometimes unwarranted, and consider models where, for example,

$$E(Y_{1i}|x_{1i}, x_{2i}, \alpha_i) = \alpha_i + \beta_0 + \beta_B \bar{x}_i + \beta_W(x_{1i} - \bar{x}_i),$$

where $\bar{x}_i = (x_{1i} + x_{2i})/2$. In this case, it is possible to distinguish between pair-specific effects and the effects of covariates for individuals in different pairs. Such models may be useful in observational studies with pairing or clustering.

11.2.2 Maximum Likelihood and Pseudolikelihood Methods

11.2.2.1 Maximum Likelihood

Let $(T_{1i}, \ldots, T_{ki})$ have continuous survivor function (s.f.) $S(t_1, \ldots, t_k)$ as in (11.2.1). Suppose the lifetimes of a random sample of individuals $i = 1, \ldots, n$ are subject to right censoring, so that potential censoring times C_{ji} $(j = 1, \ldots, k)$ may be associated with the T_{ji}. We will assume that $(C_{1i}, \ldots, C_{ki})$ is independent of $(T_{1i}, \ldots, T_{ki})$. Note that $C_{1i}, \ldots, C_{ki}$ are allowed to be related; in cases where all of the lifetimes in a cluster are subject to the same censoring time, they would be equal. Observed data for cluster i include times $(t_{1i}, \ldots, t_{ki})$, where $t_{ji} = \min(T_{ji}, C_{ji})$, and censoring indicators $(\delta_{1i}, \ldots, \delta_{ki})$, where $\delta_{ji} = I(T_{ji} = t_{ji})$.

To keep the notation and discussion as simple as possible, the following will deal with bivariate lifetimes (T_{1i}, T_{2i}); extensions to larger cluster sizes are straightforward. An observation is one of four types, since T_{1i} and T_{2i} can each be either censored or not. The likelihood function takes the form

$$L = \prod_{i=1}^{n} f(t_{1i}, t_{2i})^{\delta_{1i}\delta_{2i}} \left[\frac{-\partial S(t_{1i}, t_{2i})}{\partial t_{1i}}\right]^{\delta_{1i}(1-\delta_{2i})} \left[\frac{-\partial S(t_{1i}, t_{2i})}{\partial t_{2i}}\right]^{(1-\delta_{1i})\delta_{2i}} \cdot S(t_{1i}, t_{2i})^{(1-\delta_{1i})(1-\delta_{2i})}, \tag{11.2.14}$$

where $f(t_1, t_2)$ is the p.d.f. for (T_1, T_2). For models involving fixed covariates, $S(t_{1i}, t_{2i})$ in (11.2.14) is replaced by

$$S(t_{1i}, t_{2i}|\mathbf{x}_i) = Pr(T_1 \geq t_{1i}, T_2 \geq t_{2i}|\mathbf{x}_i). \tag{11.2.15}$$

Maximum likelihood estimation for fully parametric models can be implemented using general-purpose optimization software (Appendix D).

Nonparametric estimation of a joint survivor function $S(t_1, t_2)$ from censored data is a complex issue. When $S(t_1, t_2)$ is continuous, a unique m.l.e. does not exist. A variety of procedures have been suggested in the literature; many exhibit occasional peculiar behavior (e.g., nonmonotonicity), and asymptotics and variance estimation

are difficult. Van der Laan (1996), Hougaard (2000, Ch. 14), and Oakes (2001, Sec. 7.2) provide surveys of the area.

The portrayal of discrete bivariate distributions is also more awkward than for univariate distributions, and the descriptive value of univariate nonparametric estimates $\hat{S}(t)$ is pretty much lost for bivariate estimates $\hat{S}(t_1, t_2)$. However, nonparametric estimates are still desirable for joint and conditional probabilities, and they allow checks of parametric assumptions. A practical approach that is fairly easily implemented is to discretize the time scales for T_1 and T_2 and consider a joint probability function

$$f(t_1, t_2) = Pr(T_1 = t_1, T_2 = t_2)$$

for (t_1, t_2) on some finite grid. Terms in the likelihood function (11.2.14) are expressible in terms of $f(t_1, t_2)$; for example,

$$Pr(T_{1i} = t_{1i}, T_{2i} > t_{2i}) = \sum_{s > t_{2i}} f(t_{1i}, s)$$

is the relevant term for an individual with $\delta_{1i} = 1, \delta_{2i} = 0$. The likelihood L can be maximized numerically subject to $\sum\sum f(t_1, t_2) = 1$ to give $\hat{f}(t_1, t_2)$.

To check parametric models $S(t_1, t_2; \boldsymbol{\theta})$ one can first compare marginal distributions $S_1(t_1; \hat{\boldsymbol{\theta}})$ and $S_2(t_2; \hat{\boldsymbol{\theta}})$ with the corresponding univariate Kaplan–Meier estimates. Estimates for joint or conditional probabilities such as $Pr(T_1 \leq a, T_2 \leq b)$ or $Pr[T_1 \leq a | T_2 \in (b_1, b_2)]$ can also be compared. The main difficulty is for conditional probabilities in which the event conditioned on has low probability; in that case, nonparametric estimates will be imprecise unless there is a lot of data.

As discussed in Section 11.2.1, semiparametric regression models can also be considered; Section 11.2.3 provides an example. Semiparametric maximum likelihood as described in Section 7.4 is more difficult with multivariate marginal models. However, regression models for which the baseline hazard functions are piecewise constant are fairly easy to handle, and standard parametric results apply. As the number of pieces increases, something similar to semiparametric estimation results. In practice, estimates of regression coefficients and cumulative hazard or survivor functions usually change little as the number of pieces increases above five or six. Some semiparametric conditional PH models (11.2.9) based on random effects can be handled by existing software; we discuss this in Section 11.2.3 and the Computational Notes.

The main objective in many applications is to assess the effects of covariates on the marginal distributions for $T_1, \ldots, T_k$. One approach is to adopt a full multivariate model. A simpler alternative is to specify only the marginal distributions for the T_j and to use estimating functions that assume the T_j are independent; this is often called the working independence model approach. Because the T_j are not usually independent, it is necessary to develop variance estimation procedures that recognize this. We consider this approach next; it is less efficient than one based on a correct full model, but it is more robust and can be implemented using univariate lifetime methodology.

11.2.2.2 Pseudolikelihood Methods Based on Working Independence Models

Consider clusters that involve k lifetimes $T_1, \ldots, T_k$ and a vector of covariates $\mathbf{x}$. Suppose that no joint distribution $S(t_1, \ldots, t_k|\mathbf{x})$ is specified, but that marginal distributions $S_j(t_j|\mathbf{x})$ are. Let $S_j(t_j|\mathbf{x})$ depend on parameters $\boldsymbol{\theta}_j$ and let $\boldsymbol{\theta}$ denote all of the parameters used, with redundancies removed. It is not assumed that the $\boldsymbol{\theta}_j$ are functionally unrelated; in applications where the T_j have identical marginal distributions we would in fact have $\boldsymbol{\theta}_1 = \cdots = \boldsymbol{\theta}_k$. Let us consider the bivariate case further.

If T_1 and T_2 were independent, then $S(t_1, t_2) = S_1(t_1)S_2(t_2)$ and the likelihood function (11.2.14) based on data (t_{1i}, t_{2i}), $(\delta_{1i}, \delta_{2i})$, $\mathbf{x}_{1i}$, $\mathbf{x}_{2i}$ for independent clusters $i = 1, \ldots, n$ factors into

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_1(t_{1i}|\mathbf{x}_{1i})^{\delta_{1i}} S_1(t_{1i}|\mathbf{x}_{1i})^{1-\delta_{1i}} \prod_{i=1}^{n} f_2(t_{2i}|\mathbf{x}_{2i})^{\delta_{2i}} S_2(t_{2i}|\mathbf{x}_{2i})^{1-\delta_{2i}},$$

where $f_1(t_1|\mathbf{x})$ and $f_2(t_2|\mathbf{x})$ are the marginal p.d.f.'s. The estimating function $\mathbf{U}(\boldsymbol{\theta}) = \partial \log L/\partial \boldsymbol{\theta}$ is then

$$\mathbf{U}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j=1}^{2} \delta_{ji} \frac{\partial \log f_j(t_{ji}|\mathbf{x}_{ji})}{\partial \boldsymbol{\theta}} + (1 - \delta_{ji}) \frac{\partial \log S_j(t_{ji}|\mathbf{x}_{ji})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \mathbf{U}_i(\boldsymbol{\theta}). \tag{11.2.16}$$

If the marginal models are correctly specified, this is an unbiased estimating function, since the expectation of each of the $2n$ terms in the sum equals zero even when T_{1i} and T_{2i} are not independent. Consequently, under mild regularity conditions on the marginal models, the estimate $\hat{\boldsymbol{\theta}}$ obtained by solving $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$ is consistent and asymptotically normal, although it is not an m.l.e. unless T_{1i} and T_{2i} are independent. As described in Appendix C, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ has asymptotic covariance matrix $V(\boldsymbol{\theta})$ of "sandwich" form C(29),

$$V(\boldsymbol{\theta}) = A(\boldsymbol{\theta})^{-1} B(\boldsymbol{\theta}) A(\boldsymbol{\theta})^{-1}, \tag{11.2.17}$$

where

$$A(\boldsymbol{\theta}) = \frac{1}{n} E\left[\frac{-\mathbf{U}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right], \qquad B(\boldsymbol{\theta}) = \frac{1}{n} \mathrm{Var}[\mathbf{U}(\boldsymbol{\theta})].$$

A consistent estimate of $V(\boldsymbol{\theta})$ is obtained by replacing $A(\boldsymbol{\theta})$ and $B(\boldsymbol{\theta})$ by

$$\hat{A} = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \mathbf{U}_i(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}'}, \qquad \hat{B} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{U}_i(\hat{\boldsymbol{\theta}}) \mathbf{U}_i(\hat{\boldsymbol{\theta}})'. \tag{11.2.18}$$

In the case where $S_j(t|\mathbf{x})$ depends on $\boldsymbol{\theta}_j$ $(j = 1, 2)$ with $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ functionally independent, we can utilize existing software for univariate lifetime models. To see this, note that with $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, the matrices $A(\boldsymbol{\theta})$ and $\hat{A}$ are then block diagonal,

and

$$A(\boldsymbol{\theta})^{-1} = \begin{pmatrix} A_1(\boldsymbol{\theta}_1)^{-1} & 0 \\ 0 & A_2(\boldsymbol{\theta}_2)^{-1} \end{pmatrix}, \qquad \hat{A} = \begin{pmatrix} \hat{A}_1^{-1} & 0 \\ 0 & \hat{A}_2^{-1} \end{pmatrix},$$

where $\hat{A}_1$ and $\hat{A}_2$ are the observed information matrices from the univariate models for T_1 and T_2. Therefore, $\hat{A}_1^{-1}$ and $\hat{A}_2^{-1}$ are the standard covariance matrices for $\sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1)$ and $\sqrt{n}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2)$ from separate univariate analyses. The matrix $\hat{B}$ is also needed; it takes the form

$$\hat{B} = \frac{1}{n} \begin{pmatrix} \sum_{i=1}^{n} \hat{\mathbf{U}}_{1i} \hat{\mathbf{U}}_{1i}' & \sum_{i=1}^{n} \hat{\mathbf{U}}_{1i} \hat{\mathbf{U}}_{2i}' \\ \sum_{i=1}^{n} \hat{\mathbf{U}}_{2i} \hat{\mathbf{U}}_{1i}' & \sum_{i=1}^{n} \hat{\mathbf{U}}_{2i} \hat{\mathbf{U}}_{2i}' \end{pmatrix}, \tag{11.2.19}$$

where $\mathbf{U}_i(\boldsymbol{\theta}) = (\mathbf{U}_{1i}(\boldsymbol{\theta}_1)', \mathbf{U}_{2i}(\boldsymbol{\theta}_2)')'$ and the notation $\hat{\mathbf{U}}_{1i} = \mathbf{U}_{1i}(\hat{\boldsymbol{\theta}}_1)$, $\hat{\mathbf{U}}_{2i} = \mathbf{U}_{2i}(\hat{\boldsymbol{\theta}}_2)$ has been used. Some software provides these values as score residual vectors; otherwise, they are easily calculated. Models where $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ can also be handled in a similar way.

These methods allow univariate lifetime methodology to be applied to bivariate settings in which the marginal distributions are of primary interest. The extension to deal with models of dimension $k > 2$ is straightforward. Semiparametric models can also be considered. Rank-based procedures for semiparametric accelerated failure time models, which were discussed in Chapter 8, are fairly easy to adopt when the estimating functions for regression parameters can be written as a sum over $j = 1, \ldots, k$. Methodology for PH models is a little more complicated, particularly when the baseline hazard functions or covariate vectors for the different T_j are the same. Some software packages for the Cox PH model do, however, provide variance estimates for $\hat{\boldsymbol{\beta}}$ in the models for which the marginal survivor functions

$$S_j(t|\mathbf{x}) = \exp[-\Lambda_0(t)e^{\boldsymbol{\beta}'\mathbf{x}}] \qquad j = 1, \ldots, k \tag{11.2.20}$$

are the same; see the Computational Notes at the end of the chapter.

An alternative approach for PH models is to use piecewise-constant hazard functions, as in Section 7.4. In this case, the problem is rendered parametric and can be treated as described earlier. Estimation of baseline c.h.f.'s and of marginal survival probabilities $S_j(t|\mathbf{x})$ is also easily handled with this approach.

Finally, situations sometimes arise in which there are no covariates, and the marginal survivor functions $S_1(t), \ldots, S_k(t)$ are assumed identical. In that case it is appropriate to estimate the common survivor function $S(t)$ using the standard Kaplan–Meier estimate based on all the observations (t_{ji}, δ_{ji}), $j = 1, \ldots, k$; $i = 1, \ldots, n$. However, the standard formula (3.2.3) is not an appropriate variance estimate. Problem 11.10 shows that under some mild conditions a variance estimate

that allows for within-cluster association is

$$\hat{S}(t)^2 \sum_{i=1}^{n} \left\{ \sum_{t_\ell^* < t} \frac{d_i(t_\ell^*) - Y_i(t_\ell^*) d(t_\ell^*)/Y(t_\ell^*)}{Y(t_\ell^*) - d(t_\ell^*)} \right\}^2, \qquad (11.2.21)$$

where $t_1^*, \ldots, t_m^*$ are the distinct times at which deaths occur, $d(t_\ell^*)$ and $Y(t_\ell^*)$ are the total numbers of deaths and individuals at risk at t_ℓ^*, and $d_i(t_\ell^*)$, $Y_i(t_\ell^*)$ are the numbers of those who are in cluster i.

11.2.3 An Example

Huster et al. (1989) and many other authors have discussed data from the Diabetic Retinopathy Study (DRS). Diabetic retinopathy is a condition that arises in persons with diabetes; it involves abnormalities in the retina of the eye and is a leading cause of vision loss and blindness. In the DRS a large group of patients with diabetes and meeting certain eligibility criteria was followed for an extended period. For each patient, one eye was treated with laser photocoagulation, and the other eye was an untreated control. The response variables are the times T_j $(j = 1, 2)$ until the first occurrence of visual acuity less than 5/200 in each eye of a patient. The primary objective of the DRS was to assess the effectiveness of the laser treatment in delaying loss of visual acuity. The data set discussed here consists of results for a subset of $n = 197$ subjects who represent a 50% sample of high-risk patients in the study. Huster et al. (1989) give further discussion of the study, including the fact that the times T_j to loss of visual acuity are measured with error. As have other authors, we ignore this measurement error in the discussion that follows.

The data (see Appendix G) consist of times or censoring times (t_{1i}, t_{2i}) to the loss of visual acuity in each eye, the treatment indicator $x_{1ji} = 1$ (eye j is treated) for $j = 1, 2$, and the additional covariate $x_{21i} = x_{22i} = 1$ (diabetes is adult-onset), which indicates whether a person's diabetes was adult-onset or juvenile-onset. Of the 197 patients, 83 had adult-onset and 114 had juvenile-onset diabetes. The covariate x_1 thus takes on different values 0 and 1 for the two eyes of a given subject, whereas x_2 is the same for both eyes. The possibility of an interaction between treatment and type of diabetes will be considered through a covariate x_1x_2; its values are also different for the two eyes of a subject.

Let us first consider marginal or population-average effects for the covariates. Proportional hazards models will be considered in which the marginal hazard functions for T_{1i} and T_{2i} are of the form

$$h_{ji}(t|\mathbf{x}_{ji}) = h_0(t) \exp(\beta_1 x_{1ji} + \beta_2 x_{2ji} + \beta_3 x_{1ji} x_{2ji}), \qquad (11.2.22)$$

where $\mathbf{x}_{ji} = (x_{1ji}, x_{2ji})$ for $j = 1, 2$. The failure time distributions for the two eyes of a subject or the eyes of any two subjects have the same baseline hazard function and differ only in as much as their covariate values differ. Diagnostic checks described in Chapters 6 and 7 show the models considered here to be consistent with the data.

Table 11.2. Covariate Effects and SEs for Marginal PH Models

Model	Regression Coefficients		
	trt (*se*)	type (*se*)	trt*type (*se*)
Cox PH (independence)	−.43 (.22)	.34 (.20)	−.85 (.35)
Cox PH (robust)	−.43 (.19)	.34 (.20)	−.85 (.30)
Weibull PH (independence)	−.43 (.22)	.36 (.20)	−.87 (.35)
Weibull PH (robust)	−.43 (.19)	.36 (.20)	−.87 (.31)
Weibull BV Clayton	−.43 (.22)	.37 (.20)	−.84 (.35)
PC BV Clayton	−.42 (.19)	.36 (.22)	−.82 (.32)

Table 11.2 shows fits of three models:

1. A semiparametric PH model fitted with a working independence assumption using S-Plus function coxph; the output includes both proper robust variance estimates and variance estimates based on the assumption that T_{1i} and T_{2i} are independent, given $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$. The latter are incorrect since there is evidence of association between T_{1i} and T_{2i}, but are shown for the sake of comparison.
2. A parametric PH model for which the baseline hazard function $h_0(t) = \lambda\alpha t^{\alpha-1}$ is of Weibull form (Huster et al. 1989); pseudolikelihood estimates are shown for the independence working model approach, with independent variance estimates plus ones obtained from (11.2.17) and, for comparison, for a bivariate (BV) Clayton model (11.2.6) fitted by maximum likelihood.
3. A parametric BV Clayton model (11.2.17), in which $h_0(t)$ is piecewise-constant; six pieces, with cut points at $t = 5.43$, 10.57, 17.88, 27.60, and 43.5 were used (He 2001).

The results for the six approaches shown in Table 11.2 are in close agreement. Although the BV Clayton models fitted indicate nonindependence for T_{1i} and T_{2i}, with $\hat{\phi} = 1.01(se = .34)$ for the Weibull model, the robust and independence standard errors (SE's) for the regression coefficient estimates are not very different. As one would expect when there is positive association between T_{1i} and T_{2i}, the independence standard errors are (slightly) larger than the robust SE's for covariates that vary within individuals (i.e., treatment (trt) and trt*type). However, the BV Weibull–Clayton model actually gives SE's here that are closer to the independence SE's.

These results indicate significant effects due to treatment and type of diabetes. A significant interaction is also indicated, with adult-onset diabetics benefitting more from the treatment than juvenile-onset diabetics. The Cox PH model relative risks (hazard ratios) for treated to untreated adults is about $\exp(-.43 - .85) = .28$, and for treated to untreated juvenile cases is about $\exp(-.43) = .65$.

These regression coefficients and relative risks are marginal or population average values. That is, they give estimated relative effects for two randomly selected individuals in the population. This is a reasonable way to quantify the effect of type of diabetes; any individual has one type or the other, and the covariate effect describes

Table 11.3. Covariate Effects and SEs for Conditional PH Models

	Regression Coefficients		
Model	trt (*se*)	type (*se*)	trt*type (*se*)
1. Clayton semiparametric	−.51 (.22)	.40 (.26)	−.99 (.36)
2. Clayton PC hazards	−.52 (.23)	.40 (.27)	−.99 (.38)
3. Stratified Cox PH	−.56 (.26)	—	−.99 (.43)

the hazard ratio for two persons whose eyes are treated the same way but who have different types. For the effect of treatment, however, the paired study design allows a comparison that is individual-specific. That is, the difference in treating versus not treating the eyes of a given individual may be assessed. The marginal models just discussed do not do this; they compare the effects of the treatments for two randomly selected individuals.

We can consider individual-specific effects by fitting conditional PH models, as discussed in Example 11.2.1. Table 11.3 shows the results of fitting three models:

1. A Clayton semiparametric model (11.2.10), in which

$$\Lambda_j(t) = \Lambda_0(t)e^{\beta' \mathbf{x}_j}, \qquad j = 1, 2,$$

   and $\Lambda_0(t)$ is left unspecified; this can be fitted using S-Plus software (see the Computational Notes) for the Cox PH model using the gamma frailty option.
2. A Clayton model (11.2.10) with the same form for $\Lambda_j(t)$ as in (1), except that $\Lambda_0(t)$ is piecewise linear; this model can be fitted by ordinary parametric maximum likelihood (He 2001).
3. A Cox PH model in which $\Lambda_j(t)$ is as in (1), but where the stratified Cox partial likelihood of Section 7.1.6 and Example 7.2 is used for estimation. In this case, only the effects for covariates that vary within individuals are estimated, so the type effect is absent in Table 11.3.

These models all arise from a model for which there is an unobservable individual-specific effect α_i such that the hazard function for the jth eye ($j = 1, 2$) of individual i is of the form

$$\lambda(t|\mathbf{x}_i, \alpha_i) = \lambda_0(t)\alpha_i e^{\beta' \mathbf{x}_{ji}}. \tag{11.2.23}$$

Thus, for example, $\exp(\beta_1 + \beta_3 x_2)$ represents the hazard ratio for the treated versus untreated eye of an individual with $x_2 = 1$ (diabetes is adult-onset); the incorporation of the treatment-type interaction allows this effect to be different for individuals with the two types of diabetes.

Table 11.3 shows that the three approaches give similar results, particularly for the relative risks for treated and untreated eyes for adult-onset and juvenile-onset diabetes. The estimated treatment effects are further from zero than those for the marginal models, as expected. This reflects the fact that the conditional and marginal specifications are different models, as discussed in Example 11.2.1. Technically, both cannot be correct specifications of the process in question. However, all models are approximations to reality, and model checks of the conditional models (11.2.23) show that, like the marginal models, they describe the data quite well.

A final point is that the estimates of ϕ from the Clayton models (1) and (2) would be expected to be similar to those from the marginal models (11.2.22). This is the case, with $\hat{\phi} = 1.02(se = 0.33)$ from approach (2) and $\hat{\phi} = .93$ from approach (1).

11.3 SEQUENCES OF LIFETIMES

11.3.1 Some Models and Methods

Many event history problems can be formulated in terms of a series or sequence of lifetime variables $T_1, T_2, \ldots$ that represent the times between a specified series of events, or the lengths of the sojourns in a specified sequence of states for an individual; Examples 11.1.2 to 11.1.5 describe settings of this type. Let us consider such sequences T_j ($j = 1, 2, \ldots$), where $T_j \geq 0$. For a given $j = 1, 2, \ldots$, the variable T_j represents the same phenomenon for all individuals. In addition, the phenomena occur sequentially in real time; that is, T_j cannot be observed unless T_{j-1} has already been observed ($j = 2, 3, \ldots$).

Quite generally, models for $T_{1i}, T_{2i}, \ldots$ for an individual i, given a vector of fixed covariates $\mathbf{x}_i$, can be formulated as a sequence of conditional distributions

$$F_j(t|\mathbf{x}_i, t_i^{(j-1)}) = Pr(T_{ji} \leq t|\mathbf{x}_i, t_i^{(j-1)}), \tag{11.3.1}$$

where $t_i^{(j-1)} = (t_{1i}, \ldots, t_{j-1,i})$ and $t_i^{(0)}$ is null. If covariates are time-varying, it is convenient to specify models in terms of hazard functions,

$$\lambda_j(t|\mathbf{x}_{ji}(t), t_i^{(j-1)}) = \lim_{\Delta t \to 0} \frac{Pr(T_{ji} < t + \Delta t | T_{ji} \geq t, \mathbf{x}_{ji}(t), t_i^{(j-1)})}{\Delta t}, \tag{11.3.2}$$

where $\mathbf{x}_{ji}(t)$ represents the values of the covariates after time t has elapsed since the occurrence of $T_{j-1,i}$. The analysis of sequences of times $T_{1i}, T_{2i}, \ldots$ under such models can be based on a direct application of the univariate lifetime methodology of previous chapters. The main challenges are frequently in deciding how to represent the effects of $t_i^{(j-1)}$ in (11.3.1) or (11.3.2), and in how to interpret the effects of covariates across a whole sequence of times. Example 11.3.1, which follows, illustrates these points. In general, successive times $T_{1i}, T_{2i}, \ldots$ will not be independent (given covariate values), but if they are then modeling, analysis, and interpretation of effects is more straightforward.

Sometimes the marginal distributions of T_{ji} given $\mathbf{x}_i$ are of interest for specific values $j \geq 2$. Conventional regression models for T_{ji} given $\mathbf{x}_i$ and $t_i^{(j-1)}$ as in (11.3.1) and (11.3.2) generally give complicated forms for the marginal distributions

of T_{ji}, given $\mathbf{x}_i$ for $j \geq 2$, and of the corresponding covariate effects. It is possible to use multivariate models like those in Section 11.2.1 for sequences of times, but it is not in general possible to employ the simple pseudolikelihood methods, based on independence models, which were described in Section 11.2.2. The reason is that if the times T_{ji} $(j = 1, 2, \ldots)$ are not mutually independent given the covariates, then in most applications the potential censoring time for T_{ji} depends on $t_i^{(j-1)}$, and is therefore not independent of T_{ji}. For example, when individual i is followed for a specified length of time, C_i, then if $T_{1i} = t_{1i}$ is observed, the potential censoring time for T_{2i} is $C_i - t_{1i}$. A case where separate marginal analysis of T_{2i} would be permissible is where the observation scheme is such that both T_{1i} and T_{2i} are always fully observed.

The one family of lifetime models that gives simple representations for covariates in both conditional and marginal distributions is the log-normal family. In this case, $Y_{ji} = \log T_{ji}$ and the conditional distributions (11.3.1) are taken to be of the form

$$Y_{ji} = \boldsymbol{\beta}_j' \mathbf{x}_i + \boldsymbol{\gamma}_j' \mathbf{z}_{ji} + \epsilon_{ji}, \qquad j = 1, 2, \ldots, \tag{11.3.3}$$

where $\mathbf{z}_{ji} = (y_{1i}, \ldots, y_{j-1,i})'$, $\boldsymbol{\beta}_j$ and $\boldsymbol{\gamma}_j$ are vectors of regression coefficients, and $\epsilon_{ji} \sim N(0, \sigma_j^2)$ are independent for $j = 1, 2, \ldots$ and for $i = 1, \ldots, n$. As is well known, the marginal distribution of Y_{ji} given $\mathbf{x}_i$ for any $j = 1, 2, \ldots$ under (11.3.3) is univariate normal with a mean that is a linear function of $\mathbf{x}_i$, and the joint distribution of any set of Y_{ji} is multivariate normal. For other models, simple conditional specifications (11.3.1) or (11.3.2) generally give complex marginal specifications for $T_{2i}, T_{3i}, \ldots$; Section 11.3.2 gives an example.

The likelihood function for typical data is easily written down in terms of conditional distributions, as in (11.3.1) or (11.3.2). In general, the number of times T_{ji} that are observed will depend on the length of follow-up C_i for that individual, and can be different for two individuals who are followed for the same length of time. If observation of individual i starts at calendar time 0, which corresponds to the value $T_{1i} = 0$, and continues until calendar time C_i, then the likelihood contribution for the individual is of the form

$$\left\{ \prod_{j=1}^{k_i - 1} f(t_{ji} | \mathbf{x}_i, t_i^{(j-1)}) \right\} \left\{ f(t_{k_i i} | \mathbf{x}_i, t_i^{(k_i - 1)})^{\delta_{k_i i}} S(t_{k_i i} | \mathbf{x}_i, t_i^{(k_i - 1)})^{1 - \delta_{k_i i}} \right\}, \tag{11.3.4}$$

where k_i is the number of random variables, T_{ji}, that are observed, and $\delta_{k_i i} = I(T_{k_i i} = t_{k_i i})$ indicates whether the k_ith lifetime is observed or censored. For (11.3.4) to be valid it is sufficient that censoring be independent in a sense equivalent to that in Section 2.2.2. That is, the censoring process may depend on prior lifetimes or event history, but it cannot be anticipatory.

In some settings the initial lifetimes T_{1i} may also be subject to left truncation, as discussed in Section 2.4.1. As long as the left-truncation mechanism is independent in the sense discussed there, the same likelihood as in (11.3.4) applies, except with

$f(t_{1i} | \mathbf{x}_i)$ replaced by (assuming $k_i > 1$)

$$f(t_{1i} | \mathbf{x}_i) / S(u_i | \mathbf{x}_i),$$

where u_i is the left-truncation time for individual i. An excellent discussion of the requirements on the beginning and end of the observation period when studying sequences of events is given by Aalen and Husebye (1991).

It is usually of interest whether times $T_{1i}, T_{2i}, \ldots$ are independent, given the covariate values. The simplest way to approach this question is through conditional models (11.2.1) or (11.2.2), in which the dependence of T_{ji} on $t_i^{(j-1)}$ can be assessed. If the T_{ji} $(j = 1, 2, \ldots)$ are independent, given covariate values, then marginal interpretations and analysis are much simpler. Problem 11.12 considers the case of Kaplan–Meier estimation for marginal distributions in this setting.

11.3.2 An Example

Data on the occurrence of pulmonary exacerbations in persons with cystic fibrosis were introduced in Example 1.1.8 and have been discussed in several other examples. The time T_1 to a first exacerbation in subjects who were randomized to either the treatment rhDNase or to a placebo was considered in Examples 6.3.4 and 7.2.1. Both treatment and the covariate forced expiratory volume (fev), measured at the start of study, were important factors, with the active treatment (rhDNase) and higher fev measures associated with longer times to exacerbation.

The 645 subjects in the study were followed for 169 days and could experience more than one exacerbation. In particular, when a subject experienced an exacerbation they were treated with antibiotics, and once the exacerbation had cleared up the person was at risk of a new exacerbation. The process can be portrayed as in Figure 11.1, with a person reverting to the Normal (nonexacerbation) state after antibiotic treatment. A more extensive analysis of the study data might therefore take account of this process, and consider the durations of the successive periods of time spent in the Normal and Exacerbation states by a subject. Since the durations of the sojourns in the Exacerbation state (i.e., during antibiotic treatment) were roughly the same for most subjects, we will consider just the durations of the subsequent exacerbation-free periods $T_1, T_2, \ldots$.

A total of 139 out of 324 Placebo subjects and 104 out of 321 Treatment subjects experienced a first exacerbation, so had uncensored values for T_1. Assuming their antibiotic treatment and recovery from the exacerbation occurred before the end of the follow-up period, these subjects would then start a second Normal, or exacerbation-free, period; for some the time T_2 to the second exacerbation would be observed and for some it would be censored. Those who experienced a second exacerbation could then become at risk for a third one, and so on. Table 11.4 shows the total numbers of exacerbations across the study group.

Let us consider analyses of the successive times $T_1, T_2, \ldots$, given the covariates $x_1 = I$ (Treatment = rhDNase) and $x_2 = fev$ (centered baseline fev) considered in Examples 6.3.4 and 7.2.1. Those examples indicated that both accelerated failure

Table 11.4. Numbers of Pulmonary Exacerbations for Patients

Number of Exacerbations	Number of Patients	
	Placebo Group	Treatment Group
0	185	217
1	97	65
2	24	30
3	13	6
4	4	3
5	1	0
	324	321

time (AFT) and PH models provide reasonable descriptions of the data for T_1. We will consider both approaches.

A PH approach for second and later exacerbation times $T_2, T_3, \ldots$ could use a PH specification for the conditional hazard functions (11.3.2) for T_j given $t_1, \ldots, t_{j-1}$ and $\mathbf{x} = (x_1, x_2)'$. Here we consider such models with

$$\lambda_j(t|\mathbf{x}_i, t_i^{(j-1)}) = \lambda_{0j}(t)\exp(\boldsymbol{\beta}_j'\mathbf{x}_i + \gamma_j s_i^{(j-1)}) \tag{11.3.5}$$

for $j = 1, 2, \ldots$, where $s_i^{(j)} = t_{1i} + \cdots + t_{ji}$, with $s_i^{(0)} = 0$. These models are suggested by preliminary analysis and are supported by diagnostic checks described in Chapter 7. It is assumed that the durations of the periods during which subjects are in the exacerbation state are independent of the T_j. This appears appropriate, but if it were not true, then it would be necessary to include terms in (11.3.4) for previous exacerbation durations, in order to render the censoring process independent.

An AFT approach would be to assume conditional specifications (11.3.1) based on one of the parametric models in Chapter 6. A log-normal model for T_1 was found to be satisfactory in Example 6.3.4, so one might consider models of the form (11.3.3) for subsequent times; note that in this case the models include an intercept term, so covariate $\mathbf{x}_i$ in (11.3.3) is $(1, x_{1i}, x_{2i})'$ and $\boldsymbol{\beta}_j = (\beta_{0j}, \beta_{1j}, \beta_{2j})'$.

There is a substantial number of observed duration times only for T_1 and T_2, so we will focus on them. Table 11.5 shows the estimated regression coefficients and standard errors for the Cox PH models for T_1 and T_2, and the estimated regression coefficients and scale parameters $\hat{\sigma}_j$ $(j = 1, 2)$ for the log-normal models. To fit these models it is only necessary to carry out two univariate lifetime analyses using standard software. The results for T_1 are reproduced from the earlier analyses in Examples 7.2.1 and 6.3.4. The results for T_2 are interesting: whereas the effects of rhDNase treatment and high fev on T_1 (time to first exacerbation) were positive (associated with larger T_1), the m.l.e.'s for the regression coefficients for T_2 point in the opposite direction, though the effects are not statistically significant. This should not necessarily surprise us. A strong positive association between T_2 and T_1 is indicated by the estimated regression coefficients for t_1 and $\log t_1$ in the PH and AFT

Table 11.5. Fitted Log-Normal and PH Models for Exacerbation Times T_1 and T_2

Model		Cox PH	Log-Normal AFT
	Parameter	Estimate (SE)	Estimate (SE)
$T_1\|\mathbf{x}$	intercept	—	5.403 (.105)
	treatment	−.383 (.130)	.430 (.137)
	fev	−.0206 (.0028)	.0217 (.0029)
	scale	—	1.446 (.074)
$T_2\|\mathbf{x}, t_1$	intercept	—	3.209 (.487)
	treatment	.358 (.225)	−.227 (.211)
	fev	.0009 (.0054)	−.0045 (.0047)
	t_1 or $\log t_1$[a]	−.0143 (.0039)	.417 (.132)
	scale	—	1.228 (.105)

[a] t_1 for PH model and $\log t_1$ for AFT model.

models for T_2 given t_1 and $\mathbf{x}$. Because of the strong association between T_1 and $\mathbf{x}$, the values of $\mathbf{x}$ and t_1 in the analysis of T_2 are rather strongly correlated, so it is not possible to get an unambiguous picture of the conditional covariate effects on T_2. Examination of third exacerbations (T_3) shows a similar picture, though in that case the small number of observed exacerbations provides little information, and no significant effects even for t_1, t_2 are seen.

The question as to whether the effect of treatment (which is administered daily) tends to diminish over time is important, and the analysis of $T_2, T_3, \ldots$ given covariates and previous exacerbation times does not provide much insight. It is possible to consider the marginal distribution of T_2 (or T_3 or $T_4 \ldots$) given the covariates treatment and fev. These can be derived for either of the PH or AFT models represented in Table 11.5, but is particularly easy for the log-normal AFT model. Under the model (11.3.3) for $j = 1, 2$, the marginal distribution of Y_{2i} given $\mathbf{x}_i$ is normal,

$$Y_{2i} \sim N\left(\beta_0^{(2)} + \beta_1^{(2)}x_{1i} + \beta_2^{(2)}x_{2i}, \sigma_{(M)2}^2\right), \tag{11.3.6}$$

where, letting $\beta_r^{(1)}$ for $r = 0, 1, 2$ be the regression coefficients in the model (11.3.3) with $j = 1$,

$$\beta_r^{(2)} = \beta_{r2} + \beta_r^{(1)}\gamma_2, \qquad r = 0, 1, 2$$

and $\sigma_{(M)2}^2 = \gamma_2^2\sigma_1^2 + \sigma_2^2$. The m.l.e.'s and se's of the regression coefficients for treatment and fev are obtainable from Table 11.5 plus the asymptotic covariance matrices for those analyses. We find the estimate (and se) for treatment to be $\hat{\beta}_1^{(2)} = -.048(.215)$ and the estimate for fev to be $\hat{\beta}_2^{(2)} = .0045(.0053)$; neither covariate effect is significant.

A qualification concerning the marginal analysis just described is that the study involved follow-up of only 169 days, and under half of the subjects experienced even

a first pulmonary exacerbation. It may be a misleading extrapolation to consider the marginal distribution of Y_2 or T_2, which implicitly assumes that everyone eventually experiences a first exacerbation, and also to assume that the normal model (11.3.3) describes the upper half of the distribution for Y_1 or T_1. An alternative approach is to consider the distribution of Y_2 given that $Y_1 \leq \log L$, where L is a value, such as 169 days, that reflects the limited information about the distribution of Y_1. Similarly, we would consider probabilities

$$Pr(T_2 \leq t_2 | T_1 \leq L, \mathbf{x}) \tag{11.3.7}$$

for which $t_2 + L$ was not too much larger than 169 days.

Probabilities (11.3.7) can be calculated for either the AFT or PH model just considered. The log-normal model gives

$$Pr(T_2 \leq t_2 | T_1 \leq L, \mathbf{x}) = \frac{\int_{-\infty}^{\log L} G\left(\frac{y_2 - \beta_2'\mathbf{x} - \gamma_2 y_1}{\sigma_2}\right) \frac{1}{\sigma_1} g\left(\frac{y_1 - \beta_1'\mathbf{x}}{\sigma_1}\right) dy_1}{G\left(\frac{\log L - \beta_1'\mathbf{x}}{\sigma_1}\right)}, \tag{11.3.8}$$

where $\mathbf{x} = (1, x_1, x_2)'$, $\boldsymbol{\beta}_j = (\beta_{0j}, \beta_{1j}, \beta_{2j})'$ as in (11.3.3) and G and g are the standard normal distribution function and p.d.f., respectively. For the PH model (11.3.5), the corresponding probability is

$$Pr(T_2 \leq t_2 | T_1 \leq L, \mathbf{x}) = \frac{\int_0^L \{1 - \exp[-\Lambda_2(t_2|\mathbf{x}, t_1)]\} \lambda_1(t_1|\mathbf{x}) \exp\{-\Lambda_1(t_1|\mathbf{x})\}\, dt_1}{1 - \exp\{-\Lambda_1(L|\mathbf{x})\}}, \tag{11.3.9}$$

where $\Lambda_1(t|\mathbf{x})$ and $\Lambda_2(t|\mathbf{x}, t_1)$ are the c.h.f.'s corresponding to $\lambda_1(t|\mathbf{x})$ and $\lambda_2(t|\mathbf{x}, t_1)$ in (11.3.5).

