

Censoring and Likelihood¹

STA312 Fall 2023

¹See last slide for copyright information.

Censored Data

- Let T^* represent time to failure.
- Let U represent censoring time.
- We observe $T = \min(T^*, U)$.
- And an indicator for whether failure occurred: $\delta = 1$ if uncensored, and zero if censored.
- All this is $i = 1, \dots, n$, so we observe n pairs $(T_1, \delta_1), (T_2, \delta_2), \dots, (T_n, \delta_n)$.
- A simple data file might look like this:

Patient	Time	Uncensored (delta)
1	5	1
2	6	0
3	8	1
4	3	1
5	22	1

Simulation code

To illustrate the process

```
# Exponential distribution, true parameter lambda, right censoring
# Censoring times will be uniform(0,top)
# Observation will be censored if censoring time is less than lifetime.
rm(list=ls()); options(scipen=999)
lambda = 1/5; top=20 # True parameters
# Simulate
set.seed(9999); n = 200
delta = numeric(n) # Indicator for uncensored, initially zero
lifetime = rexp(n,rate=lambda)
censortime = runif(n,0,top)
# If censoring time is greater than lifetime, then it's NOT censored.
delta[censortime>lifetime] = 1
# Minimum of censortime and lifetime is what we can observe.
T = pmin(censortime,lifetime) # pmin is parallel minimum.
round(cbind(lifetime,censortime,T,delta)[1:10,],2) # Take a look
expodata = cbind(T,delta) # This is all you can see in practice.
```

Note that censoring time is independent of lifetime, and they share no parameters.

Output

```
round(cbind(lifetime,censortime,T,delta)[1:10,],2) # Take a look
```

	lifetime	censortime	T	delta
[1,