Calculation of estimates of (11.3.8) and (11.3.9) based on the models represented in Table 11.5 show estimated probabilities for $L = 169$, $t_2 \leq 119$, to be slightly smaller for the placebo group than for the rhDNase group, but the differences are not statistically significant. This is consistent with the marginal log-normal model for Y_2 given $\mathbf{x}$, fitted previously.

More general models for the onset of exacerbation spells, based on event history modeling, provide alternative ways to examine whether the treatment effect is persistent. These types of models are discussed in the following section. Cook and Lawless (2002, Section 6) illustrate such an approach on the data in this section.

11.4 GENERAL EVENT HISTORY PROCESSES

As described in Section 11.1, many studies involve processes in which several types of events can occur over time for individuals in some population. Section 11.3 discussed how univariate lifetime analysis can be applied to sequences of events that occur in some fixed order. Univariate methodology can also be applied to more complex processes; how this is done will be described briefly.

Processes involving multiple events or states, as in Examples 11.1.2—11.1.5, can quite generally be represented in the following way. Suppose that k different types of events can occur for an individual over (continuous) time. For a specific individual at time t, let $\mathcal{H}(t)$ denote the history of events plus covariate information up to time $t-$, and for $s < t$ let $N_j(s, t)$ denote the number of times event type j occurs in the interval $[s, t)$. Then event intensity functions are defined as

$$\lambda_j[t|\mathcal{H}(t)] = \lim_{\Delta t \to 0} \frac{Pr[N_j(t, t + \Delta t) = 1 | \mathcal{H}(t)]}{\Delta t}, \qquad j = 1, \ldots, k, \tag{11.4.1}$$

and under the assumption that no two events of any kind can occur simultaneously, they specify the events process. This assumption is made throughout this section. The intensity functions (11.4.1) can be thought of as generalizations of hazard functions; the mode-specific hazard functions (9.1.1) used for multiple failure mode problems are a special case of (11.4.1).

The forms that the intensity functions will take in a specific setting depend on the process in question. Consider the following simple but important example.

Example 11.4.1. The Illness-Death Process. Figure 11.2 portrays a process in which individuals start in state 1, and may subsequently enter states 2 and/or 3. This is often referred to as the illness–death process, since it applies to settings where an individual starts off as "healthy" (state 1), may acquire a certain illness or disease (enter state 2), and may die (enter state 3) with or without having acquired the disease. In this case, we can identify three types of events with the three types of transitions between states, say as follows:

Event type 1: $1 \rightarrow 2$ transition
Event type 2: $1 \rightarrow 3$ transition
Event type 3: $2 \rightarrow 3$ transition.

If there were no covariates, then one might consider specifications for the intensity functions $\lambda_j[t|\mathcal{H}(t)]$ for $j = 1, 2$ and $t \geq 0$ as

$$\lambda_j[t|\mathcal{H}(t)] = \alpha_j(t) Y_1(t), \qquad j = 1, 2, \tag{11.4.2}$$

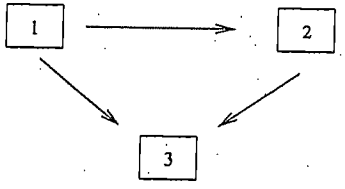

Figure 11.2. The illness–death process.

where $Y_1(t) = I$ (individual is in state 1 at time $t-$), and the $\alpha_j(t)$ are specified functions. The $Y_1(t)$ term is used in (11.4.2) because if the individual is not in state 1 just prior to time t, no transition out of state 1 can occur. Specification of $\lambda_3[t|\mathcal{H}(t)]$ would require consideration of whether it should depend on the time of entry to state 2 as well as on t. Models for which it depends only on t are called Markov models (e.g., Ross 1983).

In many applications the intensity functions for different event types are functionally independent, and do not have any parameters in common. Such situations give rise to likelihood functions that can be factored into pieces and treated using univariate lifetime methods, provided that event and censoring times are observed. To see this, suppose that an individual i is observed over the time interval $(0, C_i)$, and suppose that during this period the individual experiences m_i events, at times $t_{i1} < t_{i2} < \cdots < t_{im_i}$. For notational convenience, an event is considered to occur at C_i; this may be the event "end of follow-up" or one of the event types $1, \ldots, k$.

A key result in what follows is that for a process specified by intensities (11.4.1)

$$Pr[\text{no events in } (t, t+s)|\mathcal{H}(t)] = \exp\left\{-\int_t^{t+s} \sum_{j=1}^{k} \lambda_j[u|\mathcal{H}(u)]\,du\right\}. \quad (11.4.3)$$

This is a generalization of (9.1.2) for multiple failure modes and of (1.2.18) in Section 1.2.3. It is proved by noting that

$$Pr\left[\sum_{j=1}^{k} N_j(t, t+\Delta t) = 0|\mathcal{H}(t)\right] = 1 - \sum_{j=1}^{k} \lambda_j[t|\mathcal{H}(t)]\Delta t + o(\Delta t)$$

and using the product-integral introduced in Section 1.2.3 to evaluate (11.4.3) as

$$\prod_{[s,t)} Pr\{\text{no events in } [u, u+du)|\mathcal{H}(u)\}.$$

The likelihood contribution for individual i can now be found to be

$$\prod_{\ell=1}^{m_i} \exp\left\{-\int_{t_{i,\ell-1}}^{t_{i\ell}} \sum_{j=1}^{k} \lambda_j[t|\mathcal{H}_i(t)]\,dt\right\} \prod_{j=1}^{k} \lambda_j[t_{i\ell}|\mathcal{H}_i(t_{i\ell})]^{\delta_{i\ell j}}, \quad (11.4.4)$$

where $t_{i0} = 0$, $\delta_{i\ell j} = I$ (event of type j occurs at $t_{i\ell}$ for individual i), and it is understood that $0^0 = 1$. To obtain (11.4.4), note that the event history observed for individual i consists of gaps $(t_{i,\ell-1}, t_{i\ell})$ with no events that are "followed" by events at the times $t_{i\ell}$. This approach is completely analogous to that used to justify the likelihood function for right-censored univariate lifetime processes in Section 2.2.2.

The property that makes univariate lifetime data methods applicable is that (11.4.4) can be factored into separate pieces for each $j = 1, \ldots, k$:

$$L_{ij} = \prod_{\ell=1}^{m_i} \exp\left\{-\int_{t_{i,\ell-1}}^{t_{i\ell}} \lambda_j[t|\mathcal{H}_i(t)]\,dt\right\} \lambda_j[t_{i\ell}|\mathcal{H}_i(t_{i\ell})]^{\delta_{i\ell j}}. \quad (11.4.5)$$

Each term (11.4.5) takes the form of the likelihood for a univariate lifetime variable with hazard function $\lambda_j[t|\mathcal{H}_i(t)]$, subject to left truncation at $t_{i,\ell-1}$ and to possible right censoring at $t_{i\ell}$. The event time $t_{i\ell}$ in L_{ij} is equivalent to an observed lifetime of type j if $\delta_{i\ell j} = 1$ and to a censoring time if $\delta_{i\ell j} = 0$. The likelihood function obtained for multiple failure mode models in (9.1.11) and (9.1.13) of Chapter 9 is a special case of (11.4.5). As there, parametric univariate lifetime data analysis software can be applied directly to (11.4.5), provided that it accommodates left truncation as well as right censoring. Semiparametric models are in some cases a little more difficult to handle. However, in the case of PH or multiplicative hazards models, the justification used in Section 7.4 for semiparametric likelihood can be used to show that standard Cox partial likelihood methods can be quite generally applied to the separate event intensity functions for $j = 1, \ldots, k$.

Note that in (11.4.5), a specific event type j may not appear in some or all of the terms $L_{ij\ell}$; it will, in fact, appear only if individual i is at risk of a type j event over $(t_{i,\ell-1}, t_{i\ell}]$. The following example illustrates the likelihood calculation for a model corresponding to Figure 11.2.

Example 11.4.2. Consider the illness–death model of Example 11.4.1, in which three types of transitions, or events, are possible. Suppose that a Markov model is considered, for which the event intensity functions are

$$\lambda_j[t|\mathcal{H}(t)] = Y_j(t)\alpha_j(t), \qquad j = 1, 2, 3, \quad (11.4.6)$$

where $Y_1(t) = Y_2(t) = I$ (individual is in state 1 at time $t-$), $Y_3(t) = I$ (individual is in state 2 at time $t-$), and $\alpha_1(t)$, $\alpha_2(t)$, $\alpha_3(t)$ have separately specified parametric forms, with no covariates present. For simplicity, suppose there are prespecified follow-up times, C_i, that determine censoring.

Now, each of the event types 1, 2, 3 can occur at most once, and there can be at most two events in total, so $m_i = 1$ or 2. Let t_{i1} be the time of the first event, which is of either type 1, type 2, or end of follow-up. If it is of type 1, then $m_i = 2$ and the individual becomes at risk for a second event (of either type 3 or end of follow-up); otherwise, $m_i = 1$. It is then easily seen that under (11.4.6), the likelihood contributions L_{ij} for $j = 1, 2, 3$ from individual i are, respectively,

$$L_{i1} = \exp\left\{-\int_0^{t_{i1}} \alpha_1(t)\,dt\right\} \alpha_1(t_{i1})^{\delta_{i11}} \quad (11.4.7)$$

$$L_{i2} = \exp\left\{-\int_0^{t_{i1}} \alpha_2(t)\,dt\right\} \alpha_2(t_{i1})^{\delta_{i12}} \quad (11.4.8)$$

$$L_{i3} = \exp\left\{-\int_{t_{i1}}^{t_{i2}} \delta_{i11}\alpha_3(t)\,dt\right\} \alpha_3(t_{i2})^{\delta_{i23}\delta_{i11}}. \quad (11.4.9)$$

Note that the likelihood contributions (11.4.7) and (11.4.8) have the form for (possibly right-censored) lifetimes t_{i1}, and (11.4.9) has the form for a lifetime that is left truncated at t_{i1} and possibly right-censored at t_{i2}.

The following example considers a study that has been examined by many different authors. The rather brief treatment here is designed to illustrate the application of univariate methods to one specific model that fits the data satisfactorily.

Example 11.4.3. The Stanford Heart Transplant Study. The data set considered here concerns the survival of patients who were admitted into the Stanford University heart transplant program, from 1967 to 1973. This observational study has been considered by many authors, including Crowley and Hu (1977), Kalbfleisch and Prentice (1980, Sec. 5.5.3), Cox and Oakes (1984, Sec. 8.9), and Therneau and Grambsch (2000, Sec. 3.7.1). The data consist of information on each individual, their time of admission to the program, their time of transplant and factors associated with the transplant (if it occurred), and their time of death or end of follow-up. The data considered here are those used by Crowley and Hu (1977) and are available in electronic form (see Appendix G).

The following factors are those found most relevant, and attention is restricted to them:

$$x_1 = \text{age of patient at admission (in years)—48}$$
$$x_2 = I \text{ (previous heart surgery)}$$
$$x_3 = \text{year of admission to program}$$
$$x_4(t) = I \text{ (transplant occurred by time } t),$$

where $t \geq 0$ is time since admission to the program. One approach to analysis, discussed extensively in the references just cited, is to use multiplicative semiparametric regression models for the time T to death from admission. In this approach, the effect of the heart transplant on the hazard function for death is examined through the effect of the time-varying covariate $x_4(t)$, and possible interactions with other explanatory variables. Such an analysis can be carried out by using standard Cox model software and the methods discussed in Chapter 7. A clear picture does not emerge from such an analysis. The strongest effect indicated is a positive association between survival time and the year of admission to the program, suggesting perhaps that individuals admitted in earlier years were typically in poorer condition than those admitted later on. No other factors are strongly significant, though there is mild evidence of an interaction between year of admission (x_3) and transplant ($x_4(t)$), suggesting that a mild (but nonstatistically significant) benefit from transplant applied primarily to persons admitted in the earlier years.

An alternative approach is to consider the three-state framework portrayed in Figure 11.2, with

Table 11.6.

Transition Intensity	Covariate	Estimated Coefficient	se	Z
State 1 → State 2	x_1	.031	.014	2.20
	x_2	.047	.315	.14
	x_3	.002	.070	.03
State 1 → State 3	x_3	−.275	.106	−2.61
State 2 → State 3	x_1	.051	.021	2.43
	x_2	−.833	.447	−1.19

State 1 ≡ Admitted to program, alive, pretransplant

State 2 ≡ Alive, posttransplant

State 3 ≡ Dead.

A Markov model with event intensities $\lambda_1[t|\mathcal{H}(t)]$ for transplant, $\lambda_2[t|\mathcal{H}(t)]$ for death without transplant, and $\lambda_3[t|\mathcal{H}(t)]$ for death following transplant offers some useful insights. Let us consider the multiplicative models

$$\lambda_j[t|\mathcal{H}(t)] = Y_j(t)\lambda_{0j}(t)\exp(\boldsymbol{\beta}_j'\mathbf{x}), \qquad j = 1, 2, 3, \tag{11.4.10}$$

where $Y_j(t) = I$ (individual is at risk for event j at time t) and $\mathbf{x}$ involves only the fixed covariates x_1, x_2, x_3. It is unnecessary, and redundant, to consider the transplant indicator $x_4(t)$, because the models (11.4.10) allow the hazard function for death to be different pre- and posttransplant. That is, (11.4.10) is a little more general than the model discussed earlier; the former model is equivalent to assuming the baseline death intensities $\lambda_{02}(t)$ and $\lambda_{03}(t)$ in (11.4.10) are proportional.

This model can be fitted using software for the semiparametric Cox model, separately for each event type $j = 1, 2, 3$. For event types 1 and 2 (corresponding to transitions from States 1 to 2 and 1 to 3, respectively), there is no left truncation, and the events are like two competing modes of failure. For event Type 3 (a State 2 to 3 transition), the left truncation time is t_1, the time at which the transplant occurred. Table 11.6 shows some fitted models that describe the data well. The analysis shows that the year of admission (x_3) is significant only for Type 2 events, and that age is significant for Type 1 and Type 3 events. Prior therapy (x_2) is not significant anywhere. These results suggest that persons admitted to the program in earlier years were more likely to die before receiving a transplant and that younger persons are associated with a longer wait for transplant, but also a lower risk of death after transplant.

Parametric maximum likelihood analysis for quite general event history models is straightforward via (11.4.4) and (11.4.5). Semiparametric PH analysis as in the preceding example is also available for many situations, as noted prior to

Example 11.4.2. References provided in the Bibliographic Notes can be consulted for details and examples.

11.5 FAILURE AND RELATED PROCESSES

Time-varying processes are frequently of interest in problems involving failure or lifetimes. Situations in which external time-varying factors are treated as covariates have been discussed earlier in the book, most specifically in Sections 6.4.3 and 7.1.8. This addresses factors such as varying temperature or voltage stresses in a life test experiment on electronic parts, or the effect of air pollution on the occurrence of respiratory problems in humans. In many applications internal processes are also of interest. Measurements on the amount of deterioration in a piece of equipment or on biological markers in persons with ultimately fatal illnesses are of this type. Such processes are in many cases physically related to the failure process, and can be considered as responses, rather than as covariates. It is often important to understand and model such processes, both individually and jointly with failure. This leads to topics on stochastic processes that cannot be discussed in detail here. A brief overview of some important topics, and approaches to modeling, will be provided.

11.5.1 Some Context and Objectives

The following examples give some settings in which failure and other processes are of joint interest.

Example 11.5.1. Usage Measures and Patterns. In problems involving the reliability of equipment, the way in which the equipment is used is generally an important factor. This is often characterized through one or more usage measures $x(t)$. For example, with automobiles t might represent the age (total time in service) of the vehicle and $x_1(t)$ the distance driven by time t. Another usage factor would be the cumulative number of engine cycles $x_2(t)$ up to time t. The usage measure histories $\{\mathbf{x}(t), t \geq 0\}$ are random in populations of vehicles in service, and it is important to understand their relationship with specific types of failure (e.g., Lawless et al. 1995). In some cases it would be sensible to treat a usage measure as the primary time scale for analyzing failures. For example, the distance driven is often used for certain types of vehicle failures.

Example 11.5.2. Degradation Measures. Many failure processes are linked to some type of deterioration or degradation in the physical units for which failures occur. For example, Bogdanoff and Kozin (1985) discuss the growth of fatigue cracks in metal test specimens that are subjected to cyclic loading cycles. Crack size $Z(t)$ is measured as the length of the crack, where the time variable t is the number of test cycles. Failure time, T, is often defined in this context as the time (number of cycles), t, at which the crack size first exceeds some value, z_0. The degradation process $\{Z(t), t \geq 0\}$ thus defines failures, so by modeling this process we also

determine the failure time distribution. More importantly, in contexts where failures are rare or must be avoided, the identification and study of degradation processes can largely replace failure time considerations. For example, conditioned-based maintenance methodology uses degradation or other measures of the condition of equipment to decide when to overhaul or replace units before failures occur.

Whitmore et al. (1998) consider a more complex example concerning the production of aluminum by electrolysis in reduction cells that consist of a carbon anode in a carbon-lined steel box. The reduction cells degrade over time through physical distortion of the box and iron contamination of the aluminum resulting from cracks in the box's lining. Replacement and occasional catastrophic failures of the reduction cells is examined in terms of three degradation measures: $z_1(t)$ = percent iron contamination at time t, $z_2(t)$ = horizontal distortion (in inches) of the box at t, and $z_3(t)$ = displacement (drop) of the cathode (in inches) at t.

Example 11.5.3. Performance Measures. Sometimes there are measures that specify how well a unit or individual is functioning, or performing, at time t. For example, Tseng et al. (1995) considered the design of fluorescent lamps, whose performance at time t in service was measured by their luminosity $Z(t)$. This tends to decrease over time, and the lifetime for a lamp is often defined as the time t at which the luminosity falls below a certain value. Tseng et al. (1995) considered the effect of design and manufacturing factors on the luminosity curves $\{Z(t), t \geq 0\}$ for a certain type of lamp, and were thus able to draw inferences about the lifetimes of the lamps.

Example 11.5.4. Biological Markers. For many human illnesses or chronic conditions there are biological "markers" or measures that are associated with the course of the illness, and thus with lifetime variables related to the illness. For example, the disease pattern and time to death for persons infected with HIV is related to variations in CD4 T-lymphocyte blood-cell counts $Z_1(t)$ and to the HIV viral load or prevalence $Z_2(t)$ (e.g., Shi et al. 1996; Lee et al. 2000). Such markers are predictive of survival in the sense that the risk of death becomes high when the marker values fall into certain regions (e.g., when CD4 counts become very low), generally because they are associated with either the physical condition of the individual or the progression or virulence of the disease. However, because of factors such as the long and highly variable course of HIV disease, and evolving methods of treatment, precise predictions concerning the survival time of relatively healthy but infected individuals are not possible. Marker processes have other uses, though. One of the most important is in serving as responses in comparative treatment trials; a second is in serving as an adjustment or stratification factor in disease progression studies involving infected persons, especially when the times of infection are unknown.

The processes $Z(t)$ considered in Examples 11.5.2–11.5.4 are internal to individuals, and if lifetime T corresponds to the death of an individual the process $\{Z(t), t \geq 0\}$ terminates at T. This is also true of the usage process $\{x(t), t \geq 0\}$

in Example 11.5.1, although in applications such as designed experiments the usage process patterns can be prespecified, and thus are external. In each case the relationship between the processes and failure is of interest as a way of understanding the failure process, of preventing failure or extending lifetime, of aiding decisions about maintenance or treatment, and so on. Indeed, the future of lifetime analysis and of efforts to extend life in specific fields lies in a deeper understanding of failure and related processes. To deal with such issues it is necessary to consider joint models for failure and related processes. We discuss this next.

11.5.2 Some Approaches to Modeling and Analysis

Let us consider the following framework. Individuals are observed on some time scale $t \geq 0$, in terms of which they have some lifetime T. During their lifetime, certain internal processes $\mathcal{Z}(T) = \{\mathbf{Z}(t), 0 \leq t \leq T\}$ may be observed. In addition, certain observable external processes $X = \{\mathbf{x}(t), t \geq 0\}$ or covariates may affect T and $\mathcal{Z}(T)$. The key issue is to consider models for T and $\mathcal{Z}(T)$ given X. We outline some important approaches. References cited below and in the Bibliographic Notes can be consulted for further information.

11.5.2.1 Process-Defined Failures

In many settings failure is defined in terms of some internal process; Examples 11.5.2 and 11.5.3 provide illustrations. For simplicity consider a single positive-valued measure $Z(t)$ and suppose that failure time T is defined as

$$T = \inf\{t : Z(t) = z_T\}, \tag{11.5.1}$$

that is, the first time that $Z(t)$ reaches the specified threshold value $z_T > 0$. Given a model for the stochastic process $\{Z(t), t \geq 0\}$, the distribution of T can be determined from (11.5.1). If $Z(t)$ is monotonic in the sense that $Z(s) \leq Z(t)$ for $s \leq t$, then the relationship

$$Pr\{T > t\} = Pr\{Z(t) < z_T\} \tag{11.5.2}$$

can be used.

Models for which the distribution of T is fairly easily calculated include

1. Random growth-curve models (e.g., Carey and Koenig 1991; Lu and Meeker 1993), such as

$$Z_i(t) = \mu_i(t; \boldsymbol{\theta}_i, \boldsymbol{\beta}) \tag{11.5.3}$$

where $\boldsymbol{\theta}_i$ is a random vector of parameters specific to individual i and $\boldsymbol{\beta}$ is a vector of parameters common to all individuals. It is usually assumed that we cannot observe $Z_i(t)$, but only $Z_i^*(t) = Z_i(t) + e_i(t)$, where $\{e_i(t), t \geq 0\}$ is an error process. In many cases, $e_i(t)$ is assumed to include only measurement error, so that $e_i(t)$ and $e_i(s)$ are independent for $s \neq t$. Covariates can be included in $\mu_i(t)$.

2. Wiener processes (e.g., Doksum and Hoyland 1992; Doksum and Normand 1995; Whitmore 1995), for which $\{Z_i(t), t \geq 0\}$ is a stationary, independent increments Gaussian process with positive drift (e.g., Ross 1983). In this case, the process is not monotonic, but the distribution of T defined by (11.5.1) is inverse Gaussian, as considered in Sections 1.3.7, 4.2.2, and 6.5.2.
3. Jump processes (e.g., Taylor 1975; Cinlar 1977; Ditlevsen 1986). An example is compound Poisson shock processes, for which

$$Z(t) = \sum_{j=1}^{N(t)} Y_j = \int_0^t Y(t)\, dN(t) \qquad t \geq 0,$$

where $N(t)$ is the number of shocks over $(0, t)$ and follows a Poisson process, and Y_j is a positive-valued random variable associated with the jth shock.

To fit models of these types it is necessary to have data on the individual processes $\{Z_i(t), t \geq 0\}$ for a sample of individuals $i = 1, \ldots, n$. In most applications this is available at a discrete set of time points for individual i, say $t_{i1}, \ldots, t_{im_i}$, where t_{im_i} may correspond to T_i or to a final observation time at which failure has still not occurred. We will not pursue the details of writing down the likelihood contributions based on data $\{t_{ij}, z_i(t_{ij}); j = 1, \ldots, m_i\}$ for any of the models mentioned. These and other models are treated in references cited previously and in the Bibliographic Notes.

It should be noted that if a process $\{Z(t), t \geq 0\}$ is considered as a response without reference to any associated failure time, then the same modeling and analysis issues arise. However, when failure is present one must recognize the explicit connection between the $Z(t)$ process and failure in constructing the likelihood. The next subsection considers this for the more general setting in which T is associated with the $Z(t)$ process, but not necessarily defined by it.

11.5.2.2 Process-Related Failures

Consider the case where T and the process for $Z(t)$ are related; models for $\{T, \mathcal{Z}(T)\}$ are now wanted, where $\mathcal{Z}(t) = \{Z(s), 0 \leq s \leq t\}$. A way to approach this is through a model for the hazard function of T, given the past process history, along with a model for the evolution of the process. The hazard function is

$$h[t|\mathcal{Z}(t)] = \lim_{\Delta t \to 0} \frac{Pr\{T < t + \Delta t | T \geq t, \mathcal{Z}(t)\}}{\Delta t}. \tag{11.5.4}$$

A model for the evolution of $Z(t)$ over $(t, t + \Delta t)$, given $T \geq t$ and $\mathcal{Z}(t)$ must also be considered. Segall and Kailath (1975) and others show that if the $Z(t)$ process is predictable with respect to its past history and that of the failure process, then

(11.5.4) is rigorously defined and that for $t \geq 0, s \geq 0$,

$$Pr[T \geq t+s | T \geq t, \mathcal{Z}(t)] = E\left\{\exp\left[-\int_t^{t+s} h(u|\mathcal{Z}(u))\,du\right]\right\}, \qquad (11.5.5)$$

where the expectation is with respect to the process history $\mathcal{Z}(t+s)$, given $\mathcal{Z}(t)$ and $T \geq t$. The case where $t = 0$ gives the marginal survivor function for T, conditional on any pertinent information at $t = 0$.

The exact calculation of (11.5.5) is intractable for most models. Jewell and Kalbfleisch (1992, 1996), Singpurwalla (1995), Shi et al. (1996), and Bagdonavicius and Nikulin (2001) are among authors who consider reasonably tractable models, but few have been fitted to data.

The modeling approach just considered makes explicit the dependence of the failure time hazard function on the $Z(t)$ process. This is useful when one wants to assess the effect of covariates $\mathbf{x}$ on the $Z(t)$ process, but any effect of $\mathbf{x}$ on failure is confounded with the effect of the marker process. A different approach that is useful when the objective is to consider the marginal effect of $\mathbf{x}$ on failure is to take a model for T, given $\mathbf{x}$, combined with a model for $\mathcal{Z}(T)$, given T and $\mathbf{x}$. Some Gaussian models of this type are reasonably tractable (e.g., Cox 1999). A third approach is to link the failure and marker processes by assuming that they depend on common unobservable random effects (e.g., Faucett and Thomas 1996; Wulfsohn and Tsiatis 1997).

Typical data, mentioned in the preceding subsection, consist of observations on the $Z(t)$ process at discrete time points, along with the information that failure had or had not occurred by the last follow-up time. For the sake of discussion, suppose that a sequence of potential observation times t_{ij} $(j = 1, 2, \ldots)$ is prespecified for individual i, and that if an individual fails at t_i between two planned observation times, the value of t_i is observed. The data on individual i then consist of either

1. $\{Z(t_{ij}), j = 1, \ldots, m_i; T_i > t_{im_i}\}$, or
2. $\{Z(t_{ij}), j = 1, \ldots, m_i; T_i = t_i$, where $t_i < t_{i,m_i+1}\}$.

For maximum likelihood it is necessary to consider the probability of outcomes 1. or 2. For most models this is challenging to write down or compute. For calculations with defined failures and Wiener processes, as in Subsection 11.5.2.1, see Lu (1995); Whitmore et al. (1998) and Lee et al. (2000) extend this to a model involving a Gaussian marker process that is associated with failure.

11.5.2.3 Composite Measures

Another class of problems concerns measures that are functions of failure time, T, and process histories over $(0, T)$. We mention two specific examples below. In each setting the objective is to choose relatively simple models rather than to derive results from complex joint models for failure and the other processes.

1. **Composite Time Scales** In this case, the objective is to select a time scale other than real time. With equipment, such scales are often a function of real-time and usage processes $\{x(t), t \geq 0\}$; an example is the use of mileage driven or a combination of mileage and time in service with motor vehicles (e.g., Duchesne and Lawless 2000).
2. **Total Lifetime Cost or Quality** The emphasis here is on some accumulating measure of cost or quality associated with an individual's lifetime. For example, an individual who experiences a bout of disease may accumulate cost associated with its treatment. If $T \geq 0$ denotes the duration of the treatment period and $Z(t)$ is the total cost of treatment up to time t in that period, then the total lifetime cost is $Z = Z(T)$. Attempts are often made to find reasonably simple models for Z (e.g., Lin et al. 1997). A methodological challenge is that if data are observed in which some treatment durations T_i are censored at t_i, then the corresponding censoring value $z_i = Z(t_i)$ for Z_i is not in general the result of an independent censoring mechanism in the sense of Section 2.2.2. Equivalent problems occur with measures of total quality of life (e.g., Zhao and Tsiatis 1997) and other problems where a "mark" variable Y is observable at failure time T (Huang and Louis 1998).

BIBLIOGRAPHIC NOTES

Event history analysis is discussed in several books, including Blossfeld et al. (1989), Lancaster (1990), Andersen et al. (1993), Hougaard (2000) and Therneau and Grambsch (2000). Tuma and Hannan (1984) and Trussell et al. (1992) emphasize applications in the social sciences and demography, respectively.

The books by Kotz et al. (2000) and Joe (1997) provide extensive coverage of multivariate distributions and concepts of dependence or association. Hougaard (2000) discusses applications and statistical methods for multivariate lifetime distributions; Crowder et al. (1991, Ch. 7) and Klein and Moeschberger (1997, Ch. 3) give brief treatments with examples from engineering and medicine. Liang et al. (1995) review random effects and marginal models.

Copulas and multivariate distributions are thoroughly discussed by Joe (1997); Frees and Valdez (1998) give a useful survey, including applications to insurance. Clayton (1978) and Cook and Johnson (1981) introduced the model (11.2.6).

There has traditionally been a strong emphasis on random-effects models for multivariate lifetime data. Hougaard (1986) and Oakes (1989) are important early sources, and Hougaard (2000) covers the area well. Lindeboom and van den Berg (1994) stressed the need for multivariate random effects; see also McGilchrist (1993), Petersen et al. (1996), and Hougaard (2000, Ch. 10). There is a close connection between random-effects models and copula models (e.g., Marshall and Olkin 1988; Oakes 1989), though different regression specifications tend to be used with the two types, as discussed in Section 11.2.

Paired experiments involving lifetimes were discussed by Holt and Prentice (1974), Kalbfleisch and Prentice (1980, Sec. 8.1), and Wild (1983). Manatunga and

Oakes (1999) provide an excellent discussion of the different ways of specifying covariate effects, as in Example 11.2.1. Neuhaus and Kalbfleisch (1998) discuss marginal and conditional covariate effects in a broader context.

Maximum likelihood for parametric multivariate models is standard (e.g., Huster et al. 1989; Crowder et al. 1991; Frees and Valdez 1998). Nonparametric estimation of a continuous multivariate survivor function $S(t_1, \ldots, t_k)$ from censored data is difficult; Van der Laan (1996), Hougaard (2000, Ch. 14), and Oakes (2001, Sec. 7.2) discuss the topic.

Huster et al. (1989) considered pseudolikelihood methods for independence working models and compared their efficiency with maximum likelihood based on the bivariate Clayton model. Wei et al. (1989) and Lee et al. (1992) considered a similar approach with semiparametric PH models for the marginal distributions. Further contributions, some using weighted estimating functions, are in Liang et al. (1993), Lin (1994), Spiekerman and Lin (1998), Cai and Prentice (1995), and Prentice and Hsu (1997). He and Lawless (2002) consider PH models with flexible piecewise-constant and regression spline baseline hazard functions. Working independence approaches for models with accelerated failure time marginal distributions have been considered by Lee et al. (1993). Genest et al. (1995), Shih and Louis (1995), and Fine and Jiang (2000) consider pseudolikelihood estimation of copula parameters like ϕ in (11.2.6), along with estimation of marginal distributions. Williams (1995) and Lawless (2003) consider variance estimation for Kaplan–Meier estimates as in (11.2.21). For simple semiparametric PH frailty models, a maximum likelihood approach is available (e.g., Klein 1992; Nielsen et al. 1992; Andersen et al. 1997; Hougaard 2000, Ch. 8). For models with multivariate random-effects see Liang et al. (1995, Sec. 2) and Hougaard (2000, Ch. 10). Bandeen-Roche and Liang (1996), Lin (2000), and Lawless (2003) consider marginal methods that can be applied to sample survey data.

Sequences of gap times were considered under PH models by Cox (1972b), Gail et al. (1980), Kalbfleisch and Prentice (1980, Sec. 7.3), and Prentice et al. (1981). Follman and Goldberg (1988) and Aalen and Husebye (1991) considered simple random-effects models of PH and AFT form; Xue and Brookmeyer (1996), Cook et al. (1999), and Fong et al. (2001) consider more complex models. Dabrowska et al. (1994) provide a rigorous discussion of semiparametric PH analysis. Visser (1996), Wang and Wells (1998), and Lin et al. (1999) consider nonparametric estimation of the marginal distributions for second and subsequent gap times. Lawless and Fong (1999) and Lawless et al. (2001) discuss modeling and analysis of recurrent durations and gap times.

Statistical methods for more general event history processes, especially of Markov type, are discussed by Blossfeld et al. (1989), Andersen et al. (1993), Hougaard (2000) and Therneau and Grambsch (2000). Andersen and Keiding (2002) review multistate models; Lawless (1995) and Cook and Lawless (2002) survey methodology for recurrent events. Ascher and Feingold (1984) provide extensive discussion of repeated failures in reliability. Kalbfleisch and Lawless (1985, 1988a), Gentleman et al. (1994), and Lindsey and Ryan (1993) consider incomplete data for multistate models. Pepe and Mori (1993) discuss marginal characteristics of multistate models.

There is a large literature on joint models for failure and associated processes. Cox (1999) provides a synopsis of model types. Degradation models with defined failures are very well studied, with approaches that include damage processes (e.g., Bogdanoff and Kozin 1985; Desmond 1985; Sobczyk 1987), diffusion processes (e.g., Doksum and Hoyland 1992; Doksum and Normand 1995; Lu 1995; Whitmore 1995; Whitmore and Schenkelberg 1997), and random-effects models (Meeker and Escobar, 1998, Ch. 13; Lu et al. 1997; Sun et al. 1999). Aalen and Gjessing (2001) survey models with random boundaries or starting points. The operations research literature contains many papers, for example, on discrete time, discrete $Z(t)$ Markov models for maintenance and replacement planning (e.g., Valdez-Flores and Feldman 1989).

Jewell and Kalbfleisch (1992, 1996), Self and Pawitan (1992), Jewell and Nielsen (1993), Singpurwalla (1995), and Shi et al. (1996) discuss aspects of hazard-based models for process-related failures. Joint Gaussian models for T and $\mathcal{Z}(T)$ are considered by Taylor et al. (1994), Tsiatis et al. (1995), Cox (1999), and others. Models involving random-effects linking lifetime and marker process variables are discussed by Faucett and Thomas (1996), Wulfsohn and Tsiatis (1997), Hogan and Laird (1997a), Xu and Zeger (2001), and others. Hogan and Laird (1997b) review different approaches to modeling. Models based on multivariate diffusion processes are considered by Whitmore et al. (1998) and Lee et al. (2000).

The analysis of data on lifetimes and total lifetime cost or quality measures has been considered by Lin et al. (1997), Zhao and Tsiatis (1997), Cook and Lawless (1997), Huang and Louis (1998), Bang and Tsiatis (2000), Ghosh and Lin (2000), and others. Cook and Lawless (2002) provide additional references. A related area is the treatment of alternative time scales (e.g., Kordonsky and Gertsbakh 1993; Oakes 1995; Duchesne and Lawless 2000; Wilson 2000), in which scales that are functions of external usage or exposure processes are considered.

COMPUTATIONAL NOTES

Specialized software for specific parametric multivariate models and censored data is not widely available, so the use of general optimization methodology (Appendix D) is recommended. Marginal analysis of the semiparametric PH model as discussed by Wei et al. (1989), Lee et al. (1992), and Spiekerman and Lin (1998) is available in S-Plus (see function coxph with the cluster option) and other systems such as SAS and STATA. Lin (1994) discusses other software for marginal PH analysis. Survey data-analysis software such as SUDAAN and WESVAR also provide methods of variance estimation for working independence models, emphasizing data from sample surveys.

Conditional semiparametric PH random-effects models such as (11.2.23) are handled by S-Plus function coxph with the frailty option. Gamma and other distributions for the shared random effects are allowed.

Gap time analysis can utilize standard univariate survival analysis software, as can certain analyses for multistate models or multiple events. In addition to the examples

in Sections 11.3 and 11.4, the books by Hougaard (2000) and Therneau and Grambsch (2000) can be consulted for detailed instructions.

Not much specialized software has been developed for the models of Section 11.5, though for the analysis of some marker processes alone, software for continuous or discrete longitudinal data can be utilized.

PROBLEMS AND SUPPLEMENTS

11.1 Give expressions for the joint survivor function and p.d.f. of bivariate lifetimes T_1, T_2 in terms of the three hazard functions in (11.2.4).

(Section 11.2)

11.2 *Dependence or association measures.* Let (T_1, T_2) have continuous c.d.f. $F(t_1, t_2)$. Two measures of the association between T_1 and T_2 are Spearman's ρ and Kendall's τ, defined, respectively, as

$$\rho = \text{Corr}[S_1(T_1), S_2(T_2)]$$

$$\tau = E\{\text{sign}[(T_1 - T_1')(T_2 - T_2')]\}$$

where (T_1, T_2) and (T_1', T_2') are independent pairs from $F(t_1, t_2)$ and $\text{Corr}(X, Y)$ stands for the correlation of random variables X and Y.

(a) Show that

$$\rho = 12 \int_0^\infty \int_0^\infty F_1(t_1) F_2(t_2)\, dF(t_1, t_2) - 3$$

and

$$\tau = 4 \int_0^\infty \int_0^\infty F(t_1, t_2)\, dF(t_1, t_2) - 1.$$

Show that ρ and τ equal zero when T_1 and T_2 are independent.

(b) Write ρ and τ in terms of the associated copula $C(u_1, u_2)$ of (11.2.5). Show that ρ and τ are invariant to strictly monotone increasing transformations of T_1 or T_2.

(c) The covariance of T_1 and T_2 is equal to $E(T_1T_2) - E(T_1)E(T_2)$. Prove that (assuming it exists)

$$E(T_1T_2) = \int_0^\infty \int_0^\infty S(t_1, t_2)\, dt_1\, dt_2.$$

This generalizes a result in part (a) of Problem 1.1, and represents $\text{Cov}(T_1, T_2)$ in terms of $S(t_1, t_2)$.

(Section 11.2)

11.3 Consider the bivariate Clayton family (11.2.6).

(a) Show that Kendall's τ defined in Problem 11.2 equals $\phi/(\phi + 2)$. There is no simple closed form for Spearman's ρ.

(b) Show also that for any (t_1, t_2)

$$\lim_{\Delta t \to 0} \frac{Pr(T_2 < t_2 + \Delta t | T_1 = t_1, T_2 \geq t_2)}{Pr(T_2 < t_2 + \Delta t | T_1 > t_1, T_2 \geq t_2)} = 1 + \phi.$$

(Section 11.2; Clayton 1978; Cook and Johnson 1981; Oakes 1982)

11.4 For the multivariate Burr model (11.2.8), show that $\text{Corr}(\log T_1, \log T_2) = \psi'(\nu)/[\psi'(\nu) + \pi^2/6]$, where $\psi(\cdot)$ is the digamma function (B10).

(Section 11.2)

11.5 Another bivariate family of the form (11.2.5) has

$$S(t_1, t_2) = \exp\{-[(-\log S_1(t_1))^{1/\alpha} + (-\log S_2(t_2))^{1/\alpha}]^\alpha\},$$

where $\alpha \leq 1$. Show that the case where T_1 and T_2 are independent is given by $\alpha = 1$. Show that Kendall's τ is $1 - \alpha$, so smaller values of α correspond to a higher degree of association.

(Section 11.2; Hougaard 1986)

11.6 Consider the Clayton family (11.2.6) in terms of the p.d.f. and s.f.'s for $Y_1 = \log T_1$ and $Y_2 = \log T_2$. Graph contours of the p.d.f. $f(y_1, y_2)$ for different values of ϕ in the case where the marginal distributions $F_1(y_1)$ and $F_2(y_2)$ are standard normal. Compare the shapes of these contours with ones from the p.d.f. of a bivariate normal distribution with $N(0, 1)$ marginal distributions and correlation parameter ρ; it has p.d.f.

$$f(y_1, y_2) = \frac{1}{2\pi(1 - \rho^2)^{1/2}} \exp\left\{\frac{-(y_1^2 + y_2^2 - 2\rho y_1 y_2)}{2(1 - \rho^2)}\right\}$$

for $-\infty < y_1 < \infty$, $-\infty < y_2 < \infty$.

(Section 11.2; Cook and Johnson 1981)

11.7 Consider the paired data of Problem 5.15 on the survival times of treatment and control rats in a toxicology test. Give an alternative to the analysis suggested there by assuming that the log survival times Y_{1i} and Y_{2i} for the treatment and control rats in the ith pair have normal marginal distributions with means μ_1 and μ_2 and common variance σ^2. Use the working independence procedure of Section 11.2.2 to estimate $\mu_1 - \mu_2$ and obtain a robust standard error via (11.2.17). Consider the alternative of fitting a bivariate normal distribution for (Y_{1i}, Y_{2i}). Discuss how you could check on the equality of marginal variances in each case.

(Section 11.2)

Table 11.7. Times to Tumor for Treated (T) and Untreated ($U1, U2$) Rats

Litter	T	$U1$	$U2$	Litter	T	$U1$	$U2$
1	101.0[a]	49.0	104.0[a]	26	89.0[a]	104.0[a]	104.0[a]
2	104.0[a]	102.0[a]	104.0[a]	27	78.0[a]	104.0[a]	104.0[a]
3	104.0[a]	104.0[a]	104.0[a]	28	104.0[a]	81.0	64.0
4	77.0[a]	97.0[a]	79.0[a]	29	86.0	55.0	94.0[a]
5	89.0[a]	104.0[a]	104.0[a]	30	34.0	104.0[a]	54.0
6	88.0	96.0	104.0[a]	31	76.0[a]	87.0[a]	74.0[a]
7	104.0	94.0[a]	77.0	32	103.0	73.0	84.0
8	96.0	104.0[a]	104.0[a]	33	102.0	104.0[a]	80.0[a]
9	82.0[a]	77.0[a]	104.0[a]	34	80.0	104.0[a]	73.0[a]
10	70.0	104.0[a]	77.0[a]	35	45.0	79.0[a]	104.0[a]
11	89.0	91.0[a]	90.0[a]	36	94.0	104.0[a]	104.0[a]
12	91.0[a]	70.0[a]	92.0[a]	37	104.0[a]	104.0[a]	104.0[a]
13	39.0	45.0[a]	50.0	38	104.0[a]	101.0	94.0[a]
14	103.0	69.0[a]	91.0[a]	39	76.0[a]	84.0	78.0
15	93.0[a]	104.0[a]	103.0[a]	40	80.0	81.0	76.0[a]
16	85.0[a]	72.0[a]	104.0[a]	41	72.0	95.0[a]	104.0[a]
17	104.0[a]	63.0[a]	104.0[a]	42	73.0	104.0[a]	66.0
18	104.0[a]	104.0[a]	74.0[a]	43	92.0	104.0[a]	102.0
19	81.0[a]	104.0[a]	69.0[a]	44	104.0[a]	98.0[a]	73.0[a]
20	67.0	104.0[a]	68.0	45	55.0[a]	104.0[a]	104.0[a]
21	104.0[a]	104.0[a]	104.0[a]	46	49.0[a]	83.0[a]	77.0[a]
22	104.0[a]	104.0[a]	104.0[a]	47	89.0	104.0[a]	104.0[a]
23	104.0[a]	83.0[a]	40.0	48	88.0[a]	79.0[a]	99.0[a]
24	87.0[a]	104.0[a]	104.0[a]	49	103.0	91.0[a]	104.0[a]
25	104.0[a]	104.0[a]	104.0[a]	50	104.0[a]	104.0[a]	79.0

[a]Right-censored times.

11.8 Mantel et al. (1977) discussed data from an animal carcinogenicity study in which three rats from each of 50 litters were selected. One rat from each litter was treated with a suspected carcinogen and the other two rats were untreated controls. The response was time-to-tumor or censoring, recorded to the nearest week. The data are given in Table 11.7.

Use the working independence approach to fit Cox PH models for the marginal distributions of time to tumor T, given the treatment covariate $x = I$ (suspected carcinogen received), and test for the absence of a treatment effect. Compare the test that assumes independence of the times (T_{1l}, T_{2l}, T_{3l}) for rats in the same litter with one that uses a robust variance estimate.

(Section 11.2)

11.9 *(Continuation of Problem 11.8).* It is permissible to compare the Kaplan–Meier estimates $\hat{S}(t)$ based on the times to tumor for treated animals and untreated animals, respectively. (Why?)

(a) Plot the two estimates on the same graph; this provides another picture of the population-level treatment effect.

(b) Standard errors or confidence limits based on the usual variance estimate (3.2.3) are appropriate for the treated group, since there is only one animal per litter. If there is significant within-litter association, then because there are two untreated animals per litter, (3.2.3) is not valid for this group. Compute and compare standard errors for $\hat{S}(t)$ based on (3.2.3) and on the estimate (11.2.21), which accounts for within-litter association.

(Section 11.2)

11.10 Derive the variance estimate (11.2.21) by considering a discrete-time setting for $S(t)$, $t = 1, 2, \ldots$, where $h(t) = Pr(T = t)/Pr(T \geq t)$ is the hazard function. Proceed by noting that the working independence model that gives the standard Kaplan–Meier estimate based on all nk times t_{ji} and censoring indicators δ_{ji} $(j = 1, \ldots, k; i = 1, \ldots, n)$ can be obtained from the estimating functions

$$U_t = \sum_{i=1}^{n} \sum_{j=1}^{k} Y_{ji}(t) \left\{ \frac{d_{ji}(t) - h(t)}{h(t)[1 - h(t)]} \right\} \qquad t = 1, \ldots, M,$$

where $d_{ji}(t) = I(t_{ji} = t, \delta_{ji} = 1)$, $Y_{ji}(t) = I(t_{ji} \geq t)$, and M is the upper limit of the support for $S(t)$. Use the sandwich variance formula (C30) to get an estimate of the asymptotic covariance matrix for $\hat{\mathbf{h}} = (\hat{h}(1), \ldots, \hat{h}(M))'$, and from this obtain a variance estimate for

$$\hat{S}(t) = \prod_{u=1}^{t-1} [1 - \hat{h}(u)].$$

(Section 11.2; Williams 1995; Lawless 2003)

11.11 Consider the data in Example 11.1.2 on time T_1 to initiation and subsequent time T_2 to failure for electrical-cable-insulation specimens subjected to a high-voltage life test. Examine using methods of Section 11.2 or 11.3 whether there appears to be any association between T_1 and T_2.

(Sections 11.2, 11.3)

11.12 Aalen and Husebye (1991) gave data from a study on muscular activity (motility) of the small bowel in humans. Nineteen subjects were each given a standard meal and then followed over a period of about 14 hours. One response of interest was the time intervals or gaps between successive digestive cycles, called migrating motor complex (MMC) cycles. Data on these gap times are given in Table 11.8; the lengths of successive gaps (in minutes) are shown. The numbers of cycles observed varied from subject to subject, so the number of gaps varies; in addition, each final gap time is censored because observation terminated during the gap.

Table 11.8. Lengths of Successive Motility Cycle Gaps

Individual	Times (minutes)
1	112, 145, 39, 52, 21, 34, 33, 51, 54[a]
2	206, 147, 30[a]
3	284, 59, 186, 4[a]
4	94, 98, 84, 87[a]
5	67, 131[a]
6	124, 34, 87, 75, 43, 38, 58, 142, 75, 23[a]
7	116, 71, 83, 68, 125, 111[a]
8	111, 59, 47, 95, 110[a]
9	98, 161, 154, 55, 44[a]
10	166, 56, 122[a]
11	63, 90, 63, 103, 51, 85[a]
12	47, 86, 68, 144, 72[a]
13	120, 106, 176, 6[a]
14	112, 25, 57, 166, 85[a]
15	132, 267, 89, 86[a]
16	120, 47, 165, 64, 113, 12[a]
17	162, 141, 107, 69, 39[a]
18	106, 56, 158, 41, 41, 168, 13[a]
19	147, 134, 78, 66, 100, 4[a]

[a] Censoring time

(a) Show that there is no strong association among the successive gap times for an individual.

(b) Compare Kaplan–Meier estimates for the gap duration distributions for (1) first gaps, and (2) all subsequent gaps. Explain why it is permissible to use the Kaplan–Meier estimate for (2) if there is no association between gap times, but not if there is association.

(Section 11.3)

11.13 The data in Table 11.9 were given by McGilchrist and Aisbett (1991) and show the recurrent times to infection at the point of insertion of the catheter for 38 persons undergoing kidney dialysis. Data for the first two occurrences of infection are given; either one or both may be censored, because catheters were sometimes removed for causes other than infection. Covariates are also given: age (years), sex (1 = male, 2 = female), and kidney disease type (0 = glomerulo nephritis, 1 = acute nephritis, 2 = polycystic kidney disease, 3 = other).

Assess the effects of covariates on times to infection, bearing in mind that there may be association between times for a given patient even after conditioning on covariates.

(Sections 11.2, 11.3; McGilchrist and Aisbett 1991)

Table 11.9. Times to Two Successive Infections for Patients on Dialysis

Patient	Times	Censoring	Age	Sex	Disease Type
1	8, 16	1, 1[a]	28	1	3
2	23, 13	1, 0	48	2	0
3	22, 28	1, 1	32	1	3
4	447, 318	1, 1	31–32	2	3
5	30, 12	1, 1	10	1	3
6	24, 245	1, 1	16–17	2	3
7	7, 9	1, 1	51	1	0
8	511, 30	1, 1	55–56	2	0
9	53, 196	1, 1	69	2	1
10	15, 154	1, 1	51–52	1	0
11	7, 333	1, 1	44	2	1
12	141, 8	1, 0	34	2	3
13	96, 38	1, 1	35	2	1
14	149, 70	0, 0	42	2	1
15	536, 25	1, 0	17	2	3
16	17, 4	1, 0	60	1	1
17	185, 177	1, 1	60	2	3
18	292, 114	1, 1	43–44	2	3
19	22, 159	0, 0	53	2	0
20	15, 108	1, 0	44	2	3
21	152, 562	1, 1	46–47	1	2
22	402, 24	1, 0	30	2	3
23	13, 66	1, 1	62–63	2	1
24	39, 46	1, 0	42–43	2	1
25	12, 40	1, 1	43	1	1
26	113, 201	0, 1	57–58	2	1
27	132, 156	1, 1	10	2	0
28	34, 30	1, 1	52	2	1
29	2, 25	1, 1	53	1	0
30	130, 26	1, 1	54	2	0
31	27, 58	1, 1	56	2	1
32	5, 43	0, 1	50–51	2	1
33	152, 30	1, 1	57	2	2
34	190, 5	1, 0	44–45	2	0
35	119, 8	1, 1	22	2	3
36	54, 16	0, 0	42	2	3
37	6, 78	0, 1	52	2	2
38	63, 8	1, 0	60	1	2

[a] 1-uncensored, 0-censored

11.14 *Poisson processes for recurrent events.* Consider a process where the same event can occur repeatedly for an individual or unit. A (nonhomogeneous) Poisson process is one for which the event intensity function defined by (11.4.1) (with $k = 1$) is of the form $\lambda[t|\mathcal{H}(t)] = \rho(t)$. That is, the event intensity depends on t, but not on the history of event occurrences prior to t.

(a) Suppose that a single individual is observed over the specified time interval $(0, \tau)$. Show using (11.4.4) and (11.4.5) that the probability density for the outcome "n events occur, at times $t_1 < \cdots < t_n$" is given by

$$\left[\prod_{i=1}^{n} \rho(t_i)\right] \exp\left[-\int_0^\tau \rho(u)\,du\right]. \tag{11.6.1}$$

(b) Consider the parametric model with $\rho(t) = \exp(\alpha + \beta t)$. Write down the likelihood function $L(\alpha, \beta)$ using (11.6.1).

(c) The likelihood function for α and β is also given by (11.6.1) in the case where τ is defined as the time t_n of the nth failure, where n is prespecified. Explain why this is so.

(d) Consider the data on the times of repeated failures in aircraft air conditioning systems, given in Example 11.1.4. Fit the model of part (b) separately for each plane, by maximizing $\log L(\alpha, \beta)$. Test the hypothesis that $\beta = 0$ in each case; this examines whether there is a trend toward an increasing (or decreasing) failure rate as time passes.

(e) The Poisson process has the property that the expected number of events in $(0, t)$ is

$$E[N(0, t)] = \int_0^t \rho(u)\,du = R(t).$$

Check the fit of the models in part (d) by plotting $N(0, t)$ versus t and $R(t; \hat{\alpha}, \hat{\beta})$ versus t on the same graph for each plane.

(Section 11.4; Cox and Lewis 1966; Ascher and Feingold 1984)

11.15 Consider a discrete-time model for the setting in Subsection 11.5.2.2 of Section 11.5.2 involving process-related failures, where $t = 1, 2, 3, \ldots$, and

$$Pr\{T = t+1, Z(t+1)|\mathcal{Z}(t), T > t\}$$
$$= Pr\{Z(t+1)|\mathcal{Z}(t), T > t\} Pr\{T = t+1|\mathcal{Z}(t+1), T > t\},$$

with $\mathcal{Z}(t) = \{Z(1), \ldots, Z(t)\}$ the history process for $Z(t)$. Give an expression for $Pr\{T > t+s|T \geq t, \mathcal{Z}(t)\}$ and give a heuristic argument leading to (11.5.5).

(Section 11.5)

11.16 *Process-related failures.* Suppose that in (11.5.4) the hazard function for failure depends only on the current value $Z(t)$; that is,

$$h[t|\mathcal{Z}(t)] = g[Z(t)].$$

Outline a procedure for estimation of the function $g(z)$, assuming that you can observe the data $(t_i, \delta_i, \mathcal{Z}_i(t_i))$ for a random sample of individuals $i = 1, \ldots, n$. Here, t_i is either a failure or censoring time, with $\delta_i = 1$ indicating a failure at t_i. Discuss an ad hoc procedure for dealing with intermittent observation of the $\{Z_i(t), t \geq 0\}$ processes.

(Section 11.5; Fusaro et al. 1993)

APPENDIX A

Glossary of Notation and Abbreviations

A.1 NOTATION AND SYMBOLS

$\mathbf{a}$	Bold denotes vectors (column form)
$\mathbf{a}'$, A'	Transpose of a vector $\mathbf{a}$ or matrix A
$G(\mathbf{a})$	Function $G(a_1, \ldots, a_k)$ for vector $\mathbf{a} = (a_1, \ldots, a_k)'$
$\partial G/\partial \mathbf{a}$	Vector $(\partial G/\partial a_1, \ldots, \partial G/\partial a_k)'$
$\partial^2 G/\partial \mathbf{a} \partial \mathbf{a}'$	$k \times k$ matrix $(\partial^2 G/\partial a_i \partial a_j)$
$I(A)$	Binary indicator function (see Example 2.1.1)
$o(x)$	Denotes a function $g(x)$ satisfying $\lim_{x \to 0}[g(x)/x] = 0$
$G(t-)$	$\lim_{x \uparrow t} G(x)$
$G(t+)$	$\lim_{x \downarrow t} G(x)$
$dG(t)$	Defined for nondecreasing, right-continuous functions or processes $G(t)$ (see Section 1.2.3)
$\int_a^b dG(u)$	Riemann–Stieltjes integral (see (1.2.9) in Section 1.2.3)
$\int_a^b g(u)\, du$	Riemann integral
$\prod_{(a,b]} \{1 + dG(u)\}$	Product integral (see (1.2.11) in Section 1.2.3)
$\mathcal{H}(t)$	History of a process up to time t (see Section 2.2.2)
$N_i(t)$, $C_i(t)$, $Y_i(t)$	Processes for indicating failure, censoring and risk status (see Section 2.2.2)
$Pr(A)$	Probability mass or density for an event A
$Y \sim F(y)$	The random variable Y has distribution function $F(y)$
$E(Y)$	Expectation (mean) of Y
$E(Y_1 \mid y_2)$	Conditional expectation of Y_1, given $Y_2 = y_2$
$\mathrm{Var}(Y)$	Variance of Y

$sd(Y)$	Standard deviation of Y
$\mathrm{Cov}(Y_1, Y_2)$	Covariance of Y_1 and Y_2
$\mathrm{Corr}(Y_1, Y_2)$	Correlation of Y_1 and Y_2
y_p	pth quantile of the random variable Y (see Section 1.2.1)
$\mathrm{Asvar}(Y_n)$	Asymptotic variance for a sequence of random variables $\{Y_n\}$
$\mathrm{Ascov}(Y_n, Z_n)$	Asymptotic covariance for a sequence of random variables $\{(Y_n, Z_n)\}$
$L(\boldsymbol{\theta})$	A likelihood function for a parameter $\boldsymbol{\theta}$ (see Section 2.1)
$\ell(\boldsymbol{\theta})$	A log-likelihood function (see Section 2.2.3)
$I(\boldsymbol{\theta})$	An observed information matrix (see Section 2.2.3)
$\mathcal{I}(\boldsymbol{\theta})$	An expected (Fisher) information matrix (see Section 2.2.3)
$\hat{\boldsymbol{\theta}}$	An estimate or estimator of $\boldsymbol{\theta}$, usually the m.l.e.
$se(\hat{\psi})$	Standard error of parameter estimate $\hat{\psi}$, usually an estimate of $sd(\hat{\psi})$.
$\phi(z)$	Probability density function for standard normal distribution
$\Phi(z)$	Distribution function for standard normal distribution
$\Gamma(x)$	Gamma function (see Appendix B.2)
$I(k, x)$	Incomplete gamma function and gamma distribution c.d.f. (see Section 1.3.5 and (B12) of Appendix B)
$\Delta_\ell(t)$	$\int_{a_{\ell-1}}^{a_\ell} I(u \le t)\, du$ for $\ell = 1, 2, 3, \ldots$ and $a_0 < a_1 < a_2 < \cdots$ (see (1.3.26))

A.2 ABBREVIATIONS

AFT	Accelerated failure time
c.d.f., d.f.	(Cumulative) distribution function
c.h.f.	Cumulative hazard function
EDF	Empirical distribution function
ESF	Empirical survivor function
h.f.	Hazard function
i.i.d.	Independent and identically distributed
KM	Kaplan–Meier (estimate)
m.l.e.	Maximum likelihood estimate
ML	Maximum likelihood
NA	Nelson–Aalen (estimate)
p.d.f.	Probability density function
p.f.	Probability (mass) function
PH	Proportional hazards

PL	Product-limit (estimate)
s.f.	Survivor function
$\mathrm{Binomial}(n, p)$	Binomial distribution
$\mathrm{Exp}(\theta)$	Exponential distribution (1.3.3)
$\mathrm{Weib}(\alpha, \beta)$	Weibull distribution (1.3.5) with $\lambda = \alpha^{-1}$
$EV(u, b)$	Extreme value distribution (1.3.8)
$F_{(r,s)}$	F distribution with (r, s) degrees of freedom
$Ga(k)$	One-parameter gamma distribution (1.3.17)
$IG(\mu, \lambda)$	Inverse Gaussian distribution (1.3.23)
$\mathrm{Logist}(u, b)$	Logistic distribution (1.3.14)
$\mathrm{LLogist}(\alpha, \beta)$	Log-logistic distribution (1.3.12)
$\mathrm{LogN}(\mu, \sigma^2)$	Log-normal distribution (1.3.10)
$N(\mu, \sigma^2)$	Normal distribution with mean μ, variance σ^2
$t_{(r)}$	Student's-t distribution with r degrees of freedom
$\chi^2_{(r)}$	Chi-squared distribution with r degrees of freedom
$\mathrm{Uniform}(a, b)$ or $U(a, b)$	Uniform distribution on (a, b)

APPENDIX B

Asymptotic Variance Formulas, Gamma Functions, and Order Statistics

B.1 ASYMPTOTIC VARIANCE FORMULAS

The following results are often used in developing large-sample inference procedures. Proofs can be found, for example, in Rao (1973, Ch. 2). Here "$\xrightarrow{D}$" means "converges in distribution to."

THEOREM B1. Let $T_{1n}, \ldots, T_{kn}$ be statistics such that as $n \to \infty$

$$\sqrt{n}(T_{1n} - \theta_1, \ldots, T_{kn} - \theta_k) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma} = (\sigma_{ij})_{k\times k}$. If $g(x_1, \ldots, x_k)$ is a function whose first derivatives all exist, then as $n \to \infty$

$$\sqrt{n}\,[g(T_{1n}, \ldots, T_{kn}) - g(\theta_1, \ldots, \theta_k)] \xrightarrow{D} N\left(0, \sum_{i=1}^{k}\sum_{j=1}^{k}\sigma_{ij}\frac{\partial g}{\partial \theta_i}\frac{\partial g}{\partial \theta_j}\right) \tag{B1}$$

where $\partial g/\partial\theta_i$ means $\partial g(\theta_1, \ldots, \theta_k)/\partial\theta_i$ $(i = 1, \ldots, k)$.

Remarks

1. Often the following terminology is used:

$$\text{Asvar}\,[g(T_{1n}, \ldots, T_{kn})] = \sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial g}{\partial \theta_i}\frac{\partial g}{\partial \theta_j}\text{Ascov}(T_{in}, T_{jn}), \tag{B2}$$

where Asvar and Ascov denote variances and covariances in the asymptotic distributions of the indicated variables. Strictly speaking, this notation is

improper since the distributions of the T_{in} and $g(T_{1n}, \ldots, T_{kn})$ are degenerate as $n \to \infty$. However, it is used as a convention for indicating asymptotic-based approximations that are used for finite but large n. (In (B2), Ascov(T_{in}, T_{jn}) is $n^{-1}\sigma_{ij}$, for example.)

2. The results in this section are often used when $(T_{1n}, \ldots, T_{kn}) = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ is a vector of maximum likelihood estimates (m.l.e.'s), based on a sample of size n.
3. The results here are stated for statistics with asymptotic normal distributions. Expressions like (B2) also hold under the weaker conditions that the variances and covariances of $T_{1n}, \ldots, T_{kn}$ are $O(n^{-r})$, where $r > 0$.
4. An important special case of Theorem B1 is given by $k = 1$: if $\sqrt{n}(T_n - \theta) \xrightarrow{D} N(0, \sigma^2)$ as $n \to \infty$, then if $g(x)$ has first derivative $g'(x)$,

$$\sqrt{n}\,[g(T_n) - g(\theta)] \xrightarrow{D} N\left[0, g'(\theta)^2\sigma^2\right]. \tag{B3}$$

This implies that, in the notation of (B2),

$$\text{Asvar}[g(T_n)] = g'(\theta)^2 \text{Asvar}(T_n). \tag{B4}$$

5. The preceding results can be proved with what is sometimes referred to as the δ method, based on Taylor series expansions. For example, the function $g(T_{1n}, \ldots, T_{kn})$ has expansion

$$g(T_{1n}, \ldots, T_{kn}) = g(\theta_1, \ldots, \theta_k) + \sum_{i=1}^{k} \delta T_{in} \frac{\partial g}{\partial \theta_i} + \text{higher-order terms},$$

where $\delta T_{in} = T_{in} - \theta_i$. The results follow from this and simple convergence results for random variables. Theorems B1 and B2 below are sometimes referred to as delta theorems.

Theorem B1 can be generalized to the case in which there are several functions of $T_1, \ldots, T_{kn}$, as follows.

THEOREM B2. Let $(T_{1n}, \ldots, T_{kn})$ be statistics defined as in Theorem B1 and let $g_i(x_1, \ldots, x_k)$, $i = 1, \ldots, p$, be functions, all of whose first derivatives exist. Then the joint distribution of $\sqrt{n}[g_i(T_{1n}, \ldots, T_{kn}) - g_i(\theta_1, \ldots, \theta_k)]$, $i = 1, \ldots, p$, is asymptotically p-variate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{G\Sigma G'}$, where $\mathbf{G}$ has (i, j) entry $G_{ij} = \partial g_i / \partial \theta_j$.

Remark For two functions $g_1(x_1, \ldots, x_k)$ and $g_2(x_1, \ldots, x_k)$, the theorem gives, in the notation of (B2),

$$\text{Ascov}\,[g_1(T_{1n}, \ldots, T_{kn}), g_2(T_{1n}, \ldots, T_{kn})] = \sum_{l=1}^{k}\sum_{j=1}^{k} \frac{\partial g_1}{\partial \theta_l}\frac{\partial g_2}{\partial \theta_j}\text{Ascov}(T_{in}, T_{jn}). \tag{B5}$$

B.2 GAMMA FUNCTIONS

We summarize a few results about the gamma function and other functions and probability distributions related to it. More details on these topics can be found in the books by Abramowitz and Stegun (1965), and Johnson et al. (1994, 1995).

The *gamma function* is defined as

$$\Gamma(z) = \int_0^\infty u^{z-1}e^{-u}du \qquad z > 0. \tag{B6}$$

We note the well-known results (see Abramowitz and Stegun 1965, Ch. 6)

$$\Gamma(z+1) = z\Gamma(z) \qquad z > 0 \tag{B7}$$

$$\Gamma\left(\frac{1}{2}\right) = \pi^{1/2} = 1.77245\ldots \tag{B8}$$

$$\log\Gamma(z) = \left(z - \frac{1}{2}\right)\log z - z + \frac{1}{2}\log(2\pi) + \frac{1}{12z} - \frac{1}{360z^3} + \cdots. \tag{B9}$$

It follows from (B6) and (B7) that for z a positive integer, $\Gamma(z+1) = z!$ The *digamma function* is defined as

$$\psi(z) = \frac{d\log\Gamma(z)}{dz} = \frac{\Gamma'(z)}{\Gamma(z)} \qquad z > 0. \tag{B10}$$

The *polygamma functions* are

$$\psi^{(n)}(z) = \frac{d^n\psi(z)}{dz^n} \qquad n = 1, 2, \ldots.$$

The case $n = 1$ is called the *trigamma function*. Two useful results are

$$\begin{aligned}\psi(1) &= -\gamma = -.577215\ldots,\\ \psi'(1) &= \frac{\pi^2}{6}.\end{aligned} \tag{B11}$$

The *incomplete gamma function* is defined in this book as

$$I(k, x) = \frac{1}{\Gamma(k)}\int_0^x u^{k-1}e^{-u}du \qquad k > 0 \qquad x > 0. \tag{B12}$$

This is the distribution function for the one-parameter gamma distribution (1.3.17) denoted $Ga(k)$, and takes on values between 0 and 1. $I(k, x)$ is related to the distribution function for the χ^2 distribution with v degrees of freedom (denoted $\chi^2_{(v)}$) as follows:

$$F_v(x) = Pr(\chi^2_{(v)} \le x) \qquad x > 0$$
$$= \int_0^x \frac{z^{v/2-1}e^{-z}}{2^{v/2}\Gamma(v/2)}\,dz$$
$$= I\left(\frac{v}{2}, \frac{x}{2}\right). \tag{B13}$$

Many software packages provide values of $\Gamma(z)$, $\psi(z)$, (B12), and (B13).

B.3 ORDER STATISTICS

A few results about order statistics are given here. An extended treatment and references can be found in the book by Arnold et al. (1992).

Suppose that X has continuous probability density function (p.d.f.) $f(x)$ and distribution function $F(x)$ and that $X_1, \ldots, X_n$ is a random sample from this distribution. The X_i, rearranged in order of magnitude, and denoted

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)},$$

are called the order statistics of the sample. The joint p.d.f. of $X_{(\ell_1)}, \ldots, X_{(\ell_k)}$, where $1 \le \ell_1 < \ell_2 < \cdots < \ell_k \le n$ and $1 \le k \le n$, can be shown to be

$$\left(n!/\prod_{i=1}^{k+1}(\ell_i - \ell_{i-1} - 1)!\right)\prod_{i=1}^{k+1}\left[F(x_{(\ell_i)}) - F(x_{(\ell_{i-1})})\right]^{\ell_i - \ell_{i-1}-1}\prod_{i=1}^{k} f(x_{(\ell_i)}), \tag{B14}$$

where $x_{(\ell_1)} \le x_{(\ell_2)} \le \cdots \le x_{(\ell_k)}$ and where, for convenience, we define $\ell_0 = 0$, $\ell_{k+1} = n+1$, $x_{(\ell_0)} = -\infty$, and $x_{(\ell_{k+1})} = +\infty$.

Important special cases of (B14) are the following.

1. The joint p.d.f. of $X_{(1)}, \ldots, X_{(r)}$ $(r \le n)$ is

$$\frac{n!}{(n-r)!}\left(\prod_{i=1}^{r} f(x_{(i)})\right)\left[1 - F(x_{(r)})\right]^{n-r}. \tag{B15}$$

2. The p.d.f. of $X_{(i)}$ $(1 \le i \le n)$ is

$$\frac{n!}{(i-1)!(n-i)!} f(x_{(i)})F(x_{(i)})^{i-1}\left[1 - F(x_{(i)})\right]^{n-i}. \tag{B16}$$

3. The joint p.d.f. of $X_{(i)}$ and $X_{(j)}$, for $i < j$, is

$$\frac{n!}{(i-1)!(j-i-1)!(n-j)!} f(x_{(i)})f(x_{(j)})F(x_{(i)})^{i-1}$$
$$\times\left[F(x_{(j)}) - F(x_{(i)})\right]^{j-i-1}\left[1 - F(x_{(j)})\right]^{n-j}. \tag{B17}$$

These expressions and (B14) can be obtained directly. For example, (B17) can be found from the probability that of the n values $X_1, \ldots, X_n$, $i-1$ are less than $x_{(i)}$, one is in $(x_{(i)}, x_{(i)} + \Delta)$, $j - i - 1$ are between $x_{(i)} + \Delta$ and $x_{(j)}$, one is in $(x_{(j)}, x_{(j)} + \Delta)$, and $n - j$ are above $x_{(j)} + \Delta$.

Moments of order statistics are useful in some applications, though it is usually not possible to get simple analytical expressions. Two exceptions are for the uniform and exponential distributions, for which the following results are easily established.

1. For the uniform distribution $U(0, 1)$, with p.d.f. $f(x) = 1$ $(0 \le x \le 1)$,

$$E(X_{(i)}) = \frac{i}{n+1}, \qquad \mathrm{Var}(X_{(i)}) = \frac{i^2}{(n+1)^2(n+2)}. \tag{B18}$$

2. For the standard exponential distribution, with p.d.f. $f(x) = e^{-x}$ $(x \ge 0)$,

$$E(X_{(i)}) = \sum_{\ell=1}^{i}(n - \ell + 1)^{-1}, \qquad \mathrm{Var}(X_{(i)}) = \sum_{\ell=1}^{i}(n - \ell + 1)^{-2}. \tag{B19}$$

Several types of asymptotic results can be established for order statistics. Some involve the extreme order statistics $X_{(1)}$ and $X_{(n)}$ (e.g., Arnold et al. 1992). We mention only the case of $X_{(i)}$, where $i = np$ and $0 < p < 1$ as $n \to \infty$. The pth quantile of the distribution of X is $x_p = F^{-1}(p)$, and it can be shown that if $f(x) = F'(x)$ is continuous at x_p and $f(x_p) \ne 0$, then $\sqrt{n}(X_{(i)} - x_p)$ is asymptotically normal with mean 0 and variance

$$\frac{p(1-p)}{[f(x_p)]^2}. \tag{B20}$$

There is also a multivariate generalization of this.

APPENDIX C

Large-Sample Theory for Likelihood and Estimating Function Methods

C.1 MAXIMUM LIKELIHOOD LARGE-SAMPLE THEORY

Here we give a brief survey of important results for parametric likelihoods. Books on statistical theory (e.g., Cox and Hinkley 1974; Lehmann 1983), asymptotic theory (e.g., Barndorff-Nielsen and Cox 1994; Serfling 1980), and inference (e.g., Kalbfleisch 1985; Rao 1973; Severini 2000; Pawitan 2001) contain more details and references.

Suppose that the probability distribution of potentially observable data in a study is specified up to a parameter vector $\boldsymbol{\theta}$. Then, if Data represents the data actually observed in the study, the likelihood function for $\boldsymbol{\theta}$ based on these data is

$$L(\boldsymbol{\theta}) = Pr(\text{Data}; \boldsymbol{\theta}), \tag{C1}$$

where Pr represents the probability density or mass function from which the observed data are assumed to arise. We sometimes write $L(\boldsymbol{\theta}; \text{Data})$ instead of $L(\boldsymbol{\theta})$ to remind ourselves that L depends on the observed data. It is also customary to define $L(\boldsymbol{\theta})$ as any function that is proportional to the right side of (C1), since the methodology and theory described in the following are invariant to multiplication of $L(\boldsymbol{\theta})$ by values that are constant (i.e., do not depend on $\boldsymbol{\theta}$). More generally, a likelihood function may be based on the probability of observed data D_1, given the values of observed data D_2:

$$L(\boldsymbol{\theta}) \propto Pr(D_1|D_2; \boldsymbol{\theta}), \tag{C2}$$

where it is assumed that the probability mass or density on the right-hand side is specified up to the value of $\boldsymbol{\theta}$.

C.1.1 Asymptotic Results and Large-Sample Methods

We will outline standard maximum likelihood large-sample theory for the case in which the data consist of a random sample $y_1, \ldots, y_n$ from a distribution with probability density function (p.d.f.) $f(y; \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ is a vector of

unknown parameters taking on values in a set Ω. The y_i can be vectors, but for simplicity they will be written as scalars. The likelihood function for $\boldsymbol{\theta}$ is then

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\theta}). \tag{C3}$$

The case where the responses $y_1, \ldots, y_n$ are independent but not identically distributed, for example, because y_i depends on covariates, is of the form (C3) with $f_i(y_i; \boldsymbol{\theta})$ in place of $f(y_i; \boldsymbol{\theta})$. The following results apply to this setting as well. Let $\hat{\boldsymbol{\theta}}$ be a point in Ω at which $L(\boldsymbol{\theta})$ is maximized; $\hat{\boldsymbol{\theta}}$ is called a maximum likelihood estimate (m.l.e.) of $\boldsymbol{\theta}$. In most simple settings $\hat{\boldsymbol{\theta}}$ exists and is unique. Methods of obtaining $\hat{\boldsymbol{\theta}}$ are discussed in Appendix D. It is often convenient to work with $\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$, which is also maximized at $\hat{\boldsymbol{\theta}}$, and in most cases $\hat{\boldsymbol{\theta}}$ can be readily found by solving the so-called maximum likelihood equations $U_j(\boldsymbol{\theta}) = 0$ $(j = 1, \ldots, k)$, where

$$U_j(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_j} \qquad j = 1, \ldots, k. \tag{C4}$$

The $U_j(\boldsymbol{\theta})$'s are called scores or score functions, and the $k \times 1$ vector $\mathbf{U}(\boldsymbol{\theta}) = [U_1(\boldsymbol{\theta}), \ldots, U_k(\boldsymbol{\theta})]'$ is called the score vector.

The score vector is a sum of independent random variables, since $\ell(\boldsymbol{\theta}) = \sum \log f(y_i; \boldsymbol{\theta})$, and under mild "regularity" conditions (e.g., see Cox and Hinkley 1974, Sec. 9.2) it is asymptotically normally distributed. In addition $\mathbf{U}(\boldsymbol{\theta})$ has mean $\mathbf{0}$ and covariance matrix $\mathcal{I}(\boldsymbol{\theta})$, with entries

$$\mathcal{I}_{ij}(\boldsymbol{\theta}) = E\left(\frac{-\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right) \qquad i, j = 1, \ldots, k. \tag{C5}$$

The matrix $\mathcal{I}(\boldsymbol{\theta})$ is called the Fisher (or expected) information matrix.

Under mild regularity conditions, $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$, and n^{-1} times the observed information matrix $I(\hat{\boldsymbol{\theta}})$ is a consistent estimator of $\mathcal{I}(\boldsymbol{\theta})/n$. In addition, several other asymptotic results hold that lead to useful inference procedures.

C.1.1.1 Score Procedures

First, with a somewhat casual wording, $\mathbf{U}(\boldsymbol{\theta})$ is asymptotically $N_k[\mathbf{0}, \mathcal{I}(\boldsymbol{\theta})]$. (The strictly correct statement is that $\mathbf{BU}(\boldsymbol{\theta})$ is asymptotically standard k-variate normal, where $\mathbf{B}$ is a matrix such that $\mathbf{BB}' = \mathcal{I}(\boldsymbol{\theta})^{-1}$, but for convenience we shall use the casual wording.) This means that under the hypothesis H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

$$W(\boldsymbol{\theta}_0) = \mathbf{U}'(\boldsymbol{\theta}_0)\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{U}(\boldsymbol{\theta}_0) \tag{C6}$$

is asymptotically $\chi^2_{(k)}$. This can be used to test H_0 and to obtain confidence regions for $\boldsymbol{\theta}$ consisting of those $\boldsymbol{\theta}_0$ that make $W(\boldsymbol{\theta}_0)$ less than a specified value. Tests and estimates for a subset of the θ_j can also be obtained: suppose that $\boldsymbol{\theta}$ is partitioned as

$\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, where $\boldsymbol{\theta}_1$ is $p \times 1$ and $\boldsymbol{\theta}_2$ is $(k - p) \times 1$. Partition $\mathbf{U}(\boldsymbol{\theta})$, $\mathcal{I}(\boldsymbol{\theta})$, and $\mathcal{I}^{-1}(\boldsymbol{\theta})$ in a corresponding way:

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{U}_1(\boldsymbol{\theta}) \\ \mathbf{U}_2(\boldsymbol{\theta}) \end{pmatrix} \qquad \mathcal{I}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_{11}(\boldsymbol{\theta}) & \mathcal{I}_{12}(\boldsymbol{\theta}) \\ \mathcal{I}_{21}(\boldsymbol{\theta}) & \mathcal{I}_{22}(\boldsymbol{\theta}) \end{pmatrix} \tag{C7}$$

$$\mathcal{I}^{-1}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}^{11}(\boldsymbol{\theta}) & \mathcal{I}^{12}(\boldsymbol{\theta}) \\ \mathcal{I}^{21}(\boldsymbol{\theta}) & \mathcal{I}^{22}(\boldsymbol{\theta}) \end{pmatrix}. \tag{C8}$$

For a given value $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$, let $\tilde{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_{10})$ be the m.l.e. of $\boldsymbol{\theta}_2$, obtained by maximizing $L(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_2)$ with respect to $\boldsymbol{\theta}_2$. This gives what is termed the *profile or maximized likelihood* for $\boldsymbol{\theta}_{10}$, and the corresponding *profile log-likelihood function*

$$\begin{aligned} \ell_p(\boldsymbol{\theta}_{10}) &= \max_{\boldsymbol{\theta}_2} \ell(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_2) \\ &= \ell(\boldsymbol{\theta}_{10}, \tilde{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_{10})). \end{aligned} \tag{C9}$$

Denote $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_{10}, \tilde{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_{10}))$. Then under H_0: $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$,

$$\mathbf{U}_1(\tilde{\boldsymbol{\theta}})'\mathcal{I}^{11}(\tilde{\boldsymbol{\theta}})\mathbf{U}_1(\tilde{\boldsymbol{\theta}}) \tag{C10}$$

is asymptotically $\chi^2_{(p)}$. Tests based on (C6) are often referred to as score tests, and ones based on (C10) as partial score tests. A test based on (C6) has the convenient property of not requiring calculation of $\hat{\boldsymbol{\theta}}$.

C.1.1.2 MLE-Based Procedures

Tests and confidence intervals can also be based on the fact that $\hat{\boldsymbol{\theta}}$ is asymptotically $N_k[\boldsymbol{\theta}, \mathcal{I}^{-1}(\boldsymbol{\theta})]$ or, more accurately, that $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically $N_k[\mathbf{0}, n\mathcal{I}^{-1}(\boldsymbol{\theta})]$. Thus, under H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\mathcal{I}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \tag{C11}$$

is asymptotically $\chi^2_{(k)}$, and it can be used as a test statistic for H_0 or to generate confidence regions for $\boldsymbol{\theta}$. Since $I(\hat{\boldsymbol{\theta}})/n$ is a consistent estimator of $\mathcal{I}(\boldsymbol{\theta}_0)/n$, an asymptotically equivalent statistic to (C11) is

$$W(\boldsymbol{\theta}_0) = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'I(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0). \tag{C12}$$

This is much simpler to deal with computationally than (C11), and often there are also theoretical reasons to prefer it.

To consider tests or confidence intervals for a subset of the parameters, we partition $\boldsymbol{\theta}$, $\mathbf{U}(\boldsymbol{\theta})$, $\mathcal{I}(\boldsymbol{\theta})$, and $I(\boldsymbol{\theta})$ as in (C7) and (C8). Then under H_0: $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$,

$$(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10})'I^{11}(\hat{\boldsymbol{\theta}})^{-1}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_{10}) \tag{C13}$$

is asymptotically $\chi^2_{(p)}$. Asymptotically equivalent versions of (C13) can also be used; in particular, $I^{11}(\hat{\boldsymbol{\theta}})$ can be replaced by any of $I^{11}(\tilde{\boldsymbol{\theta}})$, $\mathcal{I}^{11}(\hat{\boldsymbol{\theta}})$, and $\mathcal{I}^{11}(\tilde{\boldsymbol{\theta}})$. Statistics like (C11), (C12), and (C13) that are based on $\hat{\boldsymbol{\theta}}$ are sometimes referred to as Wald statistics.

The estimate $\hat{\boldsymbol{\theta}}_1$ and its asymptotic covariance matrix can be obtained directly from the profile log-likelihood function (C9), provided that $\tilde{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_{10})$ is unique and the solution to $\partial \ell(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_2)/\partial \boldsymbol{\theta}_2 = \mathbf{0}$. In that case, it is easily shown under standard regularity conditions (e.g., Richards 1961; Seber and Wild 1989, Sec. 2.2.3) that $\hat{\boldsymbol{\theta}}_1$ satisfies $\partial \ell_p(\boldsymbol{\theta}_1)/\partial \boldsymbol{\theta}_1 = \mathbf{0}$, and that

$$I^{11}(\hat{\boldsymbol{\theta}}) = \left(\frac{-\partial^2 \ell_p(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1'} \right)^{-1}_{\boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}_1} \tag{C14}$$

is the estimated asymptotic covariance matrix for $\hat{\boldsymbol{\theta}}_1$. This is useful when $\tilde{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_1)$ has an algebraic closed form; if we are primarily interested in $\boldsymbol{\theta}_1$, then $\ell_p(\boldsymbol{\theta}_1)$ can effectively be used as a standard log-likelihood function.

C.1.1.3 Likelihood Ratio Procedures

A third approach for obtaining tests or confidence regions is via likelihood ratio statistics. Under H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ the statistic

$$\Lambda(\boldsymbol{\theta}_0) = -2 \log \left[\frac{L(\boldsymbol{\theta}_0)}{L(\hat{\boldsymbol{\theta}})} \right] = 2\ell(\hat{\boldsymbol{\theta}}) - 2\ell(\boldsymbol{\theta}_0) \tag{C15}$$

is asymptotically $\chi^2_{(k)}$. Similarly, if $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, then under the hypothesis H_0: $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$ the statistic

$$\Lambda(\boldsymbol{\theta}_{10}) = 2\ell(\hat{\boldsymbol{\theta}}) - 2\ell(\tilde{\boldsymbol{\theta}}) \tag{C16}$$

is asymptotically $\chi^2_{(p)}$, where $\boldsymbol{\theta}_1$ is $p \times 1$ and $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_{10}, \tilde{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_{10}))$. Note that (C16) can also be expressed as

$$\Lambda(\boldsymbol{\theta}_{10}) = 2\ell_p(\hat{\boldsymbol{\theta}}_1) - 2\ell_p(\boldsymbol{\theta}_{10}), \tag{C17}$$

where $\ell_p(\boldsymbol{\theta}_{10})$ is the profile log-likelihood function (C9).

C.1.1.4 Functions of Parameters

If one is interested in certain functions of the parameters $\boldsymbol{\theta}$, say

$$\psi_i = g_i(\boldsymbol{\theta}) \qquad i = 1, \ldots, p,$$

then inferences about $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_p)'$ can be obtained by merely associating $\boldsymbol{\psi}$ with $\boldsymbol{\theta}_1$ in the preceding discussion. However, methods based directly on the m.l.e.

$\hat{\boldsymbol{\psi}} = (g_1(\hat{\boldsymbol{\theta}}), \ldots, g_p(\hat{\boldsymbol{\theta}}))'$ are easily implemented through the fact that the asymptotic covariance matrix for $\hat{\boldsymbol{\psi}}$ is, by Theorem B2 in Appendix B, estimated consistently by

$$\hat{V}_{\boldsymbol{\psi}} = G(\hat{\boldsymbol{\theta}}) \hat{V}_{\boldsymbol{\theta}} G(\hat{\boldsymbol{\theta}})', \tag{C18}$$

where $\hat{V}_{\boldsymbol{\theta}} = I(\hat{\boldsymbol{\theta}})^{-1}$ is the estimated asymptotic covariance matrix for $\hat{\boldsymbol{\theta}}$ and $G(\boldsymbol{\theta})$ is the $p \times k$ matrix with entries

$$G(\boldsymbol{\theta})_{ij} = \partial g_i(\boldsymbol{\theta})/\partial \theta_j \qquad i = 1, \ldots, p; \qquad j = 1, \ldots, k.$$

Then, for example, it follows that under H_0: $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, the statistic

$$(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)' \hat{V}_{\boldsymbol{\psi}}^{-1} (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0) \tag{C19}$$

is asymptotically $\chi^2_{(p)}$.

C.1.1.5 Tests and Confidence Intervals

Tests based on asymptotically χ^2 statistics such as (C11), (C15), and (C19) provide evidence against the hypothesized parameter values when the observed statistics are large. For the statistic $W(\boldsymbol{\theta}_0)$ of (C11), for example, the p-value (significance level) based on the observed value $w(\boldsymbol{\theta}_0)$ is approximately $Pr(\chi^2_{(k)} \geq w(\boldsymbol{\theta}_0))$. A confidence region for $\boldsymbol{\theta}$ with approximate confidence coefficient α consists of vectors $\boldsymbol{\theta}_0$ satisfying

$$w(\boldsymbol{\theta}_0) \leq \chi^2_{(k),\alpha}. \tag{C20}$$

Confidence intervals for a single parametric function $\psi = g(\boldsymbol{\theta})$ are often wanted. The simplest approach is to use the normal approximation $\hat{\psi} \sim N(\psi, \hat{V}_\psi)$, where $\hat{V}_\psi$ is given by (C18). This yields the approximate standard normal pivotal quantity

$$Z = \frac{\hat{\psi} - \psi}{\hat{V}_\psi^{1/2}} \tag{C21}$$

and two-sided approximate $1 - \alpha$ confidence intervals $\hat{\psi} \pm z_{\alpha/2} \hat{V}_\psi^{1/2}$, where z_q is the qth quantile for $N(0, 1)$. One-sided approximate $1 - \alpha$ confidence intervals are given by $\psi \leq \hat{\psi} - z_\alpha \hat{V}_\psi^{1/2}$ and by $\psi \geq \hat{\psi} + z_\alpha \hat{V}_\psi^{1/2}$, respectively.

Confidence intervals can also be based on the likelihood ratio statistic for testing H_0: $\psi = \psi_0$. By (C16) this is

$$\Lambda(\psi_0) = 2\ell(\hat{\boldsymbol{\theta}}) - 2\ell(\tilde{\boldsymbol{\theta}}(\psi_0)), \tag{C22}$$

where $\tilde{\boldsymbol{\theta}}(\psi_0)$ maximizes $\ell(\boldsymbol{\theta})$ subject to the constraint $g(\boldsymbol{\theta}) = \psi_0$. If $\psi = \psi_0$, then $\Lambda(\psi_0)$ is asymptotically $\chi^2_{(1)}$, and so a two-sided approximate $1 - \alpha$ confidence

interval for ψ consists of the values ψ_0, satisfying

$$\Lambda(\psi_0) \leq \chi^2_{(1),1-\alpha}. \tag{C23}$$

One-sided confidence intervals can be obtained from the *signed square-root likelihood ratio statistic*

$$S(\psi_0) = \text{sign}(\hat{\psi} - \psi_0)\Lambda(\psi_0)^{1/2}. \tag{C24}$$

If $\psi = \psi_0$, then $S(\psi_0)$ is asymptotically $N(0, 1)$. The use of $S(\psi_0)$ as a pivotal leads to the same two-sided confidence intervals as (C23). It also gives one-sided approximate $1 - \alpha$ confidence intervals through the solution of either $S(\psi_0) \geq z_\alpha$ or $S(\psi_0) \leq -z_\alpha$. The corresponding intervals consist of the values ψ_0 satisfying

$$I(\psi_0 > \hat{\psi})\Lambda(\psi_0) \leq \chi^2_{(1),1-2\alpha} \tag{C25}$$

and

$$I(\psi_0 < \hat{\psi})\Lambda(\psi_0) \leq \chi^2_{(1),1-2\alpha}, \tag{C26}$$

respectively, so that (C26) gives a lower $1 - \alpha$ confidence limit and (C25) an upper $1 - \alpha$ confidence limit.

The adequacy of the asymptotic χ^2 or normal approximations used for the procedures described in this Appendix can vary substantially according to the problem and the amount of information about the parameters. It is difficult to make general statements, but the distributions of likelihood ratio statistics often tend to their limiting distributions more quickly than Wald statistics or pivotals such as (C11), (C19), and (C21). Consequently, likelihood ratio methods are often preferred, especially for small to moderate sample sizes. However, confidence intervals based on these methods may require substantial computation and are not directly available from most software packages, so the Wald procedures are heavily used. The accuracy of these procedures can be improved substantially through suitable transformations, as discussed by Anscombe (1964) and Sprott (1973). In particular, if one is interested in the parameter $\psi = g(\boldsymbol{\theta})$, and if Z in (C21) is not close to standard normal for a given sample size, then we should seek a one-to-one function $\phi = h(\psi)$ such that

$$Z_1 = \frac{\hat{\phi} - \phi}{\hat{V}_\phi^{1/2}}$$

is close to $N(0, 1)$; note that $\hat{V}_\phi = h'(\hat{\psi})^2\hat{V}_\psi$ by (B4) of Appendix B.

A final point is that some confidence interval procedures, and in particular ones based on likelihood ratio statistics, may give close to nominal coverage for two-sided intervals based on small samples, but be considerably less accurate for one-sided intervals.

A large literature exists on higher-order asymptotics for maximum likelihood (see, e.g., Barndorff-Nielsen and Cox 1994; Reid 2000). This work leads to a variety of adjustments or improved approximations for likelihood-based test statistics and approximate pivotals. For example, DiCiccio et al. (2001) consider signed squared-root likelihood ratio procedures. In addition, the adequacy of large-sample methods for certain types of problems has been assessed through simulation. It is not feasible to discuss this here, but some remarks are given in different sections of the book.

Although the preceding discussion was focused on independent observations $Y_1, \ldots, Y_n$ from distributions $f(y; \boldsymbol{\theta})$, the general results and methodology apply much more broadly. In particular, they apply to problems involving censoring and other forms of incomplete data. The main requirement for the establishment of standard asymptotic properties is that the Fisher information matrix $\mathcal{I}(\boldsymbol{\theta})$ increases at a sufficiently fast rate. It is generally sufficient when the data arise from n independent individuals or units that $n^{-1}\mathcal{I}(\boldsymbol{\theta})$ approach a positive-definite limit as n becomes arbitrarily large.

Problems for which the parameter $\boldsymbol{\theta}$ is on the boundary of Ω are more complex (e.g., Self and Liang 1987; Crowder 1990), as are problems involving threshold parameters. The latter are considered in Section 4.5.

Nonparametric maximum likelihood asymptotics are technically much more difficult, though many practical problems can be handled by considering sequences of parametric models. For general discussion and references, see, for example, Bickel et al. (1992), Andersen et al. (1993), and Murphy and van der Vaart (1999). The book by Owen (2001) discusses empirical likelihood.

C.1.2 Marginal, Conditional, and Partial Likelihoods

Sometimes, especially when there are nuisance parameters present, it is appropriate or convenient to use only a portion of the observed data in forming a likelihood function. This leads to the concepts of marginal, conditional, and partial likelihoods.

Suppose that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and that the data, denoted simply as $\mathbf{y}$, are transformed into two parts S, T. If the distribution of S depends on $\boldsymbol{\theta}_1$, but not $\boldsymbol{\theta}_2$, then the likelihood $L_S(\boldsymbol{\theta}_1)$ obtained from the marginal distribution of S is termed a *marginal likelihood*. Similarly, suppose that the conditional distribution of T, given S, depends on $\boldsymbol{\theta}_1$, but not $\boldsymbol{\theta}_2$. In this case, the likelihood $L_{T|S}(\boldsymbol{\theta}_1)$ obtained from this conditional distribution is referred to as a *conditional likelihood*. Marginal and conditional likelihoods have been discussed by several authors (e.g., Fraser 1968; Kalbfleisch and Sprott 1970, 1974); Reid (2000) provides additional references. A main problem is in assessing whether there is a substantial loss of information entailed in using a marginal or conditional likelihood in a given situation. If there is not, these provide convenient methods of making inferences about $\boldsymbol{\theta}_1$ in the absence of knowledge of $\boldsymbol{\theta}_2$. Marginal and conditional likelihoods are based on the probability of observed outcomes so, under appropriate regularity conditions, the usual asymptotic properties of maximum likelihood hold for data sequences in which the Fisher information increases sufficiently fast.

Marginal and conditional likelihood are special cases of the more general concept of partial likelihood (Cox, 1975). The basic idea behind this is as follows: suppose that the data $\mathbf{y}$ have p.d.f. $f(\mathbf{y}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, where $\boldsymbol{\theta}_1$ is of interest and $\boldsymbol{\theta}_2$ is a nuisance parameter. Suppose that $\mathbf{y}$ can be transformed into parts $S_1, T_1, S_2, T_2, \ldots, S_m, T_m$. The joint p.d.f. of $S_1, T_1, \ldots, S_m, T_m$ can be written as

$$\prod_{i=1}^{m} f_{S_i|s^{(i-1)},t^{(i-1)}}(s_i|s^{(i-1)}, t^{(i-1)}; \boldsymbol{\theta}) \prod_{i=1}^{m} f_{t_i|s^{(i)},t^{(i-1)}}(t_i|s^{(i)}, t^{(i-1)}; \boldsymbol{\theta}),$$

where $s^{(i)} = (s_1, \ldots, s_i)$ and $t^{(i)} = (t_1, \ldots, t_i)$.

If the second term depends just on $\boldsymbol{\theta}_1$, this is termed a partial likelihood for $\boldsymbol{\theta}_1$. There will typically be some loss of information involved in using a partial likelihood, though this may be difficult to assess. In addition, to use partial likelihoods we need to know that asymptotic results of the kind described for ordinary likelihoods hold. Cox (1975) outlines conditions under which this would be so, and Wong (1986) provides further discussion. For many applications involving lifetime data and event history processes, asymptotic properties of partial likelihood have been rigorously examined (e.g., Andersen et al, 1993).

C.2 ESTIMATING FUNCTION ASYMPTOTICS

Consider data consisting of independent responses y_i $(i = 1, \ldots, n)$ and associated covariate vectors $\mathbf{x}_i$ $(i = 1, \ldots, n)$, and suppose that the distribution of Y_i given $\mathbf{x}_i$ depends on a $k \times 1$ parameter vector $\boldsymbol{\theta}$. An *estimating function* for $\boldsymbol{\theta}$ is a $k \times 1$ vector $\mathbf{U}(\boldsymbol{\theta}) = \sum \mathbf{U}_i(\boldsymbol{\theta})$ of real-valued functions of y_i, $\mathbf{x}_i$, and $\boldsymbol{\theta}$, with components

$$U_r(\boldsymbol{\theta}) = \sum_{i=1}^{n} U_{ri}(y_i, \mathbf{x}_i; \boldsymbol{\theta}) \qquad r = 1, \ldots, k. \tag{C27}$$

For observed data $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$ the objective is to use the estimating function to obtain an estimate $\hat{\boldsymbol{\theta}}$ by solving the *estimating equations* $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$. Estimating equations (see Godambe 1991; Heyde 1997) provide a unification of different approaches to estimation, including maximum likelihood as well as methods that do not require the assumption of a fully specified distribution for Y_i given $\mathbf{x}_i$.

Suitably defined estimating functions yield consistent, asymptotically normal estimators $\hat{\boldsymbol{\theta}}$ and are easy to use for hypothesis tests or interval estimation. Estimating functions of many types have been considered (e.g., Huber 1967; Crowder 1987; Fahrmeir and Kaufmann 1987; Godambe and Thompson 1989; Prentice and Zhao 1991; Heyde 1997). We outline key asymptotic results in the general framework of White (1982) and Inagaki (1973).

We restrict consideration to *unbiased estimating functions* for which $E[\mathbf{U}_i(\boldsymbol{\theta})] = \mathbf{0}$, the expectation being taken with respect to the distribution of Y_i given $\mathbf{x}_i$ at the parameter value $\boldsymbol{\theta}$. Results similar to those following can also be obtained for many estimating functions that are only asymptotically unbiased.

Under suitable regularity conditions (e.g., White 1982; Crowder 1986) consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}$ as $n \to \infty$ can be established. Crucial ingredients in the results are the $k \times k$ matrices

$$A_n(\boldsymbol{\theta}) = -\frac{1}{n}\left(\frac{\partial \mathbf{U}}{\partial \boldsymbol{\theta}'}\right), \qquad B_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{U}_i(\boldsymbol{\theta})\mathbf{U}_i(\boldsymbol{\theta})'. \tag{C28}$$

In addition, define

$$A(\boldsymbol{\theta}) = \lim_{n\to\infty} E\{A_n(\boldsymbol{\theta})\}, \qquad B(\boldsymbol{\theta}) = \lim_{n\to\infty} E\{B_n(\boldsymbol{\theta})\} = \lim_{n\to\infty} \frac{1}{n}\mathrm{Var}\{\mathbf{U}(\boldsymbol{\theta})\}.$$

Under conditions that ensure that $n^{-1/2}\mathbf{U}(\boldsymbol{\theta})$ is asymptotically normal, nonsingularity of $A(\boldsymbol{\theta})$ and $B(\boldsymbol{\theta})$, and other regularity conditions, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normal with covariance matrix

$$C(\boldsymbol{\theta}) = A(\boldsymbol{\theta})^{-1}B(\boldsymbol{\theta})[A(\boldsymbol{\theta})^{-1}]'. \tag{C29}$$

A consistent estimator of $C(\boldsymbol{\theta})$ is given by the sandwich form

$$C_n(\hat{\boldsymbol{\theta}}) = A_n(\hat{\boldsymbol{\theta}})^{-1}B_n(\hat{\boldsymbol{\theta}})[A_n(\hat{\boldsymbol{\theta}})^{-1}]'. \tag{C30}$$

An alternative is to replace one or both of $A_n(\hat{\boldsymbol{\theta}})$ and $B_n(\hat{\boldsymbol{\theta}})$ with $E\{A_n(\boldsymbol{\theta})\}$ and $E\{B_n(\boldsymbol{\theta})\}$ evaluated at $\hat{\boldsymbol{\theta}}$, assuming that these expectations are tractable.

In the case of maximum likelihood estimation, $\mathbf{U}(\boldsymbol{\theta}) = \partial \log L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and it can be shown that $A(\boldsymbol{\theta}) = B(\boldsymbol{\theta}) = n^{-1}\mathcal{I}(\boldsymbol{\theta})$, the scaled Fisher information matrix. This gives $C(\boldsymbol{\theta}) = n\mathcal{I}(\boldsymbol{\theta})^{-1}$, in agreement with maximum likelihood large-sample theory.

White (1982) extends the results concerning estimating functions (C27) so that they also apply to misspecified models. Suppose that the components $U_{ri}(Y_i, X_i; \boldsymbol{\theta})$ of (C27) are independent and identically distributed (i.i.d.) for any fixed $\boldsymbol{\theta}$, with (Y_i, X_i) having distribution G. Suppose also that there exists a unique point $\boldsymbol{\theta}^*$ such that $E_G\{U_{ri}(Y_i, X_i; \theta^*)\} = 0$ for $r = 1, \ldots, k$. Then under suitable regularity conditions, the solution $\hat{\boldsymbol{\theta}}$ to $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$ is asymptotically normal. In particular, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ is asymptotically normal with mean vector $\mathbf{0}$ and covariance matrix $C(\boldsymbol{\theta}^*)$ given by (C29), with the expectation defining $A(\boldsymbol{\theta})$ and $B(\boldsymbol{\theta})$ taken with respect to G. Note that in this framework, the parameter $\boldsymbol{\theta}$ and estimating functions (C27) define an estimator $\hat{\boldsymbol{\theta}}$, but they do not necessarily bear any relationship to the "true" distribution G.

APPENDIX D

Computational Methods and Simulation

D.1 OPTIMIZATION METHODS FOR MAXIMUM LIKELIHOOD

Statistical software for lifetime data is widely available, but there are many problems that this specialized software does not handle. In this case one can turn to numerical methods software, some of which is designed to handle generic maximum likelihood problems.

The need is for procedures that can be used to maximize multiparameter log-likelihood functions $\ell(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$ lying in a parameter space Ω. Extensive accounts of optimization (maximization or minimization) methods and software are given in books such as Gill et al. (1981), Fletcher (1987), and Press et al. (1986). Lange (1999) surveys numerical methods for a wide range of statistical applications. Some general remarks are provided here, followed by comments on specific software sources.

Many likelihood functions have a unique maximum at $\hat{\boldsymbol{\theta}}$, which is a stationary point satisfying $\partial\ell/\partial\boldsymbol{\theta} = \mathbf{0}$. Numerical methods of locating such points involve an initial value $\boldsymbol{\theta}_0$ and an iterative procedure designed to give a sequence of points $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots$ converging to $\hat{\boldsymbol{\theta}}$. A distinguishing feature of different methods is in their use of the first-derivative vector $\mathbf{U}(\boldsymbol{\theta}) = \partial\ell/\partial\boldsymbol{\theta}$ and the second derivative (or Hessian) matrix $H(\boldsymbol{\theta}) = \partial^2\ell/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ for $\ell(\boldsymbol{\theta})$. Approaches to maximization include:

1. Search procedures that do not use any derivatives. The simplex algorithm (Nelder and Mead 1965) is a familiar example.
2. Methods that use first derivatives $\mathbf{U}(\boldsymbol{\theta})$. Examples include steepest ascent, quasi-Newton, and conjugate gradient methods.
3. Methods that use $H(\boldsymbol{\theta})$. Newton–Raphson iteration is a familiar example; it uses the iteration scheme

$$\boldsymbol{\theta}_j = \boldsymbol{\theta}_{j-1} - H(\boldsymbol{\theta}_{j-1})^{-1}\mathbf{U}(\boldsymbol{\theta}_{j-1}) \qquad j = 1, 2, \ldots. \tag{D.1}$$

This is derived by approximating $\mathbf{U}(\boldsymbol{\theta})$ after the $(j-1)$st iteration by the linear function $U(\boldsymbol{\theta}_{j-1}) + H(\boldsymbol{\theta}_{j-1})(\boldsymbol{\theta} - \boldsymbol{\theta}_{j-1})$, then noting that the right side of (D.1) makes this equal to $\mathbf{0}$.

The choice of method for a specific function $\ell(\boldsymbol{\theta})$ depends on its shape and the ease with which first and second derivatives can be computed. Good optimization software includes numerical methods for closely approximating derivatives, so even for approaches of types (2) and (3), it is not essential to provide algebraic expressions for $\mathbf{U}(\boldsymbol{\theta})$ or $H(\boldsymbol{\theta})$. Since we generally want the observed information matrix $I(\hat{\boldsymbol{\theta}}) = -H(\hat{\boldsymbol{\theta}})$ along with $\hat{\boldsymbol{\theta}}$, optimization software that produces this is important.

Some likelihood functions achieve their maximum on the boundary of Ω. Finding the maximum likelihood estimate (m.l.e.) may then be more challenging, but constrained search techniques are usually effective. Likelihood functions can also possess multiple stationary points. Optimization methods are usually designed to seek a local maximum, so may not necessarily converge to the global maximum $\hat{\boldsymbol{\theta}}$. From a statistical perspective it is important to understand the shape of $\ell(\boldsymbol{\theta})$. Features such as nonconvexity, multiple stationary points, or maxima on the boundary have important implications for estimation. They can, for example, indicate that there is little information about certain parameters or that confidence interval estimation based on $\hat{\boldsymbol{\theta}}$ and associated variance estimates is unwarranted or inaccurate.

Some other practical considerations should also be noted.

1. To be confident that a global maximum $\hat{\boldsymbol{\theta}}$ has been obtained, it is necessary to understand the shape of $\ell(\boldsymbol{\theta})$. Plots of $\ell(\boldsymbol{\theta})$ for models with only one or two parameters are recommended. For models with more parameters, profile plots in which all but two of the parameters are held at fixed values are useful.
2. Parameterizations that make $\ell(\boldsymbol{\theta})$ closer to quadratic tend to both improve the convergence of iterative procedures and the accuracy of large-sample methods for testing and interval estimation. Range restrictions on parameters can also be removed through reparameterization.
3. It can sometimes be helpful to break a problem into parts by writing $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and using different methods for the maximization of $\ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ with respect to $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. For example, if $\ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is roughly quadratic with $\boldsymbol{\theta}_2$ fixed, but not when $\boldsymbol{\theta}_2$ varies, a good procedure when $\boldsymbol{\theta}_2$ is of dimension one or two is to obtain the profile log-likelihood $\ell(\tilde{\boldsymbol{\theta}}_1(\boldsymbol{\theta}_2), \boldsymbol{\theta}_2)$ for $\boldsymbol{\theta}_2$ and to plot it. The approximate location of $\hat{\boldsymbol{\theta}}_2$ is then apparent from inspection.
4. Convergence criteria for iterative procedures are based on two successive iterations giving suitably small values for some combination of $|\boldsymbol{\theta}_j - \boldsymbol{\theta}_{j-1}|$, $|\ell(\boldsymbol{\theta}_j) - \ell(\boldsymbol{\theta}_{j-1})|$ or the gradient vector $\mathbf{U}(\boldsymbol{\theta}_j)$. Some likelihoods possess ridges along which $\ell(\boldsymbol{\theta})$ changes very slowly, so it is prudent to be aware of the specific criteria used. Plots of the log-likelihood are helpful in such cases.
5. Optimization software usually has default initial values, but can accept input values. This can be useful if the software fails to locate a maximum on its own. In problems where the shape of $\ell(\boldsymbol{\theta})$ is not well understood, it is a good idea to see if different initial values lead to the same $\hat{\boldsymbol{\theta}}$.

Many software packages for numerical or statistical computation possess good optimization procedures. If the software also has built-in functions for the probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of a family of parametric models under consideration, then likelihood and log-likelihood functions based on the various data structures in this book (e.g., see (2.2.3), (2.3.1), (2.4.2)) are very easy to code. The SAS procedure NLP was used for many examples in this book; it can produce Hessian matrices $H(\hat{\boldsymbol{\theta}})$ by numerical differentiation.

Critchley et al. (1988), Venzon and Moolgavkar (1988), Seber and Wild (1989), and others discuss the computation of m.l.e.'s under constraints on the parameters. These ideas can be useful when computing profile likelihoods or likelihood ratio statistics.

D.2 SIMULATION AND BOOTSTRAP METHODS

The distributions of test statistics, estimators or pivotal quantities can often be approximated closely by simulation. The following is an example.

Example D.2.1. Let $Y_1, \ldots, Y_n$ be a random sample from a specific location-scale parameter distribution (1.3.18) or (E1), with m.l.e.'s $\hat{u}$ and $\hat{b}$. As shown in Appendix E,

$$Z = (\hat{u} - u)/\hat{b} \tag{D2}$$

is a pivotal quantity. Since its distribution does not depend on u or b, we can estimate it by simulating samples from the distribution with $u = 0$, $b = 1$, as follows.

1. Generate a pseudorandom sample $y_1^*, \ldots, y_n^*$ from the fully known distribution (E1) with $u = 0$, $b = 1$. Obtain the m.l.e.'s $\hat{u}^*$, $\hat{b}^*$ based on this sample, and thus the value
$$z^* = \hat{u}^*/\hat{b}^*. \tag{D3}$$
2. Repeat this process B times, yielding values $z_1^*, \ldots, z_B^*$.
3. The z_j^* provide an estimate of the distribution of Z. In particular, the empirical c.d.f. and sample quantiles from $z_1^*, \ldots, z_B^*$ estimate the c.d.f. and quantiles of Z.

The approach of the preceding example can be used whenever there is a parametric model and some variable (e.g., a pivotal quantity or test statistic) whose distribution does not depend on unknown parameter values. By using a large enough value of B we can effectively approximate the distribution as closely as desired. Bootstrap methodology deals with similar uses of simulation, but in a more general context where arbitrarily close approximation of a distribution is not necessarily possible. Numerous books describe bootstrap theory and methods (e.g., Efron and Tibshirani 1993; Hall 1992; Davison and Hinkley 1997). A short outline of some methods for interval estimation is presented here.

D.2.1 Parametric Bootstrap

Consider independent and identically distributed (i.i.d.) data from a parametric model $Y \sim F(y; \boldsymbol{\theta})$. Confidence intervals for a specified parameter ψ are ideally based on pivotal quantities. However, for most models pivotal quantities and confidence intervals with exact prescribed coverage probabilities α do not exist. Instead, confidence intervals are based on quantities $W = g(y_1, \ldots, y_n; \psi)$, which are asymptotically pivotal: the limiting distribution of W as $n \to \infty$ does not depend on $\boldsymbol{\theta}$. In that case, confidence intervals obtained by inverting probability statements for W based on its limiting distribution typically have coverage probability $\alpha + C(\boldsymbol{\theta})n^{-1/2}$. The approximate pivotals most often used are those from maximum likelihood large-sample theory: the approximate $N(0, 1)$ pivotal (C21),

$$W = (\hat{\psi} - \psi)/se(\hat{\psi}) \tag{D4}$$

and the likelihood ratio statistic (C22) or its signed square root (C24).

The basic *parametric bootstrap* method for obtaining confidence intervals from an asymptotic pivotal quantity $W = g(y_1, \ldots, y_n; \psi)$ is as follows. Let $\hat{\boldsymbol{\theta}}$ be the m.l.e. of $\boldsymbol{\theta}$ in the assumed model $F(y; \boldsymbol{\theta})$, based on $y_1, \ldots, y_n$. Then carry out the following simulation:

1. Generate a pseudorandom bootstrap sample $y_1^*, \ldots, y_n^*$ from $F(y; \hat{\boldsymbol{\theta}})$. Obtain the value

$$w^* = g(y_1^*, \ldots, y_n^*; \hat{\psi}), \tag{D5}$$

   where $\hat{\psi}$ is the m.l.e. of ψ based on $\hat{\boldsymbol{\theta}}$.
2. Repeat this process B times, yielding values $w_1^*, \ldots, w_B^*$. The distribution of W can be estimated from $w_1^*, \ldots, w_B^*$ as in Example D2.1. The qth quantile for W is estimated by $w_{(qB)}^*$, where we assume for simplicity that qB is an integer, so that $w_{(qB)}^*$ is the (qB)-th smallest value among the w_j^*. Then for $q_2 > q_1$,

$$Pr(w_{(q_1B)}^* \le W \le w_{(q_2B)}^*) \doteq q_2 - q_1, \tag{D6}$$

   and this can be inverted to give an approximate $q_2 - q_1$ confidence interval for ψ.

Example D.2.2. Suppose W in (D4) is used. Then each bootstrap sample $y_1^*, \ldots, y_n^*$ gives an estimate $\hat{\psi}^*$ and standard error $se(\hat{\psi}^*)$ for ψ, and the value

$$w^* = (\hat{\psi}^* - \hat{\psi})/se(\hat{\psi}^*).$$

The probability statement (D6) then gives the confidence interval

$$\hat{\psi} - w_{(q_2B)}^* se(\hat{\psi}) \le \psi \le \hat{\psi} - w_{(q_1B)}^* se(\hat{\psi}). \tag{D7}$$

The method in Example D2.2 is often called the *studentized bootstrap* or the *bootstrap-t* method. However, the same procedure can be used with other approximate pivotal quantities, for example, the likelihood ratio statistic $\Lambda(\psi)$ given by (C22). The quantities W of (D4) and $\Lambda(\psi)$ have limiting $N(0, 1)$ and $\chi_{(1)}^2$ distributions, respectively, and for large n the bootstrap estimates based on $w_1^*, \ldots, w_B^*$ will reflect this. The bootstrap approximation is usually better for smaller values of n, though a large value of B may be needed when q_1 is close to zero or q_2 is close to one. It is also important for the accuracy of confidence intervals that the quantity W on which they are based be as close to pivotal as possible. With quantities of the form (D4) it is advisable to use a parameterization that facilitates this. The approach here is essentially the same as that described in Appendix C or Section 4.1.1. Parameterizations or transformations of ψ that stabilize the variance of $\hat{\psi}$ or reduce the skewness of $\Lambda(\psi)$ can improve confidence interval accuracy substantially. Procedures based on (D4), unlike those based on $\Lambda(\psi)$, are not parameterization-invariant.

Around $B = 2000$ bootstrap samples are often suggested for good accuracy, and 5000 or more are often suggested when $q_2 - q_1$ is close to one. However, large values of B cannot overcome inaccuracy due to W's distribution depending on $\boldsymbol{\theta}$.

The parametric bootstrap can also be used to check the adequacy of large-sample normal or χ^2 approximations for pivotal quantities through probability plots of the w_j^*, or comparison of estimated probabilities or quantiles with their normal or χ^2 counterparts.

D.2.2 Nonparametric Bootstrap

In some applications it is not possible to generate bootstrap samples and values (D5) from a parametric model. Sometimes it is simply difficult to simulate data from the model, but more often, the process generating the data is not fully specified. For example, censored samples $y_i = (t_i, \delta_i)$, $i = 1, \ldots, n$ often arise under a random censoring process that is not known. The *nonparametric bootstrap* replaces step (1) in the algorithm giving (D5) with

1′. Generate a bootstrap sample $y_1^*, \ldots, y_n^*$ by randomly drawing n items, with replacement, from $y_1, \ldots, y_n$.

The rest of the bootstrap procedure is as before. Under certain conditions (see Hall 1992; Efron and Tibshirani 1993; Davison and Hinkley 1997), this process provides asymptotically accurate approximations to distributions of variables such as W in (D4) or the corresponding likelihood ratio statistic $\Lambda(\psi)$.

Other nonparametric bootstrap procedures exist, but the preceding one is easy to use and generalizes to other problems involving independent data elements $d_1, \ldots, d_n$. For example, if y_i has an associated covariate vector $\mathbf{x}_i$ and is subject to right censoring, then bootstrap samples can be obtained by selecting samples of size n, with replacement, from $d_i = (y_i, \delta_i, \mathbf{x}_i)$, $i = 1, \ldots, n$.

Example D.2.3. Consider nonparametric estimation of a survivor function $S(t)$ on the basis of a censored random sample (t_i, δ_i), $i = 1, \ldots, n$, as discussed in Section 3.2.3. The approximate pivotal quantity (3.2.16) can be used to obtain confidence intervals for $S(t)$. We generate individual bootstrap samples by selecting n items (t_i^*, δ_i^*), $i = 1, \ldots, n$, with replacement from $\{(t_i, \delta_i), i = 1, \ldots, n\}$. Each sample gives a Kaplan–Meier estimate $\hat{S}^*(t)$ and associated standard error $\hat{\sigma}_s^*(t)$, from which the value

$$Z_1^* = \frac{\hat{S}^*(t) - \hat{S}(t)}{\hat{\sigma}_s^*(t)}$$

is obtained, where $\hat{S}(t)$ is the Kaplan–Meier estimate from the data (t_i, δ_i), $i = 1, \ldots, n$. Generation of B bootstrap samples and corresponding values Z_1^* provides an empirical estimate of the distribution of (3.2.16), and of its q_1 and q_2 quantiles. This gives a $q_2 - q_1$ confidence interval for $S(t)$, exactly as in (D6) and (D7). Strawderman and Wells (1997) provide extensive discussion of this and related methods.

D.2.3 Additional Remarks

A few more points about bootstrap methodology are worth noting.

1. The bootstrap works for smooth functions of the data, and is most useful for problems where maximum likelihood large-sample theory is not easily applied.
2. A variety of methods designed to improve the accuracy of bootstrap approximations has been developed, with names such as the bias-corrected accelerated (BC_a) method and the ABC method (Davison and Hinkley 1997, Ch. 5).
3. In some problems it is conventional to make inferences conditional on certain aspects of the data. In particular, with covariates it is customary to condition on the observed values $\mathbf{x}_i$ $(i = 1, \ldots, n)$. Parametric bootstrap methods do this, but the nonparametric method described here does not; it considers observations $(y_i, \mathbf{x}_i)$ as random. Nonparametric methods that condition on the $\mathbf{x}_i$ have also been proposed. In practice there is often relatively little difference between bootstrap confidence intervals developed under fixed $\mathbf{x}$ and random $\mathbf{x}$ frameworks.

APPENDIX E

Inference in Location-Scale Parameter Models

A univariate location-scale parameter distribution is one with probability density function (p.d.f.) of the form

$$f(y; u, b) = \frac{1}{b} g\left(\frac{y-u}{b}\right) \qquad -\infty < y < \infty, \tag{E1}$$

where u $(-\infty < u < \infty)$ is a location parameter, b $(b > 0)$ is a scale parameter, and $g(\,)$ is a fully specified p.d.f. defined on $(-\infty, \infty)$. The survivor function (s.f.) corresponding to (E1) is $G[(y-u)/b]$, where

$$G(x) = \int_x^\infty g(z)\, dz.$$

Tests and interval estimation for u and b, and for parameters in related regression models, are outlined here. The ideas involved actually apply to a wider class of models (e.g., Hora and Buehler 1966; Fraser 1968), but this is not needed for this book.

E.1 EQUIVARIANT STATISTICS

We allow Type 2 censoring and suppose that $y_1 \le \cdots \le y_r$ are the r smallest observations in a random sample of size n from (E1). The results here also apply when there is progressive Type 2 censored sampling, but for simplicity we shall not examine this explicitly. Suppose that $\tilde{u} = \tilde{u}(y_1, \ldots, y_r)$ and $\tilde{b} = \tilde{b}(y_1, \ldots, y_r)$ are statistics with the following properties:

$$\tilde{u}(dy_1 + c, \ldots, dy_r + c) = d\tilde{u}(y_1, \ldots, y_r) + c \tag{E2}$$

$$\tilde{b}(dy_1 + c, \ldots, dy_r + c) = d\tilde{b}(y_1, \ldots, y_r) \tag{E3}$$

for any real constants c $(-\infty < c < \infty)$ and d $(d > 0)$. Then $\tilde{u}$ and $\tilde{b}$ are termed *equivariant statistics*; they are also commonly referred to as equivariant estimators

of u and b. The requirements (E2) and (E3) are natural ones for estimators of location and scale parameters and most, if not all, of the common types of estimators satisfy them.

THEOREM E1. Let $\hat{u}$ and $\hat{b}$ be maximum likelihood estimators (m.l.e.'s) of u and b in (E1), based on a Type 2 censored sample. Then $\hat{u}$ and $\hat{b}$ are equivariant.

Proof. The likelihood function based on $y_1 \le \cdots \le y_r$ is (see (2.2.6))

$$L_{\mathbf{y}}(u, b) = \frac{1}{b^r}\left[\prod_{i=1}^{r} g\left(\frac{y_i - u}{b}\right)\right]\left[G\left(\frac{y_r - u}{b}\right)\right]^{n-r}, \tag{E4}$$

where we write $\mathbf{y} = (y_1, \ldots, y_r)$. It is easily seen that

$$L_{\mathbf{y}}(u, b) = d^r L_{\mathbf{y}'}(u', b'),$$

where $y_i' = dy_i + c$, $\mathbf{y}' = (y_1', \ldots, y_r')$, $u' = du + c$, and $b' = db$. If $L_{\mathbf{y}}(u, b)$ is maximized for u and b at $\hat{u}(\mathbf{y})$ and $\hat{b}(\mathbf{y})$, then $L_{\mathbf{y}'}(u, b)$ is maximized at $\hat{u}(\mathbf{y}') = d\hat{u}(\mathbf{y}) + c$ and $\hat{b}(\mathbf{y}') = d\hat{b}(\mathbf{y})$. This proves the result. □

E.2 PIVOTALS AND ANCILLARIES

The next theorem follows easily from the definition of equivariant estimators.

THEOREM E2. Let $\tilde{u}$ and $\tilde{b}$ be equivariant estimators, based on a Type 2 censored sample from (E1). Then

(i) $Z_1 = (\tilde{u} - u)/\tilde{b}$, $Z_2 = \tilde{b}/b$, and $Z_3 = (\tilde{u} - u)/b$ are pivotal quantities.
(ii) The quantities $a_i = (y_i - \tilde{u})/\tilde{b}$, $i = 1, \ldots, r$, form a set of ancillary statistics, only $r - 2$ of which are functionally independent.

Proof. From (E1) the random variable $W = (Y - u)/b$ has p.d.f. $g(w)$, $-\infty < w < \infty$, not depending on u or b. Thus the joint p.d.f. of $w_1 = (y_1 - u)/b, \ldots, w_r = (y_r - u)/b$ does not depend on u or b, and consequently neither does the distribution of $\tilde{u}(w_1, \ldots, w_r)$ or $\tilde{b}(w_1, \ldots, w_r)$. But since $\tilde{u}$ and $\tilde{b}$ are equivariant, it follows from (E2) and (E3) that

$$\tilde{u}(w_1, \ldots, w_r) = \frac{\tilde{u}(y_1, \ldots, y_r) - u}{b} = Z_3$$

$$\tilde{b}(w_1, \ldots, w_r) = \frac{\tilde{b}(y_1, \ldots, y_r)}{b} = Z_2.$$

This proves that Z_2 and Z_3 are pivotal; Z_1 is thus also pivotal, since $Z_1 = Z_3/Z_2$.

Regarding (ii), the a_i are clearly ancillary, since $a_i = (y_i - \tilde{u})/\tilde{b} = (w_i - Z_3)/Z_2$, and hence their distribution does not depend on u or b. Finally, the a_i satisfy two restrictions, since

$$\tilde{u}(a_1, \ldots, a_r) = [\tilde{u}(y_1, \ldots, y_r) - \tilde{u}]/\tilde{b} = 0$$

and

$$\tilde{b}(a_1, \ldots, a_r) = \tilde{b}(y_1, \ldots, y_r)/\tilde{b} = 1.$$

Thus there are just $r - 2$ functionally independent a_i. □

The following theorem concerning the distribution of the pivotals and ancillaries also provides an alternate proof of Theorem E2.

THEOREM E3. Let $\tilde{u}$ and $\tilde{b}$ be any equivariant estimators of u and b under the conditions of Theorem E2. Then the joint p.d.f. of Z_1, Z_2, $a_1, \ldots, a_{r-2}$ is of the form

$$k(\mathbf{a}, r, n) z_2^{r-1}\left(\prod_{i=1}^{r} g(a_i z_2 + z_1 z_2)\right)[G(a_r z_2 + z_1 z_2)]^{n-r}, \tag{E5}$$

where $k(\mathbf{a}, r, n)$ is a function of $a_1, \ldots, a_{r-2}$, r, and n only. The conditional p.d.f. of Z_1 and Z_2, given $\mathbf{a} = (a_1, \ldots, a_r)$, is also of the form (E5).

Proof. The joint p.d.f. of $y_1, \ldots, y_r$ is

$$\frac{n!}{(n-r)!} b^{-r}\left[\prod_{i=1}^{r} g\left(\frac{y_i - u}{b}\right)\right]\left[G\left(\frac{y_r - u}{b}\right)\right]^{n-r}.$$

Make the change of variables from $(y_1, \ldots, y_r)$ to $(\tilde{u}, \tilde{b}, a_1, \ldots, a_{r-2})$; this transformation can be written as

$$y_i = \tilde{b} a_i + \tilde{u} \qquad i = 1, \ldots, r,$$

where we note that a_r and a_{r-1} can be expressed in terms of $a_1, \ldots, a_{r-2}$. The Jacobian $\partial(y_1, \ldots, y_r)/\partial(\tilde{u}, \tilde{b}, a_1, \ldots, a_{r-2})$ is of the form $\tilde{b}^{r-2} k'(\mathbf{a}, r, n)$, where k' is a rather complicated function. The joint p.d.f. of $\tilde{u}, \tilde{b}, a_1, \ldots, a_{r-2}$ is therefore

$$\frac{n!}{(n-r)!} k'(\mathbf{a}, r, n) \frac{\tilde{b}^{r-2}}{b^r}\left[\prod_{i=1}^{r} g\left(a_i \frac{\tilde{b}}{b} + \frac{\tilde{u} - u}{b}\right)\right]\left[G\left(a_r \frac{\tilde{b}}{b} + \frac{\tilde{u} - u}{b}\right)\right]^{n-r}.$$

Making the further change of variables from $(\tilde{u}, \tilde{b}, a_1, \ldots, a_{r-2})$ to $(z_1, z_2, a_1, \ldots, a_{r-2})$, we get (E5). Finally, the conditional p.d.f. of Z_1 and Z_2, given $\mathbf{a}$, is of the same form (though the function $k(\mathbf{a}, r, n)$ is different), since dividing (E5) by the joint p.d.f. of $a_1, \ldots, a_{r-2}$ gives a new function of the same form. □

E.3 CONFIDENCE INTERVALS

The pivotals Z_1 and Z_2 can be used to obtain confidence intervals for u and b, respectively. The pth quantile of (E1) is $y_p = u + w_p b$, where w_p satisfies $G(w_p) = 1 - p$. Confidence intervals for y_p can be based on the pivotal $Z_p = [w_p b - (\tilde{u} - u)]/\tilde{b} = w_p Z_2^{-1} - Z_1$. This can also be used to get confidence intervals for the s.f. $G[(y - u)/b]$.

There are two main approaches to the construction of confidence intervals. One is to base intervals on the unconditional distributions of pivotals. For example, the probability statement $Pr(\ell_1 \le Z_1 \le \ell_2) = \gamma$ gives $(\tilde{u} - \ell_2\tilde{b}, \tilde{u} - \ell_1\tilde{b})$ as a γ confidence interval for u. Any equivariant estimators $\tilde{u}$ and $\tilde{b}$ can be used, but the properties of the intervals will depend on the estimators. A second approach, first suggested by Fisher (1934), is to base confidence intervals on the conditional distributions of Z_1, Z_2, and Z_p, given the observed value of the ancillary statistic $\mathbf{a}$. For example, the probability statement $Pr(\ell_1 \le Z_1 \le \ell_2|\mathbf{a}) = \gamma$ gives the confidence interval $(\tilde{u} - \ell_2\tilde{b}, \tilde{u} - \ell_1\tilde{b})$ for u. Note that ℓ_1 and ℓ_2 are functions of $\mathbf{a}$ in this case.

A great deal has been written about the two approaches (e.g., see Lawless 1978, for some comments and references). From a purely logical point of view the conditional approach seems more appropriate. For distributions other than the normal the unconditional distributions of pivotals are impossible to obtain analytically; note that to do this involves integrating $a_1, \ldots, a_{r-2}$ out of (E5). The conditional distributions are computationally tractable, however, since $k(\mathbf{a}, r, n)$ in (E5) is a normalizing constant that can be obtained by two-dimensional integration (numerical, if necessary) from the fact that for any $\mathbf{a}$

$$\int_0^\infty \int_{-\infty}^\infty h(z_1, z_2|\mathbf{a})\, dz_1\, dz_2 = 1$$

where $h(z_1, z_2|\mathbf{a})$ is (E5), the joint p.d.f. of Z_1 and Z_2, given $\mathbf{a}$. Fraser (1979) and Lawless (1978) discuss the computation of probabilities for (E5) in specific situations. Approximations are also available for conditional probabilities (e.g., DiCiccio 1988), and some higher-order adjustments to likelihood methods apply to conditional methods (e.g., see Barndorff-Nielsen and Cox 1994, Ch. 6).

The following properties are also of interest.

Property 1. Confidence intervals obtained by the conditional approach are unconditional confidence intervals in the usual sense. We will demonstrate this property for confidence intervals for b. Suppose $[d_1(\mathbf{y}), d_2(\mathbf{y})]$ is a confidence interval for b, obtained from the conditional probability statement $Pr[\ell_1 \le \tilde{b}/b \le \ell_2|\mathbf{a}] = \gamma$. That is, $d_1(\mathbf{y}) = \tilde{b}/\ell_2$ and $d_2(\mathbf{y}) = \tilde{b}/\ell_1$; note that ℓ_1 and ℓ_2 are functions of $\mathbf{a}$. Then

$$\begin{aligned} Pr[d_1(\mathbf{y}) \le b \le d_2(\mathbf{y})] &= E_{\mathbf{a}}\{Pr[d_1(\mathbf{y}) \le b \le d_2(\mathbf{y})|\mathbf{a}]\} \\ &= E_{\mathbf{a}}(\gamma) \\ &= \gamma. \end{aligned}$$

Property 2. Level γ confidence intervals for u, b, or x_p constructed by the conditional method are numerically equivalent to level γ Bayes posterior probability intervals, obtained with the improper prior distribution $b^{-1}du\,db$ for u and b. This is readily established; see, for example, Hora and Buehler (1966) and Lawless (1973).

Property 3. Different equivariant estimators lead to the same confidence intervals with the conditional approach. This follows from Property 2, since any conditional confidence limit produced is numerically equivalent to a unique Bayes posterior probability limit. This property does not hold for the unconditional method, where properties of confidence intervals depend on the estimators used to form the pivotals.

In spite of these points it is the unconditional approach that is typically used for inference in location-scale parameter models. The main reasons are (1) the fact that conditional methods do not apply except for Type 2 censored or complete data, (2) the need for higher dimensional numerical integration when the conditional approach is used with regression models, (3) a lack of special-purpose software for important models, (4) the fact that unconditional intervals can be obtained easily using simulation, plus the existence of some tables generated this way, and (5) the easy application of unconditional large-sample likelihood methods.

Except for very small samples there is usually little difference between conditional and unconditional confidence intervals or tests. Moreover, higher-order adjustments that are sometimes made to maximum likelihood large-sample methods may have either conditional or unconditional interpretations. In practice, it generally does not matter much which approach is taken.

E.4 REGRESSION MODELS

These results generalize in a straightforward way for linear regression models (e.g., Verhagen 1961; Fraser 1979). Suppose that Y_i is a response variable, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik})'$ is a covariate sector, and the p.d.f of Y_i given $\mathbf{x}_i$ is of the form

$$f(y_i|\mathbf{x}_i; \boldsymbol{\beta}, b) = \frac{1}{b} g\left(\frac{y_i - \boldsymbol{\beta}'\mathbf{x}_i}{b}\right) \qquad -\infty < y_i < \infty, \tag{E6}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ is a vector of unknown regression coefficients. Suppose that a random sample $y_1, \ldots, y_n$ is taken from (E6), corresponding to fixed covariate vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Let $\mathbf{X}$ be the $n \times k$ matrix with rows $\mathbf{x}_1', \ldots, \mathbf{x}_n'$ and let $\mathbf{y} = (y_1, \ldots, y_n)'$; estimators $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\mathbf{y})$ and $\tilde{b} = \tilde{b}(\mathbf{y})$ of $\boldsymbol{\beta}$ and b are now called equivariant if for any real vector $\mathbf{c} = (c_1, \ldots, c_k)'$ and scalar d $(d > 0)$

$$\tilde{\boldsymbol{\beta}}(d\mathbf{y} + \mathbf{X}\mathbf{c}) = d\tilde{\boldsymbol{\beta}}(\mathbf{y}) + \mathbf{c} \tag{E7}$$

$$\tilde{b}(d\mathbf{y} + \mathbf{X}\mathbf{c}) = d\tilde{b}(\mathbf{y}). \tag{E8}$$

The least squares and m.l.e. are readily shown to be equivariant (e.g., Verhagen 1961).

Theorem E2 generalizes for uncensored samples under the regression model (E6) to give the following theorem.

THEOREM E4. Let $\tilde{\boldsymbol{\beta}}$ and $\tilde{b}$ be any equivariant estimators of $\boldsymbol{\beta}$ and b, based on a random sample as defined earlier. Then (1) $\mathbf{Z}_1 = (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})/\tilde{b}$ and $Z_2 = \tilde{b}/b$ are pivotals, and (2) the quantities $a_i = (y_i - \tilde{\boldsymbol{\beta}}'\mathbf{x}_i)/\tilde{b}$, $i = 1, \ldots, n$ are ancillaries, only $n - k$ of which are functionally independent.

In addition, the final part of Theorem E3 generalizes to Theorem E5.

THEOREM E5. Under the same conditions as in Theorem E4, the joint p.d.f. of $\mathbf{Z}_1$, Z_2, and $\mathbf{a} = (a_1, \ldots, a_n)$ is of the form

$$k(\mathbf{a}, \mathbf{X}, n) z_2^{n-1} \prod_{i=1}^{n} g(a_i z_2 + \mathbf{x}_i' \mathbf{z}_1 z_2). \tag{E9}$$

The conditional p.d.f. of $\mathbf{Z}_1$ and Z_2, given $\mathbf{a}$, is of the same form.

As in the ordinary location-scale case, inferences about $\boldsymbol{\beta}$ and b can be based on the unconditional distribution of $\mathbf{Z}_1$ and Z_2 or on the conditional distribution of $\mathbf{Z}_1$ and Z_2, given $\mathbf{a}$. Except when (E6) is a normal model, the conditional approach requires two- or higher-dimensional numerical integration.

APPENDIX F

Martingales and Counting Processes

Martingales and counting processes are basic tools in the mathematical analysis of survival and other stochastic processes. Comprehensive treatments are provided by Fleming and Harrington (1991) and Andersen et al. (1993). This Appendix presents some basic concepts and states a few useful results for survival analysis.

F.1 MARTINGALES

Let $M = \{M(t), t \geq 0\}$ be a stochastic process such that $E\{|M(t)|\}$ exists for $t \geq 0$, and let $\{\mathcal{H}(t)\}$ denote an increasing sequence of sigma fields such that M is $\mathcal{H}(t)$-measurable. That is, $\mathcal{H}(s) \subset \mathcal{H}(t)$ for $s < t$ and $\{\mathcal{H}(t)\}$ includes the possible sample paths of M over $[0, t]$. It may also include outcomes for other factors such as covariate or censoring processes. The process M is said to be a *martingale* with respect to $\{\mathcal{H}(t)\}$ if

$$E\{M(t)|\mathcal{H}(s)\} = M(s) \qquad s \leq t. \tag{F1}$$

To accommodate processes that may have discontinuities, we assume that $M(t)$ is right-continuous with left-hand limits. We define *martingale increments*

$$dM(t) = M[(t + dt)-] - M(t-). \tag{F2}$$

It follows from (F1) that for any $s < t$,

$$\begin{aligned} E\{dM(t)|\mathcal{H}(s)\} &= 0 \\ E\{dM(s)dM(t)\} &= 0. \end{aligned} \tag{F3}$$

In what follows, processes are all assumed to be $\mathcal{H}(t)$-measurable for some $\{\mathcal{H}(t)\}$. A process $\{A(t), t \geq 0\}$ is called *predictable* if it is $\mathcal{H}(t-)$-measurable. Left-continuous processes, for which $A(t) = A(t-)$, are predictable. The *Doob–Meyer decomposition* theorem establishes that a process $\{X(t), t \geq 0\}$ that is $\mathcal{H}(t)$-

measurable can be expressed in the form

$$X(t) = A(t) + M(t), \tag{F4}$$

where $A(t)$ is a predictable process and $M(t)$ is a martingale. The $A(t)$ process is called the *compensator* for X.

The *predictable variation process* of M, denoted $\langle M \rangle$, is defined as the compensator for the process $\{M(t)^2, t \geq 0\}$. It is of interest because it has increments related to variance,

$$\begin{aligned} d\langle M \rangle(t) &= E\{dM(t)^2 | \mathcal{H}(t-)\} \\ &= \text{Var}\{dM(t) | \mathcal{H}(t-)\}. \end{aligned} \tag{F5}$$

Similarly, if M_1 and M_2 are two martingales, then their *predictable covariation process* $\langle M_1, M_2 \rangle$ is the compensator of $\{M_1(t)M_2(t), t \geq 0\}$. This has increments

$$\begin{aligned} d\langle M_1, M_2 \rangle(t) &= E\{dM_1(t)\, dM_2(t) | \mathcal{H}(t-)\} \\ &= \text{Cov}\{dM_1(t), dM_2(t) | \mathcal{H}(t-)\}. \end{aligned} \tag{F6}$$

In obtaining (F5) and (F6), the fact that $E\{dM(t)|\mathcal{H}(t-)\} = 0$ is used; see (F3). Note also that $\langle M, M \rangle = \langle M \rangle$.

Finally, we introduce *stochastic integrals* of the form

$$Z_\ell(t) = \int_0^t W_\ell(s)\, dM_\ell(s) \qquad \ell = 1, 2, \ldots, \tag{F7}$$

where $\{W_\ell(t), t \geq 0\}$ is a bounded predictable process and $\{M_\ell(t), t \geq 0\}$ is a martingale. The integrals (F7) are written in Riemann–Stieltjes form; see (1.2.9). It is easily seen that $Z_\ell = \{Z_\ell(t), t \geq 0\}$ is also a martingale with respect to $\{\mathcal{H}(t)\}$ and that the predictable covariation processes have increments

$$\begin{aligned} d\langle Z_\ell, Z_k \rangle(t) &= \text{Cov}\{Z_\ell(t), Z_k(t) | \mathcal{H}(t-)\} \\ &= W_\ell(t) W_k(t) d\langle M_\ell, M_k \rangle(t) \end{aligned} \tag{F8}$$

for either $\ell \neq k$ or $\ell = k$.

F.2 COUNTING PROCESSES AND SURVIVAL MODELS

A *counting process* is a right-continuous stochastic process that counts the number of events of some type over the time interval $[0, t]$. The heuristic discussion here considers only survival processes. In this case a process counts the deaths (either 0 or 1) for each individual: $N_i(t) = I(T_i \leq t)$, where T_i is the lifetime for individual i. In describing data subject to right censoring the counting process for observed failures

is used,

$$N_i(t) = I(T_i \leq t, \delta_i = 1). \tag{F9}$$

The *intensity function* for a counting process specifies the probability an event occurs in $[t, t + dt)$, given the process history $\mathcal{H}(t-)$. In the survival setting this is given by the hazard function. Using the definition (F9), we write

$$\begin{aligned} Pr\{dN_i(t) = 1 | \mathcal{H}(t-)\} &= d\Lambda_i(t) \\ &= E\{dN_i(t) | \mathcal{H}(t-)\}, \end{aligned} \tag{F10}$$

where $\{N_i(t), t \geq 0\}$ is $\mathcal{H}(t-)$-measurable. We refer to $\Lambda_i(t) = \int_0^t d\Lambda_i(s)$ as the *cumulative intensity process*. In the case of continuous lifetime distributions $d\Lambda_i(t) = \lambda_i(t)\, dt = Y_i(t) h_i(t)\, dt$, where $h_i(t)$ is the continuous hazard function for T_i and $Y_i(t) = I$ ($T_i \geq t$ and censoring has not occurred by t). In general we write

$$d\Lambda_i(t) = Y_i(t)\, dH_i(t) \tag{F11}$$

and assume that $Y_i(t)$ is left-continuous and $H_i(t)$ is right-continuous and predictable. This requires that any covariate processes involved in the specification of $H_i(t)$ are predictable.

Under the preceding conditions,

$$M_i(t) = N_i(t) - \Lambda_i(t), \qquad t \geq 0 \tag{F12}$$

is a martingale with respect to $\{\mathcal{H}(t)\}$ and can be used to study properties of many inference procedures. A crucial factor in doing this is the ability to express relevant processes as stochastic integrals for which the predictable covariation processes can be obtained. In particular, consider two processes of the form

$$Z_\ell(t) = \sum_{i=1}^n \int_0^t W_{i\ell}(s)\, dM_i(s), \qquad \ell = 1, 2, \tag{F13}$$

where $M_i(t)$ is given by (F12) and (F11), and the $W_{i\ell}$ processes are all bounded and predictable. It can then be seen (e.g., Fleming and Harrington 1991, Sec. 2.4—2.6) that Z_1 and Z_2 are martingales with respect to $\{\mathcal{H}(t)\}$ and, assuming that the $dN_i(t)$'s for $i = 1, \ldots, n$ are independent, given $\mathcal{H}(t-)$,

$$E\{Z_\ell(t)\} = 0 \tag{F14}$$

$$\begin{aligned} E\{Z_\ell(t)^2\} &= \text{Var}\{Z_\ell(t)\} \\ &= E \sum_{i=1}^n \int_0^t W_{i\ell}^2(s) Y_i(s) [1 - \Delta H_i(s)]\, dH_i(s) \end{aligned} \tag{F15}$$

$$E\{Z_1(t)Z_2(t)\} = \mathrm{Cov}\{Z_1(t), Z_2(t)\}$$
$$= E\sum_{i=1}^{n}\int_0^t W_{i1}(s)W_{i2}(s)Y_i(s)[1-\Delta H_i(s)]\,dH_i(s), \qquad \text{(F16)}$$

where the notation $\Delta X(s) = X(s) - X(s-)$ for right-continuous processes is introduced. That is, $\Delta X(s) = 0$, unless $X(s)$ is discontinuous at s.

We conclude with brief examples of the application of these ideas.

Example F1. Maximum Likelihood Score Functions. The likelihood score function $\mathbf{U}(\boldsymbol{\theta})$ for a parametric lifetime distribution with continuous hazard function $h(t;\boldsymbol{\theta})$, based on a censored random sample, is given by (2.2.18) as

$$\mathbf{U}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\int_0^\infty \frac{\partial \log h_i(t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\,dM_i(t),$$

where $\boldsymbol{\theta}$ is $p \times 1$,

$$dM_i(t) = dN_i(t) - Y_i(t)h_i(t;\boldsymbol{\theta})\,dt$$

and $dN_i(t) = I(T_i = t, \delta_i = 1)$. The processes $\{M_i(t), t \geq 0\}$ are martingales and the functions

$$W_{i\ell}(t) = \frac{\partial \log h_i(t;\boldsymbol{\theta})}{\partial \theta_\ell} \qquad \ell = 1, \ldots, p$$

are predictable. It follows immediately from the results (F14)–(F16) and the fact that $\Delta H_i(t;\boldsymbol{\theta}) = 0$ that $E\{\mathbf{U}(\boldsymbol{\theta})\} = \mathbf{0}$ and for $\ell, j = 1, \ldots, p$,

$$\mathrm{Cov}\{U_\ell(\boldsymbol{\theta}), U_j(\boldsymbol{\theta})\} = E\sum_{i=1}^{n}\int_0^t \frac{\partial \log h_i(t;\boldsymbol{\theta})}{\partial \theta_\ell}\frac{\partial \log h_i(t;\boldsymbol{\theta})}{\partial \theta_j}Y_i(t)h_i(t;\boldsymbol{\theta})\,dt.$$

Some algebra shows the equivalence of this to the expectation of the information matrix $I(\boldsymbol{\theta}) = -\partial \mathbf{U}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'$, and martingale asymptotics can be applied to prove standard asymptotic likelihood results (e.g., Andersen et al. 1993, Ch. 6).

Example F2. The Nelson–Aalen Estimator. The Nelson–Aalen estimator $\hat{H}(t)$ of (3.2.12) is represented as a stochastic integral with respect to the martingale $M_i(t) = N_i(t) - H_i(t)$ in (3.2.27) of Section 3.2.4. An application of (F14) then gives (3.2.29), from which a variance estimate for $\hat{H}(t)$ can be obtained.

APPENDIX G

Data Sets

G.1 TRANSFUSION-RELATED AIDS DATA

These data are given by Kalbfleisch and Lawless (1989). Here we give the subset of the data discussed in Examples 3.5.3 and 4.3.3, which is for persons aged 5–59. The following values are as given by Kalbfleisch and Lawless:

INF = month of human immunodeficiency virus (HIV) infection (month 1 is January 1978)

DIAG = duration of the Acquired Immune Deficiency Syndrome (AIDS) induction (incubation) period, in months

AGE = age + 1(in years), where age is for the patient at the time of transfusion

One individual who had an infection later than 90 months has been dropped from the data set, leaving 124 individuals.

In the analyses of Examples 3.5.3 and 4.3.3, the induction time t is defined as DIAG, and the right truncation time v equals 102.5 − INF.

G.2 TIMES TO FIRST PULMONARY EXACERBATION

These data were kindly provided by Terry Therneau, and were discussed in Examples 1.1.8, 3.2.4, 3.2.5, 6.2.3, 6.3.4, 7.2.1, and 11.3.1. A larger data set that also gives the times of second and subsequent exacerbations is available by visiting the Mayo Clinic Web site http://www.mayo.edu/hsr/biostat.html; see also Appendix D of Therneau and Grambsch (2000). See under "rhDNase for Cystic Fibrosis" for the data set.

The data discussed in the examples listed in the previous paragraph use the first forced expiratory volume (fev) measurement as variable fev, and only times to a first exacerbation.

INF	DIAG	AGE	INF	DIAG	AGE	INF	DIAG	AGE
38	15	56	47	35	23	83	12	55
27	28	57	64	18	35	84	11	55
23	34	20	15	68	6	59	36	53
25	34	46	35	48	26	65	31	51
42	17	46	67	16	59	47	49	51
33	29	53	35	48	51	81	15	36
45	17	39	36	47	22	87	9	57
33	29	54	74	10	42	85	11	59
34	29	34	15	69	30	72	24	44
26	38	56	29	55	54	53	43	46
3	61	29	62	22	38	54	43	47
53	12	46	69	16	34	17	80	54
28	38	46	62	23	54	61	36	28
34	32	26	82	4	46	48	49	60
21	46	30	75	11	59	29	68	60
37	30	25	27	59	60	30	68	29
33	34	51	56	30	26	40	58	28
17	53	33	73	13	54	75	23	41
58	13	39	76	10	50	43	55	59
49	22	57	50	36	46	86	12	36
67	4	29	49	38	59	57	41	11
35	37	57	61	26	42	53	46	37
12	60	21	59	29	52	48	51	54
19	53	52	57	31	41	60	39	55
53	20	56	68	20	60	36	63	57
36	38	56	76	12	29	48	51	27
12	62	58	19	70	17	55	44	51
50	24	52	22	67	59	60	39	35
37	37	34	26	63	48	82	18	60
34	41	32	57	33	51	71	29	44
57	18	23	58	32	56	37	63	59
28	48	42	59	32	32	60	40	49
66	10	32	41	50	46	78	22	41
45	32	60	70	21	53	52	48	56
53	24	44	58	33	33	37	64	49
54	25	58	75	17	52	22	79	33
40	39	50	39	53	37	12	89	38
54	25	50	83	10	38	47	54	53
74	5	39	29	65	53	85	16	38
52	29	24	65	29	58	84	17	49
62	19	58	41	53	58			
67	14	45	58	36	21			

G.3 BREAKING STRENGTHS OF CARBON FIBERS

The data, from Crowder (2000), are used in Example 6.4.2 and give the breaking strengths of single carbon fibers of different lengths.

Length (ℓ)	Breaking Strength (t)							
1	2.247	2.640	2.842	2.908	3.099	3.126	3.245	3.328
	3.355	3.383	3.572	3.581	3.681	3.726	3.727	3.728
	3.783	3.785	3.786	3.896	3.912	3.964	4.050	4.063
	4.082	4.111	4.118	4.141	4.216	4.251	4.262	4.326
	4.402	4.457	4.466	4.519	4.542	4.555	4.614	4.632
	4.634	4.636	4.678	4.698	4.738	4.832	4.924	5.043
	5.099	5.134	5.359	5.473	5.571	5.684	5.721	5.998
	6.060							
10	1.901	2.132	2.203	2.228	2.257	2.350	2.361	2.396
	2.397	2.445	2.454	2.454	2.474	2.518	2.522	2.525
	2.532	2.575	2.614	2.616	2.618	2.624	2.659	2.675
	2.738	2.740	2.856	2.917	2.928	2.937	2.937	2.977
	2.996	3.030	3.125	3.139	3.145	3.220	3.223	3.235
	3.243	3.264	3.272	3.294	3.332	3.346	3.377	3.408
	3.435	3.493	3.501	3.537	3.554	3.562	3.628	3.852
	3.871	3.886	3.971	4.024	4.027	4.225	4.395	5.020
20	1.312	1.314	1.479	1.552	1.700	1.803	1.861	1.865
	1.944	1.958	1.966	1.997	2.006	2.021	2.027	2.055
	2.063	2.098	2.140	2.179	2.224	2.240	2.253	2.270
	2.272	2.274	2.301	2.301	2.339	2.359	2.382	2.382
	2.426	2.434	2.435	2.478	2.490	2.511	2.514	2.535
	2.554	2.566	2.570	2.586	2.629	2.633	2.642	2.648
	2.684	2.697	2.726	2.770	2.773	2.800	2.809	2.818
	2.821	2.848	2.880	2.954	3.012	3.067	3.084	3.090
	3.096	3.128	3.233	3.433	3.585	3.585		
50	1.339	1.434	1.549	1.574	1.589	1.613	1.746	1.753
	1.764	1.807	1.812	1.840	1.852	1.852	1.862	1.864
	1.931	1.952	1.974	2.019	2.051	2.055	2.058	2.088
	2.125	2.162	2.171	2.172	2.180	2.194	2.211	2.270
	2.272	2.280	2.299	2.308	2.335	2.349	2.356	2.386
	2.390	2.410	2.430	2.431	2.458	2.471	2.497	2.514
	2.558	2.577	2.593	2.601	2.604	2.620	2.633	2.670
	2.682	2.699	2.705	2.735	2.785	2.785	3.020	3.042
	3.116	3.174						

G.4 LIFETIMES OF STEEL SPECIMENS

The data, from Crowder (2000), give the lifetimes of steel specimens tested at 14 different stress levels. The data are used in Example 6.4.3.

Stress (s)	Lifetime (t)									
38.5	60	51	83	140	109	106	119	76	68	67
	111	57	69	75	122	128	95	87	82	132
38.0	100	90	59	80	128	117	177	98	158	107
	125	118	99	186	66	132	97	87	69	109
37.5	199	105	147	113	98	118	182	131	156	78
	84	103	89	124	71	65	220	109	93	171
37.0	141	143	98	122	110	132	194	155	104	83
	125	165	146	100	318	136	200	201	251	111
36.5	118	273	192	238	105	398	108	182	130	170
	181	119	152	199	89	211	324	164	133	121
36.0	173	218	162	288	394	585	295	262	127	151
	181	209	141	186	309	192	117	203	198	255
35.5	156	173	125	852	559	442	168	286	261	227
	285	253	166	133	309	247	112	202	365	702
35.0	230	169	178	271	129	568	115	280	305	326
	1101	285	734	177	493	218	342	431	143	381
34.5	155	397	1063	738	140	364	218	461	174	326
	504	374	321	169	426	248	350	348	265	293
34.0	168	397	385	1585	224	987	358	763	610	532
	449	498	714	159	326	291	425	146	246	253
33.5	154	305	957	1854	363	457	415	559	767	210
	678	332	180	1274	528	254	835	611	482	593
33.0	184	241	273	1842	371	830	683	1306	562	166
	981	1867	493	418	2978	1463	2220	312	251	760
32.5	4257	879	799	1388	271	308	2073	227	347	669
	1154	393	250	196	548	475	1705	2211	975	2925
32.0	1144	231	523	474	4510	3107	815	6297	1580	605
	1786	206	1943	935	283	1336	727	370	1056	413
	619	2214	1826	597						

G.5 VA LUNG CANCER SURVIVAL DATA

These data give survival times for 137 advanced lung cancer patients, and are discussed in Examples 1.1.9, 3.2.4, 3.2.5, 6.2.3, 6.3.4, 7.2.1 and 11.3.1. Part of the data set (the data for the 40 patients who received therapy prior to the study) is given in Example 1.1.9.

The full data set is available in the file "veteran" on the Statlib Web site (http://lib.stat.cmu.edu/datasets/). The data are also given in the matrix cancer.vet in some versions of S-Plus.

G.6 SURVIVAL TIMES FOR PERSONS WITH LUPUS NEPHRITIS

These data, discussed by Abrahamowicz et al. (1996), are used in Problem 7.10. The data include the variables:

TIME = survival time (in months) from an initial renal biopsy

STATUS = failure indicator (0 = alive at end of follow-up, 1 = dead)

DURATION = duration (in months) of untreated kidney disease prior to the biopsy

for 87 lupus patients who underwent a renal biopsy between 1967 and 1983 and were followed up until death or the end of 1990.

The data set is available in the file "lupus" on the Statlib Web site (http://lib.stat.cmu.edu/datasets/).

G.7 PRIMARY BILIARY CIRRHOSIS (PBC) DATA

These data are discussed in Example 8.2.1. The full data set is given in Appendix D.1 in Fleming and Harrington (1991). The data set is also available in the file "pbc" on the Statlib Web site (http://lib.stat.cmu.edu/datasets/).

G.8 DIABETIC RETINOPATHY STUDY DATA

This data set is discussed in Section 11.2.3. The data set, with some additional variables, is available from the Mayo Clinic Web site referenced in Section G.2 of this Appendix under "Diabetes."

G.9 STANFORD HEART TRANSPLANT DATA

This data set is discussed in Example 11.4.3. The data were discussed by Crowley and Hu (1977) and given in the book by Kalbfleisch and Prentice (1980), where several typos occurred. The full data set given by Crowley and Hu is available in the file "stanford" on the Statlib Web site (http://lib.stat.cmu.edu/datasets/). The Kalbfleisch–Prentice data set and a summary of the differences in the two data sets is also provided. These data are also given in the S-Plus data frame "heart."

G.10 COLON CANCER RECURRENCE AND SURVIVAL DATA

This data set is based on a study by Moertel et al. (1990) and was discussed by Lin et al. (1999). Danyu Lin kindly provided the data used in Example 3.3.3 and later illustrations.

References

Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scand. J. Stat.*, **3**, 15–27.

Aalen, O. (1978a). Nonparametric estimation of partial transition probabilities in multiple decrement models. *Ann. Stat.*, **6**, 534–545.

Aalen, O. (1978b). Nonparametric inference for a family of counting processes. *Ann. Stat.*, **6**, 701–726.

Aalen, O. (1988). Heterogeneity in survival analysis. *Stat. Med.*, **7**, 1121–1137.

Aalen, O. (1994). Effects of frailty in survival analysis. *Stat. Methods Med. Res.*, **3**, 227–243.

Aalen, O. and Gjessing, H. K. (2001). Understanding the shape of the hazard rate: A process point of view (with discussion). *Stat. Sci.*, **16**, 1–22.

Aalen, O. and Husebye, E. (1991). Statistical analysis of repeated events forming renewal processes. *Stat. Med.*, **10**, 1227–1240.

Aalen, O. and Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scand. J. Stat.*, **5**, 141–150.

Abernethy, R. B. (1996). *The New Weibull Handbook*, 2nd ed. R. B. Abernethey, North Palm Beach, FL.

Abramowitz, M. and I. A. Stegun, Eds. (1965). *Handbook of Mathematical Functions*. Dover, New York.

Abrahamowicz, M., Ciampi, A., and Ramsay, J. O. (1992). Nonparametric density estimation for censored survival data; Regression-spline approach. *Can. J. Stat.*, **20**, 171–185.

Abrahamowicz, M., Mackenzie, T., and Esdaile, J. M. (1996). Time-dependent hazard ratio: Modeling and hypothesis testing with application in Lupus Nephritis. *J. Amer. Stat. Assoc.*, **91**, 1432–1439.

Adichie, J. N. (1967). Estimators of regression parameters based on rank tests. *Ann. Math. Stat.*, **38**, 894–904.

Aitchison, J. and Dunsmore, I. R. (1975). *Statistical Prediction Analysis*. Cambridge University Press, Cambridge.

Aitkin, A., and G. T. Wilson (1980). Mixture models, outliers and the EM algorithm. *Technometrics*, **22**, 325–331.

Akritas, M. G. (1988). Pearson-type goodness-of-fit tests: The univariate case. *J. Amer. Stat. Assoc.*, **83**, 222–230.

Akritas, M. G. and Torbeyns, A. F. (1997). Pearson-type goodness-of-fit tests for regression. *Can. J. Stat.*, **25**, 359–374.

Altschuler, B. (1970). Theory for the measurement of competing risks in animal experiments. *Math. Biosci.*, **6**, 1–11.

Andersen, P. K. (1982). Testing goodness-of-fit of Cox's regression and life model. *Biometrics*, **38**, 67–77. Correction: **40**, 1217.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Stat.*, **10**, 1100–1120.

Andersen, P. K. and Keiding, N. (2002). Multi-state models for event history analysis. *Stat. Methods Med. Res.*, **11**, 91–115.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1982). Linear nonparametric tests for comparison of counting processes, with application to censored survival data (with discussion). *Int. Stat. Rev.*, **50**, 219–258.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Andersen, P. K., Klein, J. P., Knudsen, K. M., and Palacios, P. T. (1997). Estimation of variance in Cox's regression model with shared gamma frailties. *Biometrics*, **53**, 1475–1484.

Anderson, J. A. and Senthilselvan, A. (1980). Smooth estimates for the hazard function. *J. Roy. Stat. Soc. B*, **42**, 322–327.

Anderson, K. M. (1991). A nonproportional hazards Weibull accelerated failure time regression model. *Biometrics*, **47**, 281–288.

Andrews, D. F. and Herzberg, A. M. (1985). *Data*. Springer-Verlag, New York.

Anscombe, F. J. (1964). Normal likelihood functions. *Ann. Inst. Stat. Math.*, **16**, 1–19.

Aranda-Ordaz, F. J. (1983). An extension of the proportional-hazards model for grouped data. *Biometrics*, **39**, 109–117.

Arjas, E. (1989). Survival models and martingale dynamics (with discussion). *Scand. J. Stat.*, **16**, 177–225.

Arjas, E. and Haara, P. (1984). A marked point process approach to censored failure data with complicated covariates. *Scand. J. Stat.*, **11**, 193–209.

Arjas, E. and Haara, P. (1992). Observation scheme and likelihood. *Scand. J. Stat.*, **19**, 111–132.

Armitage, P. (1975). *Sequential Medical Trials*, 2nd ed. John Wiley & Sons, New York.

Armitage, P. and Colton, T. (1998). *Encyclopedia of Biostatistics*. John Wiley & Sons, New York.

Arnold, B., Balakrishnan, N., and Nagaraja, H. N. (1992). *A First Course in Order Statistics*. John Wiley & Sons, New York.

Aroian, L. A. (1976). Applications of the direct method in sequential analysis. *Technometrics*, **18**, 301–306.

Ascher, H. and Feingold, H. (1984). *Repairable Systems Reliability*. Marcel Dekker, New York.

Atkinson, A. C. (1985). *Plots, Transformations and Regression*. Oxford University Press, London.

Bacchetti, P. (1990). Estimating the incubation period of AIDS by comparing population infection and diagnosis patterns. *J. Amer. Stat. Assoc.*, **85**, 1002–1008.

Bagdonavicius, V. and Nikulin, M. (1997). Transfer functionals and semiparametric regression models. *Biometrika*, **84**, 365–378.

Bagdonavicius, V. and Nikulin, M. S. (2001). Estimation in degradation models with explanatory variables. *Lifetime Data Anal.*, **7**, 85–103.

Bagdonavicius, V. and Nikulin, M. (2002). *Accelerated Life Models: Modeling and Statistical Analysis*. Chapman & Hall/CRC Press, London.

Bailey, K. R. (1983). The asymptotic joint distribution of regression and survival parameter estimates in the Cox regression model. *Ann. Stat.*, **11**, 39–58.

Bailey, K. R. (1984). Asymptotic equivalence between the Cox estimator and the general ML estimators of regression and survival parameters in the Cox model. *Ann. Stat.*, **12**, 730–736.

Bain, L. J. (1974). Analysis for the linear failure rate distribution. *Technometrics*, **16**, 551–559.

Balakrishnan, N. and Chan, P. S. (1995). Maximum likelihood estimation for the log-gamma distribution under Type II censored samples and associated inference. In *Recent Advances in Life-Testing and Reliability*, N. Balakrishnan, Ed., 409–421. CRC Press, Boca Raton.

Balasooriya, U., Gadag, V., and Saw, S. L. C. (2000). Progressively censored reliability sampling plans for the Weibull distribution. *Technometrics*, **42**, 160–167.

Baltazar-Aban, I. and Pena, E. A. (1995). Properties of hazard-based residuals and implications in model diagnostics. *J. Amer. Stat. Assoc.*, **90**, 185–197.

Bandeen-Roche, K. J. and Liang, K.-Y. (1996). Modelling failure-time associations in data with multiple levels of clustering. *Biometrika*, **83**, 29–39.

Bang, J. and Tsiatis, A. A. (2000). Estimating medical costs with censored data. *Biometrika*, **87**, 329–343.

Barlow, R. E. and Proschan F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart & Winston, New York.

Barlow, R. E., Marshall, A. W., and Proschan, F. (1963). Properties of probability distributions with monotone failure rate. *Ann. Math. Stat.*, **34**, 375–389.

Barlow, W. and Prentice, R. (1988). Residuals for relative risk regression. *Biometrika*, 75, 65–74.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994) *Inference and Asymptotics*. Chapman & Hall, London.

Barndorff-Nielsen, O. E. and Cox, D. R. (1996). Prediction and asymptotics. *Bernoulli*, **2**, 319–340.

Barnett, V. (1975). Probability plotting methods and order statistics. *Appl. Stat.*, **24**, 95–108.

Bartholomew, D. J. (1957). A problem in life testing. *J. Amer. Stat. Assoc.*, 52, 350–355.

Bartholomew, D. J. (1963). The sampling distribution of an estimate arising in life testing. *Technometrics*, **5**, 361–374.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc. London A*, **160**, 268–282.

Bartlett, M. S. and Kendall, D. G. (1946). The statistical analysis of variance-heterogeneity and the logarithmic transformation. *J. Roy. Stat. Soc. Suppl.*, **8**, 128–138.

Bartolucci, A. and Dickey, J. M. (1977). Comparative Bayesian and traditional inference for gamma-modelled survival data. *Biometrics*, **33**, 343–354.

Benichou, J. and Gail, M. H. (1990). Estimates of absolute cause-specific risk in cohort studies. *Biometrics*, **46**, 813–826.

Bennett, S. (1983a). Log-logistic regression models for survival data. *Appl. Stat.*, **32**, 165–171.

Bennett, S. (1983b). Analysis of survival data by the proportional odds model. *Stat. Med.*, **2**, 273–277.

Beran, R. (1990). Calibrating prediction regions. *J. Amer. Stat. Assoc.*, **85**, 715–723.

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer-Verlag, New York.

Berger, J. (2000). Bayesian analysis: A look at today and thoughts of tomorrow. *J. Amer. Stat. Assoc.*, **95**, 1269–1276.

Berkson, J. and Elveback, L. (1960). Competing exponential risks, with particular reference to the study of smoking and lung cancer. *J. Amer. Stat. Assoc.*, **55**, 415–428.

Berkson, J. and Gage, R. P. (1950). Calculation of survival rates for cancer. *Proc. Staff Meet. Mayo Clin.*, **25**, 270–286.

Bernstein, D. and Lagakos, S. W. (1978). Sample size and power determination for stratified clinical trials. *J. Stat. Comp. Simul.*, **8**, 65–73.

Betensky, R., Lindsey, J. C., Ryan, L. M., and Wand, M. P. (1999). Local EM estimation of the hazard function for interval-censored data. *Biometrics*, **55**, 238–245.

Bickel, P. J., Klassen, A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Inference for Semi-Parametric Models*. Johns Hopkins University Press, Baltimore, MD.

Bie, O., Borgan, O., and Liestol, K. (1987). Confidence intervals and confidence bands for the cumulative hazard rate function and their small sample properties. *Scand. J. Stat.*, **14**, 221–233.

Billmann, B., Antle, C., and Bain, L. J. (1972). Statistical inferences from censored Weibull samples. *Technometrics*, **14**, 831–840.

Birnbaum, Z. W. and Saunders, S. C. (1969). A new family of life distributions. *J. Appl. Probab.*, **6**, 319–327.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis*. Massachusetts Institute of Technology Press, Cambridge, MA.

Blischke, W. R. and Murthy, D. N. P. (2000). *Reliability: Modeling, Prediction and Optimization*. John Wiley & Sons, New York.

Blom, G. (1958). *Statistical Estimates and Transformed Beta-Variables*. John Wiley & Sons, New York.

Blossfeld, H. P. and Rohwer, G. (1995). *Techniques of Event History Modeling*. Lawrance-Erlbaum, Mahwah, NJ.

Blossfeld, H. P., Hamerle, A., and Mayer, K. U. (1989). *Event History Analysis*. Erlbaum, Hillsdale, NY.

Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. Roy. Stat. Soc. B*, **11**, 15–53.

Bogdanoff, D. and Pierce, D. A. (1973). Bayes-fiducial inference for the Weibull distribution. *J. Amer. Stat. Assoc.*, **68**, 659–664.

Bogdanoff, J. L. and Kozin, F. (1985). *Probabilistic Models of Cumulative Damage*. John Wiley & Sons, New York.

Böhning, D. (2000). *Computer-Assisted Analysis of Mixtures and Applications*. Chapman & Hall/CRC, Boca Raton.

Böhning, D., Schlattmann, P., and Dietz, E. (1996). Interval censored data: A note on the nonparametric maximum likelihood estimation of the distribution function. *Biometrika*, **83**, 462–466.

Borgan, O. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scand. J. Stat.*, **11**, 1–16. Correction: **11**, 275.

Bowman, A. W. and Wright, E. M. (2000). Graphical exploration of covariate effects on survival data through nonparametric quantile curves. *Biometrics*, **56**, 563–570.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA.

Box, G. E. P., Hunter, J. S., and Hunter, W. H. (1978). *Statistics for Experimenters*. John Wiley & Sons, New York.

Breslow, N. E. (1970). A generalized Kruskal-Wallis test for comparing k samples subject to unequal patterns of censorship. *Biometrika*, **57**, 579–594.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics*, **30**, 89–99.

Breslow, N. E. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Stat.*, **2**, 437–453.

Breslow, N. E. and Day, N. E. (1987). *Statistical Methods in Cancer Research. 2: The Design and Analysis of Cohort Studies*. I. A. R. C., Lyon.

Breslow, N. E. and Haug, C. (1972). Sequential comparison of exponential survival curves. *J. Amer. Stat. Assoc.*, **67**, 691–697.

Brindley, E. C. and Thompson, W. A. (1972). Dependence and aging aspects of multivariate survival. *J. Amer. Stat. Assoc.*, **67**, 822–830.

Brookmeyer, R. and Crowley, J. J. (1982). A confidence interval for the median survival time. *Biometrics*, **38**, 29–41.

Brookmeyer, R. and Gail, M. H. (1994). *AIDS Epidemiology: A Quantitative Approach*. Oxford University Press, New York.

Bryant, C. M. and Schmee, J. (1979). Confidence limits of MTBF for sequential test plans of MIL-STD 781. *Technometrics*, **21**, 33–42.

Bryson, M. C. and Siddiqui, M. M. (1969). Some criteria for aging. *J. Amer. Stat. Assoc.*, **64**, 1472–1483.

Buckland, W. R. (1964). *Statistical Assessment of the Life Characteristic*. Griffin, London.

Buckley, T. and James, I. (1979). Linear regression with censored data. *Biometrika*, **66**, 429–436.

Burke, M. and Yuen, K. C. (1995). Goodness-of-fit tests for the Cox model via bootstrap method. *J. Stat. Plan. Inf.*, **47**, 237–256.

Burr, I. W. (1942). Cumulative frequency distributions. *Ann. Math. Stat.*, **13**, 215–232.

Burridge, J. (1981). A note on maximum likelihood estimation for regression models using grouped data. *J. Roy. Stat. Soc. B*, **43**, 41–45.

Byar, D. P. and Green, S. B. (1980). The choice of treatment for cancer patients based on covariate information: Application to prostate cancer. *Bull. Cancer*, **67**, 477–488.

Cai, J. and Prentice, R. L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika*, **82**, 151–164.

Cain, K. C. and Lange, N. T. (1984). Approximate case influence for the proportional hazards regression model with censored data. *Biometrics*, **40**, 493–499.

Canfield, R. V. and Borgman, L. E. (1975). Some distributions of time to failure for reliability applications. *Technometrics*, **17**, 263–268.

Carbone, P. O., Kellerhouse, L. E., and Gehan, E. A. (1967). Plasmacytic myeloma: A study of the relationship of survival to various clinical manifestations and anomolous protein type in 112 patients. *Amer. J. Med.*, **42**, 937–948.

Carey, M. B. and Koenig, R. H. (1991). Reliability assessment based on accelerated degradation: A case study. *IEEE Trans. Reliab.*, **R-40**, 499–506.

Carlin, B. P. and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, London.

Carstensen, B. (1996). Regression models for interval censored data: Application to HIV infection in Danish homosexual men. *Stat. Med.*, **15**, 2177–2189.

Chaloner, K. and Larntz, K. (1992). Bayesian design for accelerated life testing. *J. Stat. Plan. Inf.*, **33**, 245–259.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Stat. Sci.*, **10**, 273–304.

Chao, M. and Glaser, R. E. (1978). The exact distribution of Bartlett's test statistic for homogeneity of variances with unequal sample sizes. *J. Amer. Stat. Assoc.*, **73**, 422–426.

Chappell, R. (1992). A note on linear rank tests and Gill and Schumacher's tests of proportionality. *Biometrika*, **79**, 199–201.

Cheng, R. C. H. and Iles, T. C. (1987). Corrected maximum likelihood in nonregular problems. *J. Roy. Stat. Soc. B*, **49**, 95–101.

Cheng, R. C. H. and Traylor, L. (1995). Non-regular maximum likelihood problems (with discussion). *J. Roy. Stat. Soc. B*, **57**, 3–44.

Cheng, S. C., Fine, J. P., and Wei, L. J. (1998). Prediction of cumulative incidence function under the proportional hazards model. *Biometrics*, **54**, 219–228.

Chernoff, H. (1953). Locally optimum designs for estimating parameters. *Ann. Math. Stat.*, **24**, 586–602.

Chernoff, H. and Lehmann, E. L. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Stat.*, **25**, 579–586.

Chhikara, R. S. and Folks, J. L. (1977). The Inverse Gaussian distribution as a lifetime model. *Technometrics*, **19**, 461–468.

Chhikara, R. S. and Folks, J. L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology and Applications*. Marcel Dekker, New York.

Chiang, C. L. (1960a). A stochastic study of the life table and its applications: I. Probability distribtuions of the biometric functions. *Biometrics*, **16**, 618–635.

Chiang, C. L. (1960b). A stochastic study of the life table and its applications: II. Sample variance of the observed expectation of life and other biometric functions. *Hum. Biol.*, **32**, 221–238.

Chiang, C. L. (1961). On the probability of death from specific causes in the presence of competing risks. *Proc. 4th Berkeley Symp.*, **4**, 169–180.

Clampi, A., Hogg, S. A., and Kates, L. (1986). Regression analysis of censored survival data with the generalized F family—an alternative to the proportional hazards model. *Stat. Med.*, **5**, 85–96.

Cinlar, E. (1977). Shock and wear models and Markov additive processes. In *The Theory and Applications of Reliability*, I. N. Shimi and C. P. Tsokos, Eds., pp. 193–214. Academic Press, New York.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrics*, **65**, 141–151.

Cleveland, W. S. (1985). *The Elements of Graphing Data*. Duxbury, Boston.

Cnaan, A. and Ryan, L. (1989). Survival analysis in natural history studies of disease. *Stat. Med.*, **8**, 1255–1268.

Collett, D. (1994). *Modelling Survival Data in Medical Research*. Chapman & Hall/CRC, Boca Raton.

Condra, L. W. (1993). *Reliability Improvement with Design of Experiments*. Marcel Dekker, New York.

Cook, J. A. (1988). Goodness of fit tests and two sample tests for grouped and censored lifetime data. Unpublished Ph.D. Thesis, University of Waterloo, Waterloo, Ont., Canada.

Cook, J. A. and Lawless, J. F. (1991). Two-sample tests with multinomial or grouped failure time data. *Biometrics*, **47**, 445–459.

Cook, R. D. and Johnson, M. E. (1981). A family of distributions for modelling nonelliptically symmetric multivariate data. *J. Roy. Stat. Soc. B*, **43**, 210–218.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall, London.

Cook, R. J. and Lawless, J. F. (1997). Marginal analysis of recurrent events and a terminating event. *Stat. Med.*, **16**, 911–924.

Cook, R. J. and Lawless, J. F. (2002). Analysis of repeated events. *Stat. Methods Med. Res.*, **11**, 141–166.

Cook, R. J., Ng, E. J. M., Mukherjee, J., and Vaughan, D. (1999). Two-state mixed renewal processes for chronic disease. *Stat. Med.*, **18**, 175–188.

Cornfield, J. (1957). The estimation of the probability of developing a disease in the presence of competing risks. *Amer. J. Public Health*, **47**, 601–607.

Coroni, C. (2002). The correct "ball bearing" data. *Lifetime Data Anal.*, to appear.

Cox, D. R. (1959). The analysis of exponentially distributed lifetimes with two types of failure. *J. Roy. Stat. Soc. B*, **21**, 411–421.

Cox, D. R. (1962). *Renewal Theory*. Methuen, London.

Cox, D. R. (1964). Some applications of exponential ordered scores. *J. Roy. Stat. Soc. B*, **26**, 103–110.

Cox, D. R. (1972a). Regression models and life tables (with discussion). *J. Roy. Stat. Soc. B*, **34**, 187–220.

Cox, D. R. (1972b). The statistical analysis of dependencies in point processes. In *Symposium on Point Processes*, P. A. W. Lewis, Ed. John Wiley & Sons, New York.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, **62**, 269–276.

Cox, D. R. (1978). Some remarks on the role in statistics of graphical methods. *Appl. Stat.*, **27**, 4–9.

Cox, D. R. (1979). A note on the graphical analysis of survival data. *Biometrika*, **66**, 188–190.

Cox, D. R. (1999). Some remarks on failure-times, surrogate markers, degradation, wear and the quality of life. *Lifetime Data Anal.*, **5**, 307–314.

Cox, D. R. and Hinkley, D. V. (1968). A note on the efficiency of least squares estimates. *J. Roy. Stat. Soc. B*, **30**, 284–289.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman & Hall, London.

Cox, D. R. and Lewis, P. A. W. (1966). *The Statistical Analysis of Series of Events*. Methuen, London.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman & Hall, London.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *J. Roy. Stat. Soc. B*, **30**, 248–275.

Cox, D. R. and Snell, E. J. (1989). *Analysis of Binary Data*, 2nd ed., Chapman & Hall, London.

Critchley, F., Ford, I., and Rijal, O. (1988). Interval estimation based on the profile likelihood: Strong Lagrangian theory with applications to discrimination. *Biometrika*, **75**, 21–28.

Crowder, M. J. (1986). On consistency and inconsistency of estimating equations. *Econ. Theory*, **2**, 305–330.

Crowder, M. J. (1987). On linear and quadratic estimating functions. *Biometrika*, **74**, 591–597.

Crowder, M. J. (1990). On some nonregular tests for a modified Weibull model. *Biometrika*, **77**, 499–506.

Crowder, M. J. (2000). Tests for a family of survival models based on extremes. In *Recent Advances in Reliability Theory*, N. Limnios and M. Nikulin, Eds., pp. 307–321. Birkhauser, Boston.

Crowder, M. J. (2001). *Classical Competing Risks*. Chapman & Hall/CRC, Boca Raton.

Crowder, M. J., Kimber, A. C., Smith, R. L., and Sweeting, T. J. (1991). *Statistical Analysis of Reliability Data*. Chapman & Hall, London.

Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant data. *J. Amer. Stat. Assoc.*, **72**, 27–36.

Crowley, J. and Storer, B. E. (1983). Comment on a paper by M. Aitkin et al. *J. Amer. Stat. Assoc.*, **78**, 277–281.

Crowley, J. and Thomas, D. R. (1975). Large sample theory for the log rank test. University of Wisconsin, Department of Statistics, Technical Report No. 415.

Currie, I. D. (1996). A note on Buckley-James estimators for censored data. *Biometrika*, **83**, 912–915.

Cuzick, J. (1985). Asymptotic properties of censored linear rank tests. *Ann. Stat.*, **13**, 133–141.

Dabrowska, D. M., Sun, G., and Horowitz, M. M. (1994). Cox regression in a Markov renewal model: An application to the analysis of bone marrow transplant data. *J. Amer. Stat. Assoc.*, **89**, 867–877.

D'Agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-Fit Techniques* (editors). Marcel Dekker, New York.

David, H. A. and Moeschberger, M. L. (1978). *Theory of Competing Risks*. Griffin, London.

Davis, D. J. (1952). An analysis of some failure data. *J. Amer. Stat. Assoc.*, **47**, 113–150.

Davis, H. T. and Feldstein, M. L. (1979). The generalized Pareto law as a model for progressively censored survival data. *Biometrika*, **66**, 299–306.

Davis, T. P. and Lawrance, A. J. (1989). The likelihood for competing risk survival analysis. *Scand. J. Stat.*, **16**, 23–28.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge.

de Boor, C. (1978). *A Practical Guide to Splines*. Springer-Verlag, New York.

Desmond, A. F. (1985). Stochastic models of failure in random environments. *Can. J. Stat.*, **13**, 171–183.

Diamond, I. R. and McDonald, J. W. (1992). The analysis of current status data. In *Demographic Applications of Event History Analysis*, J. Trussell, R. Hankinson, and J. Tilton, Eds. Clarendon Press, Oxford.

DiCiccio, T. J. (1988). Likelihood inference for linear regression models. *Biometrika*, **75**, 29–34.

DiCiccio, T. J., Monti, M. A., and Stern, S. E. (2001). Simple and accurate one-sided inference from signed roots of likelihood ratios. *Can. J. Stat.*, **29**, 67–76.

Dinse, G. E. (1986). Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause-of-death. *J. Amer. Stat. Assoc.*, **81**, 328–336.

Ditlevsen, O. (1986). Random fatigue crack growth—a first passage problem. *Eng. Fract. Mech.*, **23**, 467–477.

Doganaksoy, N. and Schmee, J. (2000). Practical aspects of corrected likelihood ratio confidence intervals: A discussion of Jeng-Meeker and Wong-Wu. *Technometrics*, **42**, 156–159.

Doksum, K. and Hoyland, A. (1992). Models for variable-stress accelerated testing experiments based on Wiener processes and the inverse Gaussian distribution. *Technometrics*, **34**, 74–82.

Doksum, K. A. and Normand, S.-L. (1995). Gaussian models for degradation processes—Part I: Methods for the analysis of biomarker data. *Lifetime Data Anal.*, **1**, 131–144.

Drolette, M. E. (1975). The effect of incomplete followup. *Biometrics*, **31**, 135–144.

Dubey, S. D. (1968). A compound Weibull distribution. *Nav. Res. Logist. Q.*, **15**, 179–188.

Duchesne, T. and Lawless, J. F. (2000). Alternative time scales and failure time models. *Lifetime Data Anal.*, **6**, 157–179.

Duchesne, T. and Stafford, J. E. (2002). A kernel density estimate for interval censored data. Unpublished manuscript.

Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. SIAM, Philadelphia.

Durbin, J. (1975). Kolmogorov–Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. *Biometrika*, **62**, 5–22.

Dyer, D. D. and Keating, J. P. (1980). On the determination of critical values for Bartlett's test. *J. Amer. Stat. Assoc.*, **75**, 313–319.

Efron, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp.*, **4**, 831–853.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Stat. Assoc.*, **72**, 555–565.

Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Stat. Assoc.*, **76**, 312–319.

Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan–Meier curve. *J. Amer. Stat. Assoc.*, **83**, 414–425.

Efron, B. and Petrosian, V. (1999). Nonparametric methods for doubly truncated data. *J. Amer. Stat. Assoc.*, **94**, 824–834.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.

Elandt-Johnson, R. C. (1977). Various estimators of conditional probabilities of death in followup studies: Summary of results. *J. Chronic Dis.*, **30**, 246–256.

Elandt-Johnson, R. C. and Johnson, N. L. (1980). *Survival Models and Data Analysis*. John Wiley & Sons, New York.

Elveback, L. (1958). Estimation of survivorship in chronic disease: The "actuarial" method. *J. Amer. Stat. Assoc.*, **53**, 420–440.

Engelhardt, M. (1975). On simple estimation of the parameters of the Weibull or extreme value distribution. *Technometrics*, **17**, 369–374.

Engelhardt, M. and Bain, L. J. (1977). Uniformly most powerful unbiased tests on the scale parameter of a gamma distribution with a nuisance shape parameter. *Technometrics*, **19**, 77–81.

Engelhardt, M. and Bain, L. J. (1978a). Construction of optimal unbiased inference procedures for the parameters of the gamma distribution. *Technometrics*, **20**, 485–489.

Engelhardt, M. and Bain, L. J. (1978b). Tolerence limits and confidence limits on reliability for the two-parameter exponential distribution. *Technometrics*, **20**, 37–39.

Epstein, B. (1954). Truncated life tests in the exponential case. *Ann. Math. Stat.*, **25**, 555–564.

Epstein, B. and Sobel, M. (1953). Life testing. *J. Amer. Stat. Assoc.*, **48**, 486–502.

Epstein, B. and Sobel, M. (1954). Some theorems relevant to life testing from an exponential distribution. *Ann. Math. Stat.*, **25**, 373–381.

Epstein, B. and Sobel, M. (1955). Sequential life tests in the exponential case. *Ann. Math. Stat.*, **26**, 82–93.

Escobar, L. A. and Meeker, W. Q. (1992). Assessing influence in regression analysis with censored data. *Biometrika*, **48**, 507–528.

Escobar, L. A. and Meeker, W. Q. (1994). Fisher information matrix for the extreme value, normal, and logistic distributions and censored data. *Appl. Stat.*, **43**, 533–540.

Etezadi-Amoli, J. and Ciampi, A. (1987). Extended hazard regression for censored survival data with covariates: A spline approximation for the baseline hazard function. *Biometrics*, **43**, 181–192.

Fahrmeir, L. and Kaufmann, H. (1987). Regression models for nonstationary categorical time series. *J. Time Ser. Anal.*, **8**, 147–160.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modeling Based on Generalized Linear Models*. Springer-Verlag, New York.

Falls, L. W. (1970). Estimation of parameters in compound Weibull distributions. *Technometrics*, **12**, 399–407.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, **38**, 1041–1046.

Farewell, V. T. and Prentice, R. L. (1977). A study of distributional shape in life testing. *Technometrics*, **19**, 69–75.

Farewell, V. T., Lawless, J. F., Gladman, D. D., and Urowitz, M. B. (2002). Analysis of the effect of lost-to-followup on the estimation of mortality from patient registry data. Unpublished manuscript.

Faucett, C. L. and Thomas, D. C. (1996). Simultaneouly modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Stat. Med.*, **15**, 1663–1685.

Feigl, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, **21**, 826–838.

Ferguson, T. S. and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Stat.*, **7**, 163–186.

Fertig, K. W. and Mann, N. R. (1980). Life test sampling plans for two-parameter Weibull populations. *Technometrics*, **22**, 165–177.

Fine, J. P. (1999). Analyzing competing risks data with transformation models. *J. Roy. Stat. Soc. B*, **61**, 817–830.

Fine, J. P. and Jiang, H. (2000). On association in a copula with time transformations. *Biometrika*, **87**, 559–572.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, **42**, 845–865.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London A*, **222**, 309–368.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc. A*, **144**, 285–307.

Flehinger, B. J., Reiser, B. and Yaschin, E. (1998). Survival with competing risks and masked causes of failures. *Biometrika*, **85**, 151–164.

Fleming, T. R. (1978a). Nonparametric estimation for non-time-homogeneous Markov processes in the problem of competing risks. *Ann. Stat.*, **6**, 1057–1070.

Fleming, T. R. (1978b). Asymptotic distribution results in competing risks estimation. *Ann. Stat.*, **6**, 1071–1079.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. John Wiley & Sons, New York.

Fleming, T. R., O'Fallon, J. R., O'Brien, P. C., and Harrington, D. P. (1980). Modified Kolmogorov–Smirnov test procedures with application to arbitrarily right censored data. *Biometrics*, **36**, 607–626.

Fleming, T. R., Harrington, D. P., and O'Sullivan, M. (1987). Supremum versions of the logrank and generalized Wilcoxon statistics. *J. Amer. Stat. Assoc.*, **82**, 312–320.

Fletcher, R. (1987). *Practical Methods of Optimization*, 2nd ed. John Wiley & Sons, New York.

Folkes, E. B. (1979). Some methods for studying the mixture of two normal (log-normal) distributions. *J. Amer. Stat. Assoc.*, **74**, 561–575.

Follman, D. and Goldberg, M. (1988). Distinguishing heterogeneity from decreasing hazard rates. *Technometrics*, **30**, 389–396.

Fong, D. Y. T., Lam, K. F., Lawless, J. F., and Lee, Y. W. (2001). Dynamic random effects models for times between repeated events. *Lifetime Data Anal.*, **7**, 345–362.

Fraser, D. A. S. (1968). *The Structure of Inference*. John Wiley & Sons, New York.

Fraser, D. A. S. (1979). *Inference and Linear Models*. McGraw-Hill, New York.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *N. Amer. Act. J.* **2**, 1–25.

Freireich, E. O. et al. (1963). The effect of 6-mercaptopurine on the duration of steroid induced remission in acute leukemia. *Blood*, **21**, 699–716.

Friedman, M. (1982). Piecewise exponential models for survival data with covariates. *Ann. Stat.*, **10**, 101–113.

Frydman, H. (1994). A note on nonparmetric estimation of the distribution function from interval-censored and truncated observations. *J. Roy. Stat. Soc. B*, **56**, 71–74.

Fryer, J. G. and Holt, D. (1976). On the robustness of the power function of the one-sample test for the negative exponential distribution. *Commun. Stat.*, **A5**, 723–734.

Fuchs, H. J. et al. (1994). The effect of aerosolized recombinant human DNase on respiratory exacerbations and pulmonary function in patients with cystic fibrosis. *N. Engl. J. Med.*, **331**, 637–642.

Furth, J., Upton, A. C., and Kimball, A. W. (1959). Late pathologic effects of atomic detonation and their pathogenesis. *Radiat. Res. Suppl.*, **1**, 243–264.

Fusaro, R. E., Nielsen, J. P., and Scheike, T. H. (1993). Marker-dependent hazard estimation: An application to AIDS. *Stat. Med.*, **12**, 843–865.

Gail, M. H. (1975). A review and critique of some models used in competing risk analysis. *Biometrics*, **31**, 209–222.

Gail, M. H. and Gastwirth, J. L. (1978). A scale-free goodness of fit test for the exponential distribution based on the Gini Statistic. *J. Roy. Stat. Soc. B*, **40**, 350–357.

Gail, M. H. and Ware, J. (1979). Comparing observed life table data with a known survival curve in the presence of random censorship. *Biometrics*, **35**, 385–391.

Gail, M. H., Santner, T. J., and Brown C. C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics*, **36**, 255–266.

Gaver, D. P. and Acar, M. (1979). Analytical hazard representations for use in reliability, mortality and simulation studies. *Commun. Stat.*, **8**, 91–111.

Gaynor, J. J., Feuer, E. J., Tan, C. C., et al. (1993). On the use of cause-specific failure and conditional probabilities: Examples from clinical oncology data. *J. Amer. Stat. Assoc.*, **88**, 400–409.

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, **52**, 203–233.

Gehan, E. A. (1969). Estimating survival functions from the life table. *J. Chronic Dis.*, **21**, 629–644.

Gehan, E. A. and Siddiqui, M. M. (1973). Simple regression methods for survival time studies. *J. Amer. Stat. Assoc.*, **68**, 848–856.

Geisser, S. (1993). *Predictive Inference: An Introduction*. Chapman & Hall, London.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman & Hall, London.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure for dependence parameters in multivariate families of distributions. *Biometrika*, **82**, 543–552.

Gentleman, R. and Crowley, J. (1991). Graphical methods for censored data. *J. Amer. Stat. Assoc.*, **86**, 678–683.

Gentleman, R. and Geyer, C. J. (1994). Maximum likelihood for interval censored data: Consistency and computation. *Biometrika*, **81**, 618–623.

Gentleman, R. G., Lawless, J. F., Lindsey, J., and Yan, P. (1994). Multistate Markov models for analyzing incomplete disease history data, with illustrations for HIV disease. *Stat. Med.*, **13**, 805–821.

Ghosh, D. and Lin, D. Y. (2000). Nonparmetric analysis of recurrent events and death. *Biometrics*, **56**, 554–562.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. I. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Gill, R. D. (1980). *Censoring and Stochastic Integrals*. Mathematical Centre Tracts 124. Mathematisch Centrum, Amsterdam.

Gill, R. D. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Stat.*, **11**, 49–58.

Gill, R. D. and Johansen, S. (1990). A survey of product-integration with a view towards application in survival analysis. *Ann. Stat.*, **18**, 1501–1555.

Gill, R. D. and van der Vaart, A. W. (1993). Non- and semi-parametric maximum likelihood estimators and the von Mises method, Part II. *Scand. J. Stat.*, **20**, 171–188.

Gill, P. E., Murray, W., and Wright, M. H. (1981). *Practical Optimization*. Academic Press, New York.

Glaser, R. E. (1980). Bathtub and related failure rate characterizations. *J. Amer. Stat. Assoc.*, **75**, 667–672.

Glasser, M. (1965). Regression analysis with dependent variable censored. *Biometrics*, **21**, 300–307.

Godambe, V. P., Ed. (1991). *Estimating Functions*. Oxford University Press, Oxford.

Godambe, V. P. and Thompson, M. E. (1989). An extension of quasilikelihood estimation (with discussion). *J. Stat. Plan. Inf.*, **22**, 137–172.

Goldthwaite, L. (1961). Failure rate study for the lognormal lifetime model. *Proc. Seventh Nat. Symp. Reliab. Qual. Control*, 208–213.

Gore, S. M., Pocock, S. J., and Kerr, G. R. (1984). Regression models and nonproportional hazards in the analysis of breast cancer survival. *Appl. Stat.*, **33**, 176–195.

Gould, A. (1986). *Some Issues in the Regression Analysis of Survival Data*. Unpublished Ph.D. Thesis, University of Waterloo, Waterloo, Ont., Canada.

Gould, A. and Lawless, J. F. (1988). Consistency and efficiency of regression coefficient estimates in location-scale models. *Biometrika*, **73**, 535–540.

Govindarajulu, Z. (1964). A supplement to Mendenhall's bibliography on life testing and related topics. *J. Amer. Stat. Assoc.*, **59**, 1231–1291.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, **81**, 515–526.

Gray, R. J. and Pierce, D. A. (1985). Goodness-of-fit tests for censored data. *Ann. Stat.*, **13**, 552–563.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. Chapman & Hall, London.

Greenwood, M. (1926). The natural duration of cancer. *Reports of Public Health and Medical Subjects*, Volume 33. Her Majesty's Stationery Office, London.

Griffiths, D. A. (1980). Interval estimation for the three-parameter log-normal distribution via the likelihood function. *Appl. Stat.*, **29**, 58–68.

Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser Verlag, Basel.

Grüger, J., Kay, R., and Schumacher, M. (1991). The validity of inferences based on incomplete observations in disease state models. *Biometrics*, **47**, 595–605.

Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press, New York.

Guo, G. (1993). Event-history analysis for left-truncated data. *Sociol. Methodol.*, **23**, 217–243.

Habib, M. G. and Thomas, D. R. (1986). Chi-square goodness-of-fit tests for randomly censored data. *Ann. Stat.*, **14**, 759–765.

Hager, H. W. and Bain, L. J. (1970). Inferential procedures for the generalilzed gamma distribution. *J. Amer. Stat. Assoc.*, **65**, 1601–1609.

Hahn, G. J. and Meeker, W. Q. (1991). *Statistical Intervals: A Guide for Practitioners*. John Wiley & Sons, New York.

Hajek, J. and Sidak, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansions*. Springer-Verlag, New York.

Hall, W. J., Rogers, W. H., and Pregibon, D. (1982). Outliers matter in survival analysis. *Rand Corporation Technical Report P-6761*, Santa Monica, CA.

Hall, P., Peng, L., and Tajvidi, N. (1999). On prediction intervals based on predictive likelihood or bootstrap methods. *Biometrika*, **86**, 871–880.

Halperin, M. (1952). Maximum-likelihood estimation in truncated samples. *Ann. Math. Stat.*, **23**, 226–238.

Hamada, M. (1995). Analysis of experiments for reliability improvement and robust reliability. *Recent Advances in Life-Testing and Reliability*, Ch. 9, N. Balakrishnan, Ed. CRC Press, Boca Raton.

Hamada, M. S. and Wu, C. F. J. (1991). Analysis of censored data from highly fractionated experiments. *Technometrics*, **33**, 25–38.

Hamada, M. and Wu, C. F. J. (1995). Analysis of censored data from fractionated experiments: A Bayesian approach. *J. Amer. Stat. Assoc.*, **90**, 467–477.

Hamerle, A. (1991). On the treatment of interrupted spells and initial conditions in event history analysis. *Sociol. Methods Res.*, **19**, 388–414.

Harrell, F. E. and Goldstein, R. (1997). A survey of microcomputer survival analysis software: The need for an integrated framework. *Amer. Stat.*, **51**, 360–373.

Harrington, D. P. and Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika*, **69**, 553–566.

Harter, J. L. and Moore, A. H. (1968). Maximum likelihood estimation, from doubly censored samples, of the parameters of the first asymptotic distribution of extreme values. *J. Amer. Stat. Assoc.*, **63**, 889–901.

Harter, J. L. and Moore, A. H. (1976). An evaluation of exponential and Weibull test plans, *IEEE Trans. Reliab.*, **R-25**, 100–104.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman & Hall, London.

He, W. (2001). *Methods for the Analysis of Multivariate Failure Times*. Unpublished Ph.D. Thesis, University of Waterloo, Waterloo, Ont., Canada.

He, W. and Lawless, J. F. (2002). Flexible maximum likelihood for bivariate proportional hazards models. Unpublished manuscript.

Heitjan, D. F. (1989). Inference from grouped continuous data: A review. *Stat. Sci.*, **4**, 164–183.

Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks*. John Wiley & Sons, New York.

Heyde, C. C. (1997). *Quasi-Likelihood and Its Application*. Springer-Verlag, New York.

Hill, B. M. (1963). Information for estimating the proportions in mixtures of exponential and normal distributions. *J. Amer. Stat. Assoc.*, **58**, 918–932.

Hjort, N. L. (1990a). Goodness of fit tests in models for life history data based on cumulative hazard rates. *Ann. Stat.*, **18**, 1221–1258.

Hjort, N. L. (1990b). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Stat.*, **18**, 1259–1294.

Hjorth, J. (1980). A reliability distribution with increasing, decreasing, constant, and bathtub-shaped failure rates. *Technometrics*, **22**, 99–108.

Hoel, D. G. (1972). A representation of mortality by competing risks. *Biometrics*, **28**, 475–488.

Hogan, J. W. and Laird, N. (1997a). Mixture models for the joint distribution of repeated measures and event times. *Stat. Med.*, **16**, 239–257.

Hogan, J. W. and Laird, N. (1997b). Model based approaches to analyzing incomplete longitudinal and failure time data. *Stat. Med.*, **16**, 259–272.

Hogg, R. V. (1956). On the distribution of the likelihood ratio. *Ann. Math. Stat.*, **27**, 529–532.

Holford, T. R. (1976). Life tables with concomitant information. *Biometrics*, **32**, 587–597.

Hollander, M. and Pena, E. (1989). Families of confidence bands for the survival function under the general random censorship model and the Koziol-Green model. *Can. J. Stat.*, **17**, 59–74.

Holt, J. D. (1978). Competing risk analysis with special reference to matched pair experiments. *Biometrika*, **65**, 159–166.

Holt, J. D. and Prentice, R. L. (1974). Survival analysis in twin studies and matched pair experiments. *Biometrika*, **61**, 17–30.

Hora, R. B. and Buehler, R. J. (1966). Fiducial theory and invariant estimation. *Ann. Math. Stat.*, **37**, 643–656.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, **73**, 387–396.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer-Verlag, New York.

Hougaard, P., Harvald, B., and Holm, N. V. (1992). Measuring the similarities between the lifetimes of adult Danish twins born between 1881–1930. *J. Amer. Stat. Assoc.*, **87**, 17–24.

Hu, X. J. and Lawless, J. F. (1996). Estimation from truncated lifetime data with supplementary information on covariates and censoring times. *Biometrika*, **83**, 747–761.

Hu, X. J., Lawless, J. F., and Suzuki, K. (1998). Nonparametric estimation of a lifetime distribution when censoring times are missing. *Technometrics*, **40**, 3–13.

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Stat.*, **24**, 540–568.

Huang, J. and Wellner, J. A. (1997). Interval censored survival data: A review of recent progress. In *Proceedings of the 1st Seattle Symposium in Biostatistics*, D. Y. Lin and T. R. Fleming, Eds., pp. 123–169. Springer-Verlag, New York.

Huang, Y. and Louis, T. A. (1998). Nonparametric estimation of the joint distribution of survival time and mark variables. *Biometrika*, **85**, 785–798.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 221–233.

Huster, W. J., Brookmeyer, R., and Self, S. G. (1989). Modelling paired survival data with covariates. *Biometrics*, **45**, 145–156.

Hyde, J. (1977). Testing survival under right censoring and left truncation. *Biometrika*, **64**, 225–230.

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer-Verlag, New York.

Inagaki, N. (1973). Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator. *Ann. Inst. Stat. Math.*, **25**, 1–25.

Jacobsen, M. (1984). Maximum likelihood estimation in the multiplicative intensity model: A survey. *Int. Stat. Rev.*, **52**, 193–207.

James, I. R. (1986). On estimating equations with censored data. *Biometrika*, **73**, 35–42.

James, I. R. and Smith, P. J. (1984). Consistency results for linear regression with censored data. *Ann. Stat.*, **12**, 590–600.

Jardine, A. K. S. (1979). Solving industrial replacement problems. *Proceedings of the 1979 Annual Reliability and Maintainability Symposium*, pp. 136–141. Institute of Electrical and Electronics Engineers, New York.

Jeng, S-L. and Meeker, W. Q. (2000). Comparisons of approximate confidence interval procedures for Type I censored data. *Technometrics*, **42**, 135–148.

Jewell, N. P. and Kalbfleisch, J. D. (1992). Marker processes and applications to AIDS. In *AIDS Epidemiology: Methodological Issues*, N. P. Jewell, K. Dietz, and V. T. Farewell, Eds. Birkhauser, Boston.

Jewell, N. P. and Kalbfleisch, J. D. (1996). Marker processes in survival analysis. *Lifetime Data Anal.*, **2**, 15–29.

Jewell, N. P. and Nielsen, J. P. (1993). A framework for consistent prediction rules based on markers. *Biometrika*, **80**, 153–164.

Jewell, N. P. and van der Laan, M. (1997). Singly and doubly censored current status data with extensions to multi-state counting processes. In *Proceedings of the 1st Seattle Symposium*

in Biostatistics, D. Y. Lin and T. R. Fleming, Eds., pp. 171–184. Springer-Verlag, New York.

Jewell, N. P. and Shiboski, S. (1990). Statistical analysis of HIV infectivity based on partner studies. *Biometrics*, **46**, 1133–1150.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

Johansen, S. (1978). The product limit estimate as a maximum likelihood estimate. *Scand. J. Stat.*, **5**, 195–199.

Johansen, S. (1983). An extension of Cox's regression model. *Int. Stat. Rev.*, **51**, 258–262.

Johnson, N. L. (1977). Approximate relationships among some estimators of mortality probabilities. *Biometrics*, **33**, 542–545.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. 1. John Wiley & Sons, New York.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol. 2. John Wiley & Sons, New York.

Johnson, R. A. and Mehrotra, K. G. (1972). Locally most powerful rank tests for the two-sample problem with censored data. *Ann. Math. Stat.*, **43**, 823–831.

Joly, P., Commenges, D., and Leteneur, L. (1998). A penalized likelihood approach for arbitrarily censored and truncated data: Application to age-specific incidence of dementia. *Biometrics*, **54**, 185–194.

Jones, D. R. and Whitehead, J. (1979). Sequential forms of the log rank and modified Wilcoxon tests for censored data. *Biometrika*, **66**, 105–113.

Jones, M. P. (1997). A class of semiparametric regressions for the accelerated failure time model. *Biometrika*, **84**, 73–84.

Jones, M. P. and Crowley, J. J. (1989). A general class of nonparametric tests for survival analysis. *Biometrics*, **45**, 157–170.

Jones, M. P. and Crowley, J. J. (1990). Asymptotic properties of a general class of nonparametric tests for survival analysis. *Ann. Stat.*, **18**, 1203–1220.

Jorgensen, B. (1981). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Springer-Verlag, Heidelberg.

Jurečková, J. (1971). Nonparametric estimate of regression coefficients. *Ann. Math. Stat.*, **42**, 1328–1338.

Kalbfleisch, J. D. (1974). Some efficiency calculations for survival distributions. *Biometrika*, **61**, 31–38.

Kalbfleisch, J. D. (1978). Nonparametric Bayesian analysis of survival time data. *J. Roy. Stat. Soc. B*, **40**, 214–221.

Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *J. Amer. Stat. Assoc.*, **80**, 863–871.

Kalbfleisch, J. D. and Lawless, J. F. (1988a). Likelihood analysis of multi-state models for disease incidence and mortality. *Stat. Med.*, **7**, 149–160.

Kalbfleisch, J. D. and Lawless, J. F. (1988b). Estimation of reliability in field performance studies (with discussion). *Technometrics*, **30**, 365–388.

Kalbfleisch, J. D. and Lawless, J. F. (1989). Inference based on retrospective ascertainment. An analysis of the data on transfusion related AIDS. *J. Amer. Stat. Assoc.*, **84**, 360–372.

Kalbfleisch, J. D. and Lawless, J. F. (1991). Regression models for right-truncated data, with applications to AIDS incubation times and reporting lags. *Stat. Sin.*, **1**, 19–32.

Kalbfleisch, J. D. and Lawless, J. F. (1992). Some useful methods for truncated data. *J. Qual. Technol.*, **24**, 145–152.

Kalbfleisch, J. D. and MacKay, R. J. (1978). Censoring and the immutable likelihood. University of Waterloo, Department of Statistics, Technical Report 78–09.

Kalbfleisch, J. D. and MacKay, R. J. (1979). On constant-sum models for censored survival data. *Biometrika*, **66**, 87–90.

Kalbfleisch, J. D. and McIntosh, A. (1977). Efficiency in survival distributions with time-dependent covariables. *Biometrika*, **64**, 47–50.

Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika*, **60**, 267–279.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York.

Kalbfleisch, J. D. and Prentice, R. L. (2002) *The Statistical Analysis of Failure Time Data*, 2nd ed., John Wiley & Sons, New York.

Kalbfleisch, J. D. and Sprott, D. A. (1970). Application of likelihood methods to models involving a large number of parameters (with discussion). *J. Roy. Stat. Soc. B*, **32**, 175–208.

Kalbfleisch, J. D. and Sprott, D. A. (1974). Marginal and conditional likelihoods, *Sankhya A*, **35**, 311–328.

Kalbfleisch, J. G. (1985). *Probability and Statistical Inference*, 2nd ed., Vol. 2. Springer-Verlag, New York.

Kao, J. H. K. (1959). A graphical estimation of mixed Weibull parameters in life testing of electron tubes. *Technometrics*, **1**, 389–407.

Kao, P., Kao, E. P. C., and Mogg, J. M. (1979). A simple procedure for computing perfomance characteristics of truncated sequential tests with exponential life times. *Technometrics*, **21**, 229–232.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete obsesrvations. *J. Amer. Stat. Assoc.*, **53**, 457–481.

Kay, R. (1977). Proportional hazard regression models and the analysis of censored survival data. *Appl. Stat.*, **26**, 227–237.

Kay, R. (1979). Some further asymptotic efficiency calculations for survival data regression models. *Biometrika*, **66**, 91–96.

Kay, R. (1984). Goodness of fit methods for the proportional hazards model. *Rev. Epidemiol. Santé Publique*, **32**, 185–198.

Kay, R. (1986). Treatment effects in Competing Risks analysis of prostate cancer data. *Biometrics*, **42**, 203–211.

Keiding, N. (1991). Age-specific incidence and prevalence: A statistical perspective (with discussion). *J. Roy. Stat. Soc. A*, **154**, 371–412.

Keiding, N. (1992). Independent delayed entry. In *Survival Analysis: State of the Art*, J. P. Klein and P. K. Goel, Eds., pp. 309–326. Kluwer, Dordrecht.

Keiding, N. and Gill, R. D. (1990). Random truncation models and Markov processes. *Ann. Stat.*, **18**, 582–602.

Keiding, N., Begtrup, K., Scheike, T. H., and Hasibeder, G. (1996). Estimation from current-status data in continuous time. *Lifetime Data Anal.*, **2**, 119–129.

Kellerer, A. M. and Chmelevsky, D. (1983). Small-sample properties of censored-data rank tests. *Biometrics*, **39**, 675–682.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Ann. Math. Stat.*, **27**, 887–906.

Kim, D. K. (1997). Regression analysis of interval-censored survival data with covariates using log-linear models. *Biometrics*, **53**, 1274–1283.

Kimball, A. W. (1960). Estimation of mortality intensities in animal experiments. *Biometrics*, **16**, 505–521.

Kimball, A. W. (1969). Models for the estimation of competing risks from grouped data. *Biometrics*, **25**, 329–337.

Kimber, A. C. (1990). Exploratory data analysis for possibly censored data from skewed distributions. *Appl. Stat.*, **39**, 21–30.

Klein, J. P. (1991). Small sample moments of some estimators of the variance of the Kaplan–Meier and Nelson–Aalen estimators. *Scand. J. Stat.*, **18**, 333–340.

Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*, **48**, 795–806.

Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis*. Springer-Verlag, New York.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (1998). *Loss Models*. John Wiley & Sons, New York.

Kooperberg, C. and Clarkson, D. B. (1997). Hazard regression with interval-censored data. *Biometrics*, **53**, 1485–1494.

Kooperberg, C. and Stone, C. J. (1992). Logspline density estimation for censored data. *J. Comput. Graph. Stat.*, **1**, 301–328.

Kooperberg, C., Stone, C. J. and Truong, Y. L. (1995). Hazard regression. *J. Amer. Stat. Assoc.*, **90**, 78–94.

Kopecky, K. J. and Pierce, D. A. (1979). Efficiency of smooth goodness of fit tests. *J. Amer. Stat. Assoc.*, **74**, 393–397.

Kordonsky, K. B. and Gertsbakh, I. (1993). Choice of the best time scale for system reliability analysis. *Eur. J. Oper. Res.*, **65**, 235–246.

Kotz, S., Johnson, N. L., and Reid, C. B. (1988). *Encyclopedia of Statistical Sciences*. John Wiley & Sons, New York.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). *Continuous Multivariate Distributions*, 2nd ed. John Wiley & Sons, New York.

Koziol, J. A. and Petkau, A. J. (1978). Sequential testing of the equality of two survival distributions using the modified Savage statistic. *Biometrika*, **65**, 615–623.

Krall, J., Uthoff, V., and Harley, J. (1975). A step-up procedure for selecting variables associated with survival. *Biometrics*, **31**, 49–57.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion analysis of variance. *J. Amer. Stat. Assoc.*, **47**, 583–621.

Kuzma, J. W. (1967). A comparison of two life table methods. *Biometrics*, **23**, 51–64.

Lachin, J. M. and Foulkes, M. A. (1986). Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to followup, noncompliance, and stratification. *Biometrics*, **42**, 507–519.

Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics*, **35**, 139–156.

Lagakos, S. W. (1981). The graphical evaluation of explanatory variables in proportional hazard regression models. *Biometrika*, **68**, 93–98.

Lagakos, S. W. and Williams, J. S. (1978). Models for censored survival analysis: A cone class of variable-sum models. *Biometrika*, **65**, 181–189.

Lai, T. L. and Ying, Z. (1991). Rank regression methods for left-truncated and right-censored data. *Ann. Stat.*, **19**, 531–556.

Laird, N. and Olivier, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *J. Amer. Stat. Assoc.*, **76**, 231–240.

Lakatos, E. (1988). Sample sizes based on the log–rank statistic in complex clinical trials. *Biometrics*, **44**, 229–241.

Lancaster, T. (1990) *The Econometric Analysis of Transition Data*. Cambridge University Press, Cambridge.

Lancaster, T. and Nickell, S. (1980). The analysis of reemployment probabilities for the unemployed. *J. Roy. Stat. Soc. A*, **143**, 141–165.

Lange, K. (1999). *Numerical Analysis for Statisticians*. Springer-Verlag, New York.

Larson, M. G. and Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Appl. Stat.*, **34**, 201–211.

Latta, R. B. (1981). A Monte Carlo study of some two sample rank tests with censored data. *J. Amer. Stat. Assoc.*, **76**, 713–719.

Lawless, J. F. (1971). A prediction problem concerning samples from the exponential distribution, with application in life testing. *Technometrics*, **4**, 725–730.

Lawless, J. F. (1972). Confidence interval estimation for the parameters of the Weibull distribution. *Util. Math.*, **2**, 71–87.

Lawless, J. F. (1973). On the estimation of safe life when the underlying life distribution is Weibull. *Technometrics*, **15**, 857–865.

Lawless, J. F. (1975). Construction of tolerance bounds for the extreme value and Weibull distributions. *Technometrics*, **17**, 255–261.

Lawless, J. F. (1976). Confidence interval estimation in the inverse power law model. *Appl. Stat.*, **25**, 128–138.

Lawless, J. F. (1978). Confidence interval estimation for the Weibull and extreme value distributions. *Technometrics*, **20**, 355–364.

Lawless, J. F. (1980). Inference in the generalized gamma and log gamma distributions. *Technometrics*, **22**, 409–419.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, New York.

Lawless, J. F. (1986). A note on lifetime regression models. *Biometrika*, **73**, 509–512.

Lawless, J. F. (1995). The analysis of recurrent events for multiple subjects. *Appl. Stat.*, **44**, 487–498.

Lawless, J. F. (2000). Introduction to two classics in reliability theory. *Technometrics*, **42**, 5–6.

Lawless, J. F. (2003). Event history analysis and longitudinal surveys. In *Analysis of Survey Data*, R. L. Chambers and C. J. Skinner, Eds. John Wiley & Sons, Chichester.

Lawless, J. F. and Fong, D. Y. T. (1999). State duration models in clinical and observational studies. *Stat. Med.*, **18**, 2365–2376.

Lawless, J. F. and Mann, N. R. (1976). Tests for homogeneity for extreme value scale parameters. *Commun. Stat.*, **A5**, 389–405.

Lawless, J. F. and Singhal, K. (1978). Efficient screening of nonnormal regression models. *Biometrics*, **34**, 318–327.

Lawless, J. F., Hu, J., and Cao, J. (1995). Methods for the estimation of failure distributions and rates from automobile warranty data. *Lifetime Data Anal.*, **1**, 227–240.

Lawless, J. F., Kalbfleisch, J. D., and Wild, C. J. (1999). Semiparametric methods for response-selective and missing data problems in regression. *J. Roy. Stat. Soc. B*, **61**, 413–438.

Lawless, J. F., Wigg, M. B., Tuli, S., Drake, J., and Lamberti-Pasculli, M. (2001). Analysis of repeated failures or durations, with application to shunt failures for patients with paediatric hydrocephalus. *Appl. Stat.*, **50**, 449–465.

Lee, E. T., Desu, M. M., and Gehan, E. A. (1975). A Monte Carlo study of the power of some two-sample tests. *Biometrika*, **62**, 425–432.

Lee, E. W., Wei, L. J., and Amato, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis: State of the Art*, J. P. Klein and P. J. Goel, Eds., pp. 237–247. Kluwer Academic, Dordrecht.

Lee, E. W., Wei, L. J., and Ying, Z. (1993). Linear regression analysis for highly stratified failure time data. *J. Amer. Stat. Assoc.*, **88**, 557–565.

Lee, K. C., Harrell, F. E., Tolley, H. D., and Rosati, R. A. (1983). A comparison of test statistics for assessing the effects of concomitant variables in survival analysis. *Biometrics*, **39**, 341–350.

Lee, M. L. T., De Gruttola, V., and Schoenfeld, D. (2000). A model for markers and latent health status. *J. Roy. Stat. Soc. B*, **62**, 747–762.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco, CA.

Lehmann, E. L. (1983). *Theory of Point Estimation*. John Wiley & Sons, New York.

Lemon, G. (1975). Maximum likelihood estimation for the three-parameter Weibull distribution, based on censored samples. *Technometrics*, **17**, 247–254.

Leurgans, S. (1983). Three classes of censored data rank tests: strengths and weaknesses under censoring. *Biometrika*, **70**, 654–658.

Leurgans, S. (1984). Asymptotic behavior of two-sample rank tests in the presence of random censoring. *Ann. Stat.*, **12**, 572–589.

Li, G. and Doss, H. (1993). Generalized Pearson-Fisher chi-square goodness-of-fit tests, with application to models with life history data. *Ann. Stat.*, **21**, 772–797.

Liang, K.-Y., Self, S. G., and Chang, Y. C. (1993). Modelling marginal hazards in multivariate failure time data. *J. Roy. Stat. Soc. B*, **55**, 441–453.

Liang, K.-Y., Self, S. G., Bandeen-Roche, K. J., and Zeger, S. L. (1995). Some recent developments for regression analysis of multivariate failure time data. *Lifetime Data Anal.*, **1**, 403–415.

Lieblein, J. and Zelen, M. (1956). Statistical investigation of the fatigue life of deep groove ball bearings. *J. Res. Nat. Bur. Stand.*, **57**, 273–316.

Lin, D. Y. (1994). Cox regression analysis of mulvariate failure time data: The marginal approach. *Stat. Med.*, **13**, 2233–2247.

Lin, D. Y. (2000). On fitting Cox's proportional hazards model to survey data. *Biometrika*, **87**, 37–47.

Lin, D. Y. and Geyer, C. J. (1992). Computational methods for semiparametric linear regression with censored data. *J. Comput. Graph. Stat.*, **1**, 77–90.

Lin, D. Y. and Spiekerman, C. F. (1996). Model checking techniques for parametric regression with censored data. *Scand. J. Stat.*, **23**, 157–177.

Lin, D. Y. and Wei, L. J. (1991). Goodness-of-fit tests for the general Cox regression model. *Stat. Sin.*, **1**, 1–17.

Lin, D. Y. and Ying, Z. (1995). Semiparametric inference for the accelerated life model with time-dependent covariates. *J. Stat. Plan. Inf.*, **44**, 47–63.

Lin, D. Y. and Ying, Z. (1997). Additive hazards regression models for survival data. In *Proceedings of the 1st Seattle Symposium in Biostatistics*, D. Y. Lin and T. R. Fleming, Eds., pp. 184–198. Springer-Verlag, New York.

Lin, D. Y., Wei, L. J., and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, **80**, 557–572.

Lin, D. Y., Feuer, E. J., Etzioni, R., and Wax, Y. (1997). Estimating medical costs from incomplete followup data. *Biometrics*, **84**, 339–348.

Lin, D. Y., Sun, W., and Ying, Z. (1999). Nonparametric estimation of the gap time distribution for serial events with censored data. *Biometrika*, **86**, 59–70.

Lindeboom, M. and Van Den Berg, G. J. (1994). Heterogeneity in models for bivariate survival: The importance of the mixing distribution. *J. Roy. Stat. Soc. B*, **56**, 49–60.

Lindsey, J. C. and Ryan, L. M. (1993). A three-state multiplicative model for rodent tumorigenicity experiments. *Appl. Stat.*, **42**, 283–300.

Lindsey, J. C. and Ryan, L. M. (1998). Tutorial in biostatistics: Methods for interval-censored data. *Stat. Med.*, **17**, 219–238.

Lindsey, J. K. (1998). A study of interval censoring in parametric regression models. *Lifetime Data Anal.*, **4**, 329–354.

Lininger, L., Gail, M., Green, S., and Byar, D. (1979). Comparison of four tests for equality of survival curves in the presence of stratification and censoring. *Biometrika*, **66**, 419–428.

Link, C. L. (1984). Confidence intervals for the survival function using Cox's proportional-hazard model with covariates. *Biometrics*, **40**, 601–610.

Littell, A. S. (1952). Estimation of the T-year survival rate from followup studies over a limited period of time. *Hum. Biol.*, **24**, 87–116.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis of Missing Data*. John Wiley & Sons, New York.

Lloyd, E. H. (1952). Least-squares estimation of location and scale parameters using order statistics. *Biometrika*, **39**, 88–95.

Loader, C. (1999). *Local Regression and Likelihood*. Springer-Verlag, New York.

Lockhart, R. A. and Stephens, M. A. (1994). Estimation and tests of fit for the three-parameter Weibull distribution. *J. Roy. Stat. Soc. B*, **56**, 491–500.

Louis, T. A. (1977). Sequential allocation in clinical trials comparing two exponenetial curves. *Biometrics*, **33**, 627–634.

Louis, T. A. (1981). Nonparametric analysis of an accelerated failure time model. *Biometrika*, **68**, 381–390.

Loynes, R. M. (1980). The empirical distribution function of residuals from generalized regression. *Ann. Stat.*, **8**, 285–298.

Lu, C. J. and Meeker, W. Q. (1993). Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, **34**, 161–174.

Lu, J. C., Park, J., and Yang, Q. (1997). Statistical inference of a time-to-failure distribution derived from linear degradation data. *Technometrics*, **39**, 391–400.

Lu, Jin (1995). A reliability model based on degradation and lifetime data, Ph.D. Thesis, McGill University, Montreal, P.Q., Canada.

Lunn, M. and McNeil, D. (1995). Applying Cox regression to competing risks. *Biometrics*, **51**, 524–532.

LuValle, M. J. (1993). Experimental design and graphical analysis for checking acceleration models. *Microelectron Reliab.*, **33**, 741–763.

LuValle, M. J., Welsher, T. L., and Svoboda, K. (1988). Acceleration transforms and statistical kinetic models. *J. Stat. Phys.*, **52**, 311–320.

Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Mon. Not. Roy. Astron. Soc.*, **155**, 95–118.

Machado, S. G. and Bailey, K. R. (1985). Assessment of interaction between carcinogens in long-term factorially designed animal experiments. *Biometrics*, **41**, 539–545.

Mackenzie, G. (1996). Regression models for survival data: The generalized time-dependent logistic family. *The Statistician*, **45**, 21–34.

Maller, R. A. and Zhou, S. (1996). *Survival Analysis With Long Term Survivors*. John Wiley & Sons, New York.

Manatunga, A. K. and Oakes, D. (1999). Parametric analysis for matched pair survival data. *Lifetime Data Anal.*, **5**, 371–387.

Mann, N. R. (1968). Point and interval estimation procedures for the two-parameter Weibull and extreme value distributions. *Technometrics*, **10**, 231–256.

Mann, N. R. (1969). Optimum estimators for linear functions of location and scale parameters. *Ann. Math. Stat.*, **40**, 2149–2155.

Mann, N. R. (1977). An F approximation for two-parameter Weibull and log-normal tolerance bounds based on possibly censored data. *Nav. Res. Logist. Q.*, **24**, 187–196.

Mann, N. R. and Fertig, K. W. (1973). Tables for obtaining confidence bounds and tolerance bounds based on best linear invariant estimates of parameters of the extreme value distribution. *Technometrics*, **15**, 87–101.

Mann, N. R. and Fertig, K. W. (1975). A goodness-of-fit test for the two parameter vs. three parameter Weibull; confidence bounds for the threshold. *Technometrics*, **17**, 237–245.

Mann, N. R., Scheuer, E. M., and Fertig, K. W. (1973). A new goodness of fit test for the two-parameter Weibull or extreme value distribution. *Commun. Stat.*, **2**, 383–400.

Mann, N. R., Schafer, R. E., and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Lifetime Data*. John Wiley & Sons, New York.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *J. Amer. Stat. Assoc.*, **58**, 690–700.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Rep.*, **50**, 163–170.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from restrospective studies of disease. *J. Nat. Cancer Inst.*, **22**, 719–748.

Mantel, N., Bohidar, N. R., and Ciminera, J. L. (1977). Mantel-Haenszel analyses of litter-matched time-to-response data, with modifications for recovery of interlitter information. *Cancer Res.*, **37**, 3863–3868.

Manton, K. G. and Stallard, E. (1988). *Chronic Disease Modelling*. Griffin, London.

Marshall, A. W. and Olkin, I. (1988). Families of multivariate distributions. *J. Amer. Stat. Assoc.*, **83**, 834–841.

Martz, H. F. and Waller, R. A. (1982). *Bayesian Reliability Analysis*. John Wiley & Sons, New York.

Marzec, L. and Marzec, P. (1997). Generalized martingale-residual processes for goodness-of-fit inference in Cox's type regression models. *Ann. Stat.*, **25**, 683–714.

Matthews, D. E. (1988). Likelihood-based confidence intervals for functions of many parameters. *Biometrika*, **75**, 139–144.

May, S. and Hosmer, D. W. (1998). A simplified method of calculating an overall goodness-of-fit test for the Cox proportional hazards model. *Lifetime Data Anal.*, **4**, 109–120.

McCool, J. I. (1970). Inferences on Weibull percentiles and shape parameter for maximum likelihood estimates. *IEEE Trans. Reliab.*, **R-19**, 2–9.

McCool, J. I. (1974). Inferential techniques for Weibull populations. Wright-Patterson Air Force Base, Ohio, ARL Technical Report ARL-74-0180.

McCool, J. I. (1975a). Multiple comparisons for Weibull parameters. *IEEE Trans. Reliab.*, **R-24**, 186–192.

McCool, J. I. (1975b). Inferential techniques for Weibull populations II. Wright-Patterson Air Force Base, Ohio, ARL Technical Report ARL-75-0233.

McCool, J. I. (1979). Analysis of single classificiation experiments based on censored samples from the two-parameter Weibull distribution. *J. Stat. Plan. Inf.*, **3**, 39–68.

McCool, J. I. (1980). Confidence limits for Weibull regression with censored data. *IEEE Trans. Reliab.*, **R-29**, 145–150.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman & Hall, London.

McGilchrist, C. A. (1993). REML estimation for survival models with frailty. *Biometrics*, **49**, 221–225.

McGilchrist, C. A. and Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics*, **47**, 461–466.

McKeague, I. W. and Utikal, K. J. (1991). Goodness-of-fit tests for additive hazards and proportional hazards models. *Scand. J. Stat.*, **18**, 177–195.

Meeker, W. Q. (2002). *SPLIDA User's Manual*. In preparation.

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. John Wiley & Sons, New York.

Meeker, W. Q. and LuValle, M. J. (1995). An accelerated life test model based on reliability kinetics. *Technometrics*, **37**, 133–146.

Meeker, W. Q. and Nelson, W. (1977). Weibull variances and confidence limits by maximum likelihood for singly censored data. *Technometrics*, **19**, 473–476.

Meeter, C. A. and Meeker, W. Q. (1994). Optimum accelerated life tests with a nonconstant scale parameter. *Technometrics*, **36**, 71–83.

Mehrotra, K. G., Johnson, R. A., and Bhattacharya, G. K. (1977). Locally most powerful rank tests for multiple-censored data. *Commun. Stat.*, **A6**, 459–469.

Mehrotra, K. G., Michalek, J. E., and Mihalko, D. (1982). A relationship between two forms of linear rank procedures for censored data. *Biometrika*, **69**, 674–676.

Meier, P. (1975). Estimation of a distribution function from incomplete observations. In *Perspectives in Probability and Statistics*, J. Gani, Ed. Applied Probability Trust, Sheffield, England.

Mendenhall, W. (1958). A bibliography on life testing and related topics. *Biometrika*, **45**, 521–543.

Mendenhall, W. and Hader, R. J. (1958). Estimation of parameters of mixed exponential distributed failure times from censored life test data. *Biometrika*, **45**, 504–520.

Miller, R. G. (1976). Least squares regression with censored data. *Biometrika*, **63**, 449–464.

Miller, R. G. (1983). What price Kaplan–Meier? *Biometrics*, **39**, 1077–1081.

Miyakawa, M. (1984). Analysis of incomplete data in a competing risks model. *IEEE Trans. Reliab.*, **R-33**, 293–296.

Moertel, C. G., Fleming, T. R., and McDonald, J. S. (1990). Levamisole and Fluorouracil for adjuvant therapy of restricted colon carcinoma. *N. Engl. J. Med.*, **322**, 352–358.

Moeschberger, M. L. (1974). Life tests under competing causes of failure. *Technometrics*, **16**, 39–47.

Moreau, T., O'Quigley, J., and Mesbah, M. (1985). A global goodness-of-fit statistic for the proportional hazards model. *Appl. Stat.*, **34**, 212–218.

Moreau, T., O'Quigley, J., and Lellouch, J. (1986). On D. Schoenfeld's approach for testing the proportional hazards assumption. *Biometrika*, **73**, 513–515.

Morton, R. (1978). Regression analysis of life tables and related nonparametric tests. *Biometrika*, **65**, 329–333.

Muller, H. G., Wang, J. L., and Copra, W. B. (1997). From lifetables to hazard rates: The transformation approach. *Biometrika*, **84**, 881–892.

Murphy, S. A. and van der Vaart, A. W. (1999). On profile likelihood (with discussion). *J. Amer. Stat. Assoc.*, **95**, 449–485.

Naes, T. (1982). The asymptotic distribution of the estimator for the regression parameter in Cox's regression model. *Scand. J. Stat.*, **9**, 107–115.

Nagelkerke, N. J. D., Oosting, J., and Hart, A. A. M. (1984). A simple test for goodness of fit of Cox's proportional hazards model. *Biometrics*, **40**, 483–486.

Nair, V. N. (1981). Plots and tests for goodness of fit with randomly censored data. *Biometrika*, **68**, 99–103.

Nair, V. N. (1984). Confidence bands for survival functions with censored data: A comparative study. *Technometrics*, **14**, 265–275.

Namboodiri, K. and Suchindran, C. M. (1987). *Life Table Techniques and Their Applications*. Academic Press, Orlando.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Comput. J.*, **7**, 380–383.

Nelson, W. B. (1969). Hazard plotting for incomplete failure data. *J. Qual. Technol.*, **1**, 27–52.

Nelson, W. B. (1970a). Statistical methods for accelerated lifetest data—the inverse power law model. General Electric Co. Technical Report 71-C-011, Schenectady, NY.

Nelson, W. B. (1970b). Hazard plotting methods for analysis of life data with different failure modes. *J. Qual. Technol.*, **2**, 126–149.

Nelson, W. B. (1972a). Graphical analysis of accelerated life test data with the inverse power law model. *IEEE Trans. Reliab.*, **R-21**, 2–11.

Nelson, W. B. (1972b). Theory and applications of hazard plotting for censored failure data. *Technometrics*, **14**, 945–965.

Nelson, W. B. (1982). *Applied Life Data Analysis*. John Wiley & Sons, New York.

Nelson, W. B. (1984). Fitting of fatigue curves with nonconstant standard deviation to data with runouts. *J. Test. Eval.*, **12**, 69–77.

Nelson, W. B. (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*. John Wiley & Sons, New York.

Nelson, W. B. and Doganaksoy, N. (1995). Statistical analysis of life or strength data from specimens of various sizes using the power-(log) normal model. In *Recent Advances in Life-Testing and Reliability*, N. Balakrishnan, Ed., pp. 377–408. CRC Press, Boca Raton.

Nelson, W. B. and Hahn, G. J. (1972). Linear estimation of a regression relationship from censored data. Part I—simple methods and their applications. *Technometrics*, **14**, 247–269.

Nelson, W. B. and Kielpinski, T. (1976). Theory for optimum censored accelerated life tests for normal and lognormal distributions. *Technometrics*, **18**, 105–114.

Nelson, W. B. and Meeker, W. Q. (1978). Theory for optimum accelerated censored life tests for Weibull and extreme value distributions. *Technometrics*, **20**, 171–177.

Nelson, W. B. and Schmee, J. (1979). Inference for (log) normal life distributions from small singly censored samples and blue's. *Technometrics*, **21**, 43–54.

Neuhaus, J. and Kalbfleisch, J. D. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics*, **54**, 628–645.

Neyman, J. (1937), 'Smooth' tests for goodness of fit. *Skand. Aktuarietiosk.*, **20**, 149–199.

Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sorensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Stat.*, **19**, 25–43.

Oakes, D. (1977). The asymptotic information in censored survival data. *Biometrika*, **64**, 441–448.

Oakes, D. (1982). A model for association in bivariate survival data. *J. Roy. Stat. Soc. B*, **44**, 414–422.

Oakes, D. (1989). Bivariate survival models induced by frailties. *J. Amer. Stat. Assoc.*, **84**, 487–493.

Oakes, D. (1995). Multiple time scales in survival analysis. *Lifetime Data Anal.*, **1**, 7–18.

Oakes, D. (2001). Biometrika Centenary: Survival analysis. *Biometrika*, **88**, 99–142.

Oleinick, A. and Mantel, N. (1970). Family studies in sytematic lupus erythematosis: *J. Chronic Dis.*, **22**, 617–625.

O'Quigley, J. and Pessione, F. (1989). Score tests for homogeneity of regression effect in the proportional hazards model. *Biometrics*, **45**, 135–144.

O'Reilly, F. J. and Rueda, R. (1992). Goodness of fit for the inverse Gaussian distribution. *Can. J. Stat.*, **20**, 387–397.

O'Sullivan, F. (1988). Fast computation of fully automated log-density and log-hazard estimators. *SIAM J. Sci. Stat. Comput.*, **9**, 363–379.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, Boca Raton.

Owen, D. B. (1968). A survey of properties and applications of the noncentral t-distribution. *Technometrics*, **10**, 445–478.

Park, H.-I. (1997). A note on the relation between two forms of linear rank statistics for right censored and grouped data. *Biometrika*, **84**, 987–988.

Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Clarendon Press, Oxford.

Peace, K. and Flora, R. (1978). Size and power assessments of tests of hypotheses on survival parameters. *J. Amer. Stat. Assoc.*, **73**, 129–132.

Pearson, E. S. and Hartley, H. O. (1972). *Biometrika Tables for Statisticians*, Vol. 2. Cambridge University Press, Cambridge, England.

Pena, E. A. (1998). Smooth goodness-of-fit tests for composite hypothesis in hazard based models. *Ann. Stat.*, **26**, 1935–1971.

Pepe, M. S. and Fleming, T. R. (1989). Weighted Kaplan–Meier statistics: A class of distance tests for censored survival data. *Biometrics*, **45**, 497–507.

Pepe, M. S. and Fleming, T. R. (1991). Weighted Kaplan–Meier statistics: Large sample and optimality considerations. *J. Roy. Stat. Soc. B*, **53**, 341–352.

Pepe, M. S. and Mori, M. (1993). Kaplan–Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Stat. Med.*, **12**, 737–751.

Persson, T. and Rootzen, H. (1977). Simple and highly efficient estimators for a Type I censored normal sample. *Biometrika*, **64**, 123–128.

Petersen, J. H., Andersen, P. K., and Gill, R. D. (1996). Variance components models for survival data. *Stat. Neerl.*, **50**, 193–211.

Peterson, A. V. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: Applications to competing risks. *Proc. Natl. Acad. Sci. U.S.A.*, **73**, 11–13.

Peterson, A. V. (1977). Expressing the Kaplan–Meier estimator as a function of empirical subsurvival functions. *J. Amer. Stat. Assoc.*, **72**, 854–858.

Peto, R. (1972). Discussion of paper by D. R. Cox. *J. Roy. Stat. Soc. B*, **34**, 205–207.

Peto, R. (1973). Experimental survival curves for interval censored data. *Appl. Stat.*, **22**, 86–91.

Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant procedures (with discussion). *J. Roy. Stat. Soc. A*, **135**, 185–206.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J., and Smith P. G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. Part I: Introduction and design. *Br. J. Cancer*, **34**, 585–612.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J., and Smith P. G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. Part II: Analysis and examples. *Br. J. Cancer*, **35**, 1–39.

Pettitt, A. N. and Bin Daud, I. (1989). Case-weighted measures of influence for proportional hazards regression. *Appl. Stat.*, **38**, 51–67.

Piantadosi, J. (1997). *Clinical Trials: A Methodologic Perspective*. John Wiley & Sons, New York.

Pierce, D. A. (1973). Fiducial, frequency and Bayesian inference on reliability for the two-parameter negative exponential distribution. *Technometrics*, **15**, 249–253.

Pierce, D. A. and Kopecky, K. J. (1979). Testing goodness of fit for the distribution of errors in regression models. *Biometrika*, **66**, 1–5.

Pierce, D. A., Stewart, W. H., and Kopecky, K. J. (1979). Distribution-free analysis of grouped survival data. *Biometrics*, **35**, 785–793.

Pike, M. C. (1966). A method of analysis of certain classes of experiments in carcinogenesis. *Biometrics*, **22**, 142–161.

Pike, M. C. (1970). A note on Kimball's paper "Models for the estimation of competing risks from grouped data." *Biometrics*, **26**, 579–581.

Prentice, R. L. (1973). Exponential survival with censoring and explanatory variables. *Biometrika*, **60**, 279–288.

Prentice, R. L. (1974). A log gamma model and its maximum likelihood estimation. *Biometrika*, **61**, 539–544.

Prentice, R. L. (1975). Discrimination among some parametric models. *Biometrika*, **62**, 607–614.

Prentice, R. L. (1978). Linear rank tests with right-censored data. *Biometrika*, **65**, 167–179.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, **73**, 1–11.

Prentice, R. L. and Gloeckler, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, **34**, 57–67.

Prentice, R. L. and Hsu, L. (1997). Regression on hazard ratios and cross ratios in multivariate failure time analysis. *Biometrika*, **84**, 349–363.

Prentice, R. L. and Marek, P. (1979). A qualitative discrepancy between censored data rank tests. *Biometrics*, **35**, 861–867.

Prentice, R. L. and Self, S. G. (1983). Asymptotic distribution theory for Cox-type regressions models with general relative risk form. *Ann. Stat.*, **11**, 804–813.

Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*, **47**, 825–839.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, **34**, 541–554.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, **68**, 373–379.

Press, W. H., Fleming, B. P., Teukolsky, S. A., and Vetterling, W. T. (1986). *Numerical Recipes*. Cambridge University Press, Cambridge.

Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics*, **5**, 375–383.

Pyke, R. (1965). Spacings (with discussion). *J. Roy. Stat. Soc. B*, **27**, 395–449.

Quantin, C., Moreau, T., Asselain, B., Maccario, S., and Lellouch, J. (1996). A regression survival model for testing the proportional hazards hypothesis. *Biometrics*, **52**, 874–885.

Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Stat.*, **11**, 453–466.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. John Wiley & Sons, New York.

Rayner, J. and Best, D. (1989). *Smooth Tests of Goodness of Fit*. Oxford University Press, Oxford.

Rayner, J. and Best, D. (1990). Smooth tests of goodness of fit: An overview. *Int. Stat. Rev.*, **58**, 9–17.

Reid, N. (2000). Likelihood. *J. Amer. Stat. Assoc.*, **95**, 1335–1340.

Reid, N. and Crépeau, H. (1985). Influence functions for proportional hazards regression. *Biometrika*, **72**, 1–9.

Rice, J. and Rosenblatt, M. (1976). Estimation of the log survivor function and hazard function. *Sankhyā Ser. A*, **38**, 60–78.

Richards, F. S. G. (1961). A method of maximum-likelihood estimation. *J. Roy. Stat. Soc. B*, **23**, 469–476.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Stat.*, **18**, 303–328.

Robins, J. and Tsiatis, A. A. (1992). Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika*, **79**, 311–319.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton, NJ.

Rockette, H., Antle, C., and Klimko, L. (1974). Maximum likelihood estimation with the Weibull model. *J. Amer. Stat. Assoc.*, **69**, 246–249.

Rosenberg, P. S. (1995). Hazard function estimation using B-splines. *Biometrics*, **51**, 874–887.

Ross, S. M. (1983). *Stochastic Processes*. John Wiley & Sons, New York.

Rubinstein, L. V., Gail, M. H., and Santner, T. J. (1981). Planning the duration of a clinical trial with loss to follow-up and a period of continued observation. *J. Chronic Dis.*, **34**, 469–479.

Ryan, T. P. (1997). *Modern Regression Methods*. John Wiley & Sons, New York.

Sacher, G. A. (1956). On the statistical nature of mortality, with special reference to chronic radiation mortality. *Radiation*, **67**, 250–257.

Sampford, M. R. and Taylor, J. (1959). Censored observations in randomized block experiments. *J. Roy. Stat. Soc. B*, **21**, 214–237.

Sarhan, A. E. and Greenberg, B. G. (1962). *Contributions to Order Statistics*. John Wiley & Sons, New York.

Savage, I. R. (1956). Contributions to the theory of rank order statistics—the two sample case. *Ann. Math. Stat.*, **27**, 590–615.

Schafer, R. E. and Sheffield, T. S. (1976). On procedures for comparing two Weibull populations. *Technometrics*, **18**, 231–235.

Schafft, H. A., Staton, T. C., Mandel, J., and Shott, J. D. (1987). Reproduceability of electromigration measurements. *IEEE Trans. Electron Devices*, **ED-34**, 673–681.

Schmee, J. and Nelson, W. (1977). Estimates and approximate confidence limits for (log) normal life distributions from singly censored samples by maximum likelihood. General Electric C. R. & D. TIS Report 76CRD250. Schenectady, New York.

Schneider, H. (1989). Failure-censored variables-sampling plans for lognormal and Weibull distributions. *Technometrics*, **14**, 679–691.

Schoenfeld, D. A. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika*, **67**, 145–153.

Schoenfeld, D. A. (1981). The asymptotic properties of comparative tests for comparing survival distributions. *Biometrika*, **68**, 316–319.

Schoenfeld, D. A. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, **69**, 239–241.

Schoenfeld, D. A. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*, **39**, 499–503.

Scholz, F. W. (1980). Towards a unified definition of maximum likelihood. *Can. J. Stat.*, **8**, 193–203.

Schumacher, M. (1984). Two-sample tests of Cramer-von Mises and Kolmogorov-Smirnov type for randomly censored data. *Int. Stat. Rev.*, **52**, 263–281.

Seal, H. L. (1977). Studies in the history of probability and statistics XXXV. Multiple decrements or competing risks. *Biometrika*, **64**, 429–439.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. John Wiley & Sons, New York.

Segall, A. and Kailath, T. (1975). The modeling of randomly modulated jump processes. *IEEE Trans. Inf. Theory*, **IT-21**, 135–143.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Stat. Assoc.*, **82**, 605–610.

Self, S. and Pawitan, Y. (1992). Modeling a marker of disease progression and onset of disease. In *AIDS Epidemiology: Methodological Issues*, N. Jewell, K. Dietz, and V. Farewell, Eds. Birkhauser, Boston.

Sellke, T. and Siegmund, D. (1983). Sequential analysis of the proportional hazards model. *Biometrika*, **70**, 315–326.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

Severini, T. (2000). *Likelihood Methods in Statistics*. Oxford University Press, Oxford.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591–611.

Shellabarger, C. J., McKnight, B., Stone, J. P., and Holtzman, S. (1980). Interaction of dimethylbenzanthracene and diethylstilbestrol on mammary adenocarcinoma formation in female ACI rats. *Cancer Res.*, **40**, 1808–1811.

Shi, M., Taylor, J. M. G., and Munoz, A. (1996). Models for residual time to AIDS. *Lifetime Data Anal.*, **2**, 31–49.

Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, **51**, 1384–1399.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes*. John Wiley & Sons, New York.

Silvapulle, M. J. (1985). Asymptotic behavior of robust estimators of regression and scale parameters with fixed carriers. *Ann. Stat.*, **13**, 1490–1497.

Silvapulle, M. J. and Burridge, J. (1986). Existence of maximum likelihood estimates in regression models for grouped and ungrouped data. *J. Roy. Stat. Soc. B*, **48**, 100–106.

Singpurwalla, N. D. (1995). Survival in dynamic environments. *Stat. Sci.*, **10**, 86–103.

Skouras, K. and David, A. P. (1998). On efficient point prediction systems. *J. Roy. Stat. Soc. B*, **60**, 765–780.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, **72**, 67–90.

Smith, R. L. (1995). Likelihood and modified likelihood estimation for distributions with unknown endpoints. In *Recent Advances in Life-Testing and Reliability*, N. Balakrishnan, Ed., pp. 455–474. CRC Press, Boca Raton.

Smyth, G. K. (1989). Generalized linear models with varying dispersion. *J. Roy. Stat. Soc. B*, **51**, 47–60.

Sobczyk, K. (1987). Stochastic models for fatigue damage of material. *Adv. Appl. Prob.*, **19**, 652–673.

Spiekerman, C. F. and Lin, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Amer. Stat. Assoc.*, **93**, 1164–1175.

Sprott, D. A. (1973). Normal likelihoods and relation to a large sample theory of estimation. *Biometrika*, **60**, 457–465.

Stacy, E. W. (1962). A generalization of the gamma distribution. *Ann. Math. Stat.*, **33**, 1187–1192.

Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *J. Amer. Stat. Assoc.*, **69**, 730–737.

Stone, G. C. (1978). Statistical analysis of accelerated aging tests on solid electrical insulation. Unpublished M.A.Sc. Thesis, University of Waterloo, Waterloo, Ont., Canada.

Stone, G. C. and Lawless, J. F. (1979). The application of Weibull statistics to insulation aging tests. *IEEE Trans. Elec. Insul.*, **EI-14**, 233–239.

Storer, B. E. and Crowley, J. (1985). A diagnostic for Cox regression and general conditional likelihoods. *J. Amer. Stat. Assoc.*, **80**, 139–147.

Strawderman, R. L. and Wells, M. T. (1997). Accurate bootstrap confidence limits for the cumulative hazard and survivor functions under random censoring. *J. Amer. Stat. Assoc.*, **92**, 1356–1374.

Struthers, C. A. (1984). Asymptotic properties of linear rank tests with censored data. Unpublished Ph.D. Thesis, University of Waterloo, Waterloo, Ont., Canada.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazards models. *Biometrika*, **73**, 363–369.

Sukhatme, P. V. (1937). Tests of significance for samples of the χ^2 population with two degrees of freedom. *Ann. Eugen.*, **8**, 52–56.

Sun, J. (1995). Empirical estimation of a distribution function with truncation and doubly interval-censored data and its application to AIDS studies. *Biometrics*, **51**, 1096–1104.

Sun, J. (1997). Self-consistency estimation of distributions based on truncated doubly censored survival data with applications to AIDS cohort studies. *Lifetime Data Anal.*, **3**, 305–313.

Sun, Y. (1997). Weak convergence of the generalized parametric empirical processes and goodness-of-fit tests for parametric models. *Commun. Stat.—Theor. Meth.*, **26**, 2393–2413.

Sun, Y., Chow, S.-C., Li, G., and Chen, K.-W. (1999). Assessing distributions of estimated drug shelf lives in stability analysis. *Biometrics*, **55**, 203–206.

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of mean survival curves from incomplete observations. *J. Amer. Stat. Assoc.*, **61**, 897–902.

Susarla, V. and Van Ryzin, J. (1980). Large sample theory for an estimator of the mean survival time from censored samples. *Ann. Stat.*, **8**, 1002–1016.

Suzuki, K. (1985a). Nonparametric estimation of lifetime distribution from a record of failures and follow-ups. *J. Amer. Stat. Assoc.*, **72**, 854–858.

Suzuki, K. (1985b). Estimation of lifetime parameters from incomplete field data. *Technometrics*, **27**, 263–271.

Suzuki, K. (1995). Role of field performance and its analysis. In *Recent Advances in Life-Testing and Reliability*, N. Balakrishnan, Ed., pp. 141–154. CRC Press, Boca Raton.

Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics*, **56**, 227–236.

Tadikamalla, P. R. (1980). A look at the Burr and related distributions. *Int. Stat. Rev.*, **48**, 337–344.

Taguchi, G. (1986). *Introduction to Quality Engineering*. Asian Productivity Organisation, Tokyo.

Taguchi, G. (1987). *Introduction to Off-Line Quality Control*. Unipub/Kraus International Publications. White Plains, NY.

Tanner, M. A. and Wong, W. H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Stat.*, **11**, 994–998.

Tanner, M. A. and Wong, W. H. (1984). Data-based nonparametric estimation of the hazard function with applications to model diagnostics and exploratory analysis. *J. Amer. Stat. Assoc.*, **79**, 174–182.

Tanner, M. A. and Wong, W. H. (1987). An application of imputation to an estimation problem in grouped lifetime analysis. *Technometrics*, **29**, 23–32.

Tarone, R. E. (1975). Tests for trend in life table analysis. *Biometrika*, **62**, 679–682.

Tarone, R. E. (1981). On the distribution of the maximum of the logrank statistic and the modified Wilcoxon statistic. *Biometrics*, **37**, 79–85.

Tarone, R. E. and Ware, J. (1977). On distribution-free tests for equality of survival distributions. *Biometrika*, **64**, 156–160.

Taylor, H. M. (1975). Optimal replacement under additive damage and other failure models. *Nav. Res. Logist. Q.*, **22**, 1–18.

Taylor, J. M. G., Cumberland, W. G., and Sy, J. P. (1994). A stochastic model for analysis of longitudinal AIDS data. *J. Amer. Stat. Assoc.*, **89**, 727–736.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York.

Therneau, T. M. and Hamilton, S. A. (1997). rhDNase as an example of recurrent event analysis. *Stat. Med.*, **16**, 2029–2047.

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika*, **77**, 147–160.

Thoman, D. R. and Bain, L. J. (1969). Two-sample tests in the Weibull distribution. *Technometrics*, **11**, 805–816.

Thoman, D. R., Bain, L. J., and Antle, C. E. (1969). Inferences on the parameters of the Weibull distribution. *Technometrics*, **11**, 445–460.

Thoman, D. R., Bain, L. J., and Antle, C. E. (1970). Reliability and tolerance limits in the Weibull distribution. *Technometrics*, **12**, 363–371.

Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *J. Amer. Stat. Assoc.*, **70**, 865–871.

Thomas, D. R. and Pierce, D. A. (1979). Neyman's smooth goodness of fit tests when the hypothesis is composite. *J. Amer. Stat. Assoc.*, **74**, 441–445.

Thompson, W. A. (1977). On the treatment of grouped observations in life studies. *Biometrics*, **33**, 463–470.

Tibshirani, R. J. and Ciampi, A. (1983). A family of proportional- and additive-hazards models for survival data. *Biometrics*, **39**, 141–148.

Tiku, M. L. (1981). Goodness of fit statistics based on the spacings of complete or censored samples. *Aust. J. Stat.*, **22**, 260–275.

Tiku, M. L. and Singh, M. (1981). Testing the two parameter Weibull distribution. *Commun. Stat. A.*, **10**, 907–918.

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons, New York.

Trussell, J., Hankinson, R., and Tilton, J. (1992). *Demographic Applications of Event History Analysis*. Clarendon Press, Oxford.

Tsai, W.-Y. (1990). Testing the assumption of independence of truncation time and failure time. *Biometrika*, **77**, 169–177.

Tseng, S. T., Hamada, M., and Chiao, C. H. (1995). Using degradation data from a factorial experiment to improve fluorescent lamp reliability. *J. Qual. Technol.*, **27**, 363–369.

Tsiatis, A. A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 20–22.

Tsiatis, A. A. (1981a). A large sample study of Cox's regression model. *Ann. Stat.*, **9**, 93–108.

Tsiatis, A. A. (1981b). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika*, **68**, 311–315.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Stat.*, **18**, 354–372.

Tsiatis, A. A., DeGruttola, V., and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *J. Amer. Stat. Assoc.*, **90**, 27–37.

Tuli, S., Drake, J., Lawless, J., Wigg, M., and Lamberti-Pasculli, M. (2000). Risk factors for repeat cerebrospinal shunt failures in pediatric hydrocephalus. *J. Neurosurg.*, **92**, 31–38.

Tuma, N. B. and Hannan, M. T. (1984). *Social Dynamics*. Academic Press, Orlando, FL.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Stat. Soc. B*, **38**, 290–295.

Turnbull, B. W. and Weiss, L. (1978). A likelihood ratio statistic for testing goodness of fit with randomly censored data. *Biometrics*, **34**, 367–375.

Tutz, G. and Pritscher, L. (1996). Nonparametric estimation of discrete hazard functions. *Lifetime Data Anal.*, **2**, 291–308.

Valdez-Flores, C. and Feldman, R. M. (1989). A survey of preventive maintenance models for stochastically deteriorating single-unit systems. *Nav. Res. Logist. Q.*, **36**, 419–446.

Van der Laan, M. J. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. *Ann. Stat.*, **24**, 596–627.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, **16**, 439–454.

Venables, W. N. and Ripley, B. D. (1999). *Modern Applied Statistics with S-Plus*, 3rd ed. Springer-Verlag, New York.

Venzon, A. and Moolgavkar, S. (1988). A method for computing profile likelihood-based confidence intervals. *Appl. Stat.*, **37**, 87–94.

Verhagen, A. M. W. (1961). The estimation of regression and error-scale parameters when the joint distribution of the errors is of any continuous form and known apart from a scale parameter. *Biometrika*, **48**, 125–132.

Verweij, P. J. M., van Houwelingen, H. C. and Stijnen, T. (1998). A goodness-of-fit test for Cox's proportional hazards model based on martingales. *Biometrics*, **54**, 1517–1526.

Visser, M. (1996). Nonparametric estimation of the bivariate survival function with an application to vertically transmitted AIDS. *Biometrika*, **83**, 507–518.

Viveros, R. and Balakrishnan, N. (1994). Interval estimation of parameters of life distributions from progressively censored data. *Technometrics*, **36**, 84–91.

Wang, M.-C., Jewell, N. P., and Tsai, W.-Y. (1986). Asymptotic properties of the product limit estimate under random truncation. *Ann. Stat.*, **14**, 1597–1605.

Wang, W. and Wells, M. T. (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika*, **85**, 561–572.

Watson, A. S. and Smith, R. L. (1985). An examination of statistical theories for fibrous materials in light of experimental data. *J. Mater. Sci.*, **20**, 3260–3270.

Watson, G. S. and Leadbetter, M. R. (1964a). Hazard analysis I. *Biometrika*, **51**, 175–184.

Watson, G. S. and Leadbetter, M. R. (1964b). Hazard analysis II. *Sankhyā Ser. A*, **26**, 101–116.

Watson, G. S. and Wells, W. R. (1961). On the possibility of improving the mean useful life of items by eliminating those with short lives. *Technometrics*, **3**, 281–298.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *J. Amer. Stat. Assoc.*, **84**, 1065–1073.

Wei, L. J., Ying, Z. and Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika*, **77**, 845–851.

Weibull, W. (1951). A statistical distribution function of wide applicability. *J. Appl. Mech.*, **18**, 293–297.

Weisberg, S. (1985). *Applied Linear Regression*, 2nd ed. John Wiley & Sons, New York.

Weissfeld, L. A. and Schneider, H. (1990). Influence diagnostics for the Weibull model fit to censored data. *Stat. Prob. Lett.*, **9**, 67–73.

Wellek, S. (1990). A nonparametric model for product-limit estimation under right censoring and left truncation. *Commun. Stat.-Stoch. Mod.*, **6**, 561–592.

Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the NPMLE from censored data. *J. Amer. Stat. Assoc.*, **92**, 945–959.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25.

White, J. S. (1969). The moments of log-Weibull order statistics. *Technometrics*, **11**, 373–386.

Whitehead, J. (1992). *The Design and Analysis of Sequential Clinical Trials*, 2nd ed. John Wiley & Sons, Chichester.

Whitmore, G. A. (1975). The inverse Gaussian distribution as a model of hospital stay. *Health Serv. Res.*, **10**, 297–302.

Whitmore, G. A. (1983). A regression model for censored inverse Gaussian data. *Can. J. Stat.*, **11**, 305–315.

Whitmore, G. A. (1986). Normal-gamma mixtures of inverse Gaussian distributions. *Scand. J. Stat.*, **13**, 211–220.

Whitmore, G. A. (1995). Estimating degradation by a Wiener diffusion process subject to measurement error. *Lifetime Data Anal.*, **1**, 307–319.

Whitmore, G. A. and Schenkelberg, F. (1997). Modelling accelerated degradation data using Wiener diffusion with a time scale transformation. *Lifetime Data Anal.*, **3**, 1–19.

Whitmore, G. A., Crowder, M. J., and Lawless, J. F (1998). Failure inference from a marker process based on a bivariate Wiener model. *Lifetime Data Anal.*, **4**, 229–251.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, **1**, 80–83.

Wild, C. J. (1983). Failure time models with matched data. *Biometrika*, **70**, 633–641.

Wilk, M. B., Gnanadesikan, R., and Huyett, M. J. (1962). Probability plots for the gamma distribution. *Technometrics*, **4**, 1–20.

Wilks, S. S. (1962). *Mathematical Statistics*. John Wiley & Sons, New York.

Williams, J. S. and Lagakos, S. W. (1977). Models for censored survival analysis: Constant sum and variable sum models. *Biometrika*, **64**, 215–224.

Williams, R. L. (1995). Product-limit survival functions with correlated survival times. *Lifetime Data Anal.*, **1**, 171–186.

Wilson, S. P. (2000). Failure models indexed by time and usage. In *Recent Advances in Reliability Theory*, N. Limnios and M. Nikulin, Eds., pp. 229–243. Birkhauser, Boston.

Winter, B. B., Földes, A., and Rejtő, L. (1978). Glivenko–Cantelli theorems for the product limit estimate. *Probl. Control Inf. Theory*, **7**, 213–225.

Wong, A. C. M. and Wu, J. (2000). Practical small-sample asymptotics for distributions used in life-data analysis. *Technometrics*, **42**, 149–155.

Wong, W. H. (1986). Theory of partial likelihood. *Ann. Stat.*, **14**, 88–123.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Stat.*, **13**, 163–177. Correction: **15**, 883 (1987).

Wu, C. F. J. and Hamada, M. S. (1999). *Experiments: Planning, Analysis and Parameter Design Optimization*. John Wiley & Sons, New York.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, **53**, 330–339.

Xu, J. and Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Appl. Stat.*, **50**, 375–387.

Xue, X. and Brookmeyer, R. (1996). Bivariate frailty model for the analysis of multivariate failure time. *Lifetime Data Anal.*, **2**, 277–289.

Yandell, B. S. (1983). Nonparametric inference for rates with censored survival data. *Ann. Stat.*, **11**, 1119–1135.

Yang, Z. (1977). Life expectancy under random censorship. *Stochastic Processes and Their Applications*, **6**, 33–39.

Ying, Z. (1993). A large sample study of rank regression for censored regression data. *Ann. Stat.*, **21**, 76–99.

Younes, N. and Lachin, J. (1997). Link-based models for survival data with interval and continuous time censoring. *Biometrics*, **53**, 1199–1211.

Zelen, M. and Dannemiller, M. (1961). The robustness of life testing procedures derived from the exponential distribution. *Technometrics*, **3**, 29–49.

Zhao, H. and Tsiatis, A. A. (1997). A consistent estimator for the distribution of quality adjusted survival time. *Biometrika*, **84**, 339–348.

Zippin, C. and Armitage, P. (1966). Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. *Biometrics*, **22**, 665–672.

Author Index

Subject Index

BERRY, CHALONER, and GEWEKE · Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner
BERNARDO and SMITH · Bayesian Theory
BHAT and MILLER · Elements of Applied Stochastic Processes, *Third Edition*
BHATTACHARYA and JOHNSON · Statistical Concepts and Methods
BHATTACHARYA and WAYMIRE · Stochastic Processes with Applications
BILLINGSLEY · Convergence of Probability Measures, *Second Edition*
BILLINGSLEY · Probability and Measure, *Third Edition*
BIRKES and DODGE · Alternative Methods of Regression
BLISCHKE AND MURTHY · Reliability: Modeling, Prediction, and Optimization
BLOOMFIELD · Fourier Analysis of Time Series: An Introduction, *Second Edition*
BOLLEN · Structural Equations with Latent Variables
BOROVKOV · Ergodicity and Stability of Stochastic Processes
BOULEAU · Numerical Methods for Stochastic Processes
BOX · Bayesian Inference in Statistical Analysis
BOX · R. A. Fisher, the Life of a Scientist
BOX and DRAPER · Empirical Model-Building and Response Surfaces
*BOX and DRAPER · Evolutionary Operation: A Statistical Method for Process Improvement
BOX, HUNTER, and HUNTER · Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building
BOX and LUCEÑO · Statistical Control by Monitoring and Feedback Adjustment
BRANDIMARTE · Numerical Methods in Finance: A MATLAB-Based Introduction
BROWN and HOLLANDER · Statistics: A Biomedical Introduction
BRUNNER, DOMHOF, and LANGER · Nonparametric Analysis of Longitudinal Data in Factorial Experiments
BUCKLEW · Large Deviation Techniques in Decision, Simulation, and Estimation
CAIROLI and DALANG · Sequential Stochastic Optimization
CHAN · Time Series: Applications to Finance
CHATTERJEE and HADI · Sensitivity Analysis in Linear Regression
CHATTERJEE and PRICE · Regression Analysis by Example, *Third Edition*
CHERNICK · Bootstrap Methods: A Practitioner's Guide
CHILÈS and DELFINER · Geostatistics: Modeling Spatial Uncertainty
CHOW and LIU · Design and Analysis of Clinical Trials: Concepts and Methodologies
CLARKE and DISNEY · Probability and Random Processes: A First Course with Applications, *Second Edition*
*COCHRAN and COX · Experimental Designs, *Second Edition*
CONGDON · Bayesian Statistical Modelling
CONOVER · Practical Nonparametric Statistics, *Second Edition*
COOK · Regression Graphics
COOK and WEISBERG · Applied Regression Including Computing and Graphics
COOK and WEISBERG · An Introduction to Regression Graphics
CORNELL · Experiments with Mixtures, Designs, Models, and the Analysis of Mixture Data, *Third Edition*
COVER and THOMAS · Elements of Information Theory
COX · A Handbook of Introductory Statistical Methods
*COX · Planning of Experiments
CRESSIE · Statistics for Spatial Data, *Revised Edition*
CSÖRGÖ and HORVÁTH · Limit Theorems in Change Point Analysis
DANIEL · Applications of Statistics to Industrial Experimentation
DANIEL · Biostatistics: A Foundation for Analysis in the Health Sciences, *Sixth Edition*
*DANIEL · Fitting Equations to Data: Computer Analysis of Multifactor Data, *Second Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.

DAVID · Order Statistics, *Second Edition*
*DEGROOT, FIENBERG, and KADANE · Statistics and the Law
DEL CASTILLO · Statistical Process Adjustment for Quality Control
DETTE and STUDDEN · The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis
DEY and MUKERJEE · Fractional Factorial Plans
DILLON and GOLDSTEIN · Multivariate Analysis: Methods and Applications
DODGE · Alternative Methods of Regression
*DODGE and ROMIG · Sampling Inspection Tables, *Second Edition*
*DOOB · Stochastic Processes
DOWDY and WEARDEN · Statistics for Research, *Second Edition*
DRAPER and SMITH · Applied Regression Analysis, *Third Edition*
DRYDEN and MARDIA · Statistical Shape Analysis
DUDEWICZ and MISHRA · Modern Mathematical Statistics
DUNN and CLARK · Applied Statistics: Analysis of Variance and Regression, *Second Edition*
DUNN and CLARK · Basic Statistics: A Primer for the Biomedical Sciences, *Third Edition*
DUPUIS and ELLIS · A Weak Convergence Approach to the Theory of Large Deviations
*ELANDT-JOHNSON and JOHNSON · Survival Models and Data Analysis
ETHIER and KURTZ · Markov Processes: Characterization and Convergence
EVANS, HASTINGS, and PEACOCK · Statistical Distributions, *Third Edition*
FELLER · An Introduction to Probability Theory and Its Applications, Volume I, *Third Edition*, Revised; Volume II, *Second Edition*
FISHER and VAN BELLE · Biostatistics: A Methodology for the Health Sciences
*FLEISS · The Design and Analysis of Clinical Experiments
FLEISS · Statistical Methods for Rates and Proportions, *Second Edition*
FLEMING and HARRINGTON · Counting Processes and Survival Analysis
FULLER · Introduction to Statistical Time Series, *Second Edition*
FULLER · Measurement Error Models
GALLANT · Nonlinear Statistical Models
GHOSH, MUKHOPADHYAY, and SEN · Sequential Estimation
GIFI · Nonlinear Multivariate Analysis
GLASSERMAN and YAO · Monotone Structure in Discrete-Event Systems
GNANADESIKAN · Methods for Statistical Data Analysis of Multivariate Observations, *Second Edition*
GOLDSTEIN and LEWIS · Assessment: Problems, Development, and Statistical Issues
GREENWOOD and NIKULIN · A Guide to Chi-Squared Testing
GROSS and HARRIS · Fundamentals of Queueing Theory, *Third Edition*
*HAHN and SHAPIRO · Statistical Models in Engineering
HAHN and MEEKER · Statistical Intervals: A Guide for Practitioners
HALD · A History of Probability and Statistics and their Applications Before 1750
HALD · A History of Mathematical Statistics from 1750 to 1930
HAMPEL · Robust Statistics: The Approach Based on Influence Functions
HANNAN and DEISTLER · The Statistical Theory of Linear Systems
HEIBERGER · Computation for the Analysis of Designed Experiments
HEDAYAT and SINHA · Design and Inference in Finite Population Sampling
HELLER · MACSYMA for Statisticians
HINKELMAN and KEMPTHORNE: · Design and Analysis of Experiments, Volume 1: Introduction to Experimental Design
HOAGLIN, MOSTELLER, and TUKEY · Exploratory Approach to Analysis of Variance
HOAGLIN, MOSTELLER, and TUKEY · Exploring Data Tables, Trends and Shapes

*Now available in a lower priced paperback edition in the Wiley Classics Library.

*HOAGLIN, MOSTELLER, and TUKEY · Understanding Robust and Exploratory Data Analysis
HOCHBERG and TAMHANE · Multiple Comparison Procedures
HOCKING · Methods and Applications of Linear Models: Regression and the Analysis of Variables
HOEL · Introduction to Mathematical Statistics, *Fifth Edition*
HOGG and KLUGMAN · Loss Distributions
HOLLANDER and WOLFE · Nonparametric Statistical Methods, *Second Edition*
HOSMER and LEMESHOW · Applied Logistic Regression, *Second Edition*
HOSMER and LEMESHOW · Applied Survival Analysis: Regression Modeling of Time to Event Data
HØYLAND and RAUSAND · System Reliability Theory: Models and Statistical Methods
HUBER · Robust Statistics
HUBERTY · Applied Discriminant Analysis
HUNT and KENNEDY · Financial Derivatives in Theory and Practice
HUSKOVA, BERAN, and DUPAC · Collected Works of Jaroslav Hajek—with Commentary
IMAN and CONOVER · A Modern Approach to Statistics
JACKSON · A User's Guide to Principle Components
JOHN · Statistical Methods in Engineering and Quality Assurance
JOHNSON · Multivariate Statistical Simulation
JOHNSON and BALAKRISHNAN · Advances in the Theory and Practice of Statistics: A Volume in Honor of Samuel Kotz
JUDGE, GRIFFITHS, HILL, LÜTKEPOHL, and LEE · The Theory and Practice of Econometrics, *Second Edition*
JOHNSON and KOTZ · Distributions in Statistics
JOHNSON and KOTZ (editors) · Leading Personalities in Statistical Sciences: From the Seventeenth Century to the Present
JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions, Volume 1, *Second Edition*
JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions, Volume 2, *Second Edition*
JOHNSON, KOTZ, and BALAKRISHNAN · Discrete Multivariate Distributions
JOHNSON, KOTZ, and KEMP · Univariate Discrete Distributions, *Second Edition*
JUREČKOVÁ and SEN · Robust Statistical Procedures: Aymptotics and Interrelations
JUREK and MASON · Operator-Limit Distributions in Probability Theory
KADANE · Bayesian Methods and Ethics in a Clinical Trial Design
KADANE AND SCHUM · A Probabilistic Analysis of the Sacco and Vanzetti Evidence
KALBFLEISCH and PRENTICE · The Statistical Analysis of Failure Time Data, *Second Edition*
KASS and VOS · Geometrical Foundations of Asymptotic Inference
KAUFMAN and ROUSSEEUW · Finding Groups in Data: An Introduction to Cluster Analysis
KEDEM and FOKIANOS · Regression Models for Time Series Analysis
KENDALL, BARDEN, CARNE, and LE · Shape and Shape Theory
KHURI · Advanced Calculus with Applications in Statistics, *Second Edition*
KHURI, MATHEW, and SINHA · Statistical Tests for Mixed Linear Models
KLUGMAN, PANJER, and WILLMOT · Loss Models: From Data to Decisions
KLUGMAN, PANJER, and WILLMOT · Solutions Manual to Accompany Loss Models: From Data to Decisions
KOTZ, BALAKRISHNAN, and JOHNSON · Continuous Multivariate Distributions, Volume 1, *Second Edition*
KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Volumes 1 to 9 with Index

*Now available in a lower priced paperback edition in the Wiley Classics Library.

KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Supplement Volume
KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 1
KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 2
KOVALENKO, KUZNETZOV, and PEGG · Mathematical Theory of Reliability of Time-Dependent Systems with Practical Applications
LACHIN · Biostatistical Methods: The Assessment of Relative Risks
LAD · Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction
LAMPERTI · Probability: A Survey of the Mathematical Theory, *Second Edition*
LANGE, RYAN, BILLARD, BRILLINGER, CONQUEST, and GREENHOUSE · Case Studies in Biometry
LARSON · Introduction to Probability Theory and Statistical Inference, *Third Edition*
LAWLESS · Statistical Models and Methods for Lifetime Data, *Second Edition*
LAWSON · Statistical Methods in Spatial Epidemiology
LE · Applied Categorical Data Analysis
LE · Applied Survival Analysis
LEE and WANG · Statistical Methods for Survival Data Analysis, *Third Edition*
LEPAGE and BILLARD · Exploring the Limits of Bootstrap
LEYLAND and GOLDSTEIN (editors) · Multilevel Modelling of Health Statistics
LIAO · Statistical Group Comparison
LINDVALL · Lectures on the Coupling Method
LINHART and ZUCCHINI · Model Selection
LITTLE and RUBIN · Statistical Analysis with Missing Data, *Second Edition*
LLOYD · The Statistical Analysis of Categorical Data
MAGNUS and NEUDECKER · Matrix Differential Calculus with Applications in Statistics and Econometrics, *Revised Edition*
MALLER and ZHOU · Survival Analysis with Long Term Survivors
MALLOWS · Design, Data, and Analysis by Some Friends of Cuthbert Daniel
MANN, SCHAFER, and SINGPURWALLA · Methods for Statistical Analysis of Reliability and Life Data
MANTON, WOODBURY, and TOLLEY · Statistical Applications Using Fuzzy Sets
MARDIA and JUPP · Directional Statistics
MASON, GUNST, and HESS · Statistical Design and Analysis of Experiments with Applications to Engineering and Science, *Second Edition*
McCULLOCH and SEARLE · Generalized, Linear, and Mixed Models
McFADDEN · Management of Data in Clinical Trials
McLACHLAN · Discriminant Analysis and Statistical Pattern Recognition
McLACHLAN and KRISHNAN · The EM Algorithm and Extensions
McLACHLAN and PEEL · Finite Mixture Models
McNEIL · Epidemiological Research Methods
MEEKER and ESCOBAR · Statistical Methods for Reliability Data
MEERSCHAERT and SCHEFFLER · Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice
*MILLER · Survival Analysis, *Second Edition*
MONTGOMERY, PECK, and VINING · Introduction to Linear Regression Analysis, *Third Edition*
MORGENTHALER and TUKEY · Configural Polysampling: A Route to Practical Robustness
MUIRHEAD · Aspects of Multivariate Statistical Theory
MURRAY · X-STAT 2.0 Statistical Experimentation, Design Data Analysis, and Nonlinear Optimization

*Now available in a lower priced paperback edition in the Wiley Classics Library.

MYERS and MONTGOMERY · Response Surface Methodology: Process and Product Optimization Using Designed Experiments, *Second Edition*
MYERS, MONTGOMERY, and VINING · Generalized Linear Models. With Applications in Engineering and the Sciences
NELSON · Accelerated Testing, Statistical Models, Test Plans, and Data Analyses
NELSON · Applied Life Data Analysis
NEWMAN · Biostatistical Methods in Epidemiology
OCHI · Applied Probability and Stochastic Processes in Engineering and Physical Sciences
OKABE, BOOTS, SUGIHARA, and CHIU · Spatial Tesselations: Concepts and Applications of Voronoi Diagrams, *Second Edition*
OLIVER and SMITH · Influence Diagrams, Belief Nets and Decision Analysis
PANKRATZ · Forecasting with Dynamic Regression Models
PANKRATZ · Forecasting with Univariate Box-Jenkins Models: Concepts and Cases
*PARZEN · Modern Probability Theory and Its Applications
PEÑA, TIAO, and TSAY · A Course in Time Series Analysis
PIANTADOSI · Clinical Trials: A Methodologic Perspective
PORT · Theoretical Probability for Applications
POURAHMADI · Foundations of Time Series Analysis and Prediction Theory
PRESS · Bayesian Statistics: Principles, Models, and Applications
PRESS · Subjective and Objective Bayesian Statistics, *Second Edition*
PRESS and TANUR · The Subjectivity of Scientists and the Bayesian Approach
PUKELSHEIM · Optimal Experimental Design
PURI, VILAPLANA, and WERTZ · New Perspectives in Theoretical and Applied Statistics
PUTERMAN · Markov Decision Processes: Discrete Stochastic Dynamic Programming
*RAO · Linear Statistical Inference and Its Applications, *Second Edition*
RENCHER · Linear Models in Statistics
RENCHER · Methods of Multivariate Analysis, *Second Edition*
RENCHER · Multivariate Statistical Inference with Applications
RIPLEY · Spatial Statistics
RIPLEY · Stochastic Simulation
ROBINSON · Practical Strategies for Experimenting
ROHATGI and SALEH · An Introduction to Probability and Statistics, *Second Edition*
ROLSKI, SCHMIDLI, SCHMIDT, and TEUGELS · Stochastic Processes for Insurance and Finance
ROSENBERGER and LACHIN · Randomization in Clinical Trials: Theory and Practice
ROSS · Introduction to Probability and Statistics for Engineers and Scientists
ROUSSEEUW and LEROY · Robust Regression and Outlier Detection
RUBIN · Multiple Imputation for Nonresponse in Surveys
RUBINSTEIN · Simulation and the Monte Carlo Method
RUBINSTEIN and MELAMED · Modern Simulation and Modeling
RYAN · Modern Regression Methods
RYAN · Statistical Methods for Quality Improvement, *Second Edition*
SALTELLI, CHAN, and SCOTT (editors) · Sensitivity Analysis
*SCHEFFE · The Analysis of Variance
SCHIMEK · Smoothing and Regression: Approaches, Computation, and Application
SCHOTT · Matrix Analysis for Statistics
SCHUSS · Theory and Applications of Stochastic Differential Equations
SCOTT · Multivariate Density Estimation: Theory, Practice, and Visualization
*SEARLE · Linear Models
SEARLE · Linear Models for Unbalanced Data
SEARLE · Matrix Algebra Useful for Statistics
SEARLE, CASELLA, and McCULLOCH · Variance Components
SEARLE and WILLETT · Matrix Algebra for Applied Economics

*Now available in a lower priced paperback edition in the Wiley Classics Library.

SEBER · Linear Regression Analysis
SEBER · Multivariate Observations
SEBER and WILD · Nonlinear Regression
SENNOTT · Stochastic Dynamic Programming and the Control of Queueing Systems
*SERFLING · Approximation Theorems of Mathematical Statistics
SHAFER and VOVK · Probability and Finance: It's Only a Game!
SMALL and McLEISH · Hilbert Space Methods in Probability and Statistical Inference
SRIVASTAVA · Methods of Multivariate Statistics
STAPLETON · Linear Statistical Models
STAUDTE and SHEATHER · Robust Estimation and Testing
STOYAN, KENDALL, and MECKE · Stochastic Geometry and Its Applications, *Second Edition*
STOYAN and STOYAN · Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics
STYAN · The Collected Papers of T. W. Anderson: 1943–1985
SUTTON, ABRAMS, JONES, SHELDON, and SONG · Methods for Meta-Analysis in Medical Research
TANAKA · Time Series Analysis: Nonstationary and Noninvertible Distribution Theory
THOMPSON · Empirical Model Building
THOMPSON · Sampling, *Second Edition*
THOMPSON · Simulation: A Modeler's Approach
THOMPSON and SEBER · Adaptive Sampling
THOMPSON, WILLIAMS, and FINDLAY · Models for Investors in Real World Markets
TIAO, BISGAARD, HILL, PEÑA, and STIGLER (editors) · Box on Quality and Discovery: with Design, Control, and Robustness
TIERNEY · LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics
TSAY · Analysis of Financial Time Series
UPTON and FINGLETON · Spatial Data Analysis by Example, Volume II: Categorical and Directional Data
VAN BELLE · Statistical Rules of Thumb
VIDAKOVIC · Statistical Modeling by Wavelets
WEISBERG · Applied Linear Regression, *Second Edition*
WELSH · Aspects of Statistical Inference
WESTFALL and YOUNG · Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment
WHITTAKER · Graphical Models in Applied Multivariate Statistics
WINKER · Optimization Heuristics in Economics: Applications of Threshold Accepting
WONNACOTT and WONNACOTT · Econometrics, *Second Edition*
WOODING · Planning Pharmaceutical Clinical Trials: Basic Statistical Principles
WOOLSON and CLARKE · Statistical Methods for the Analysis of Biomedical Data, *Second Edition*
WU and HAMADA · Experiments: Planning, Analysis, and Parameter Design Optimization
YANG · The Construction Theory of Denumerable Markov Processes
*ZELLNER · An Introduction to Bayesian Inference in Econometrics
ZHOU, OBUCHOWSKI, and McCLISH · Statistical Methods in Diagnostic Medicine

*Now available in a lower priced paperback edition in the Wiley Classics Library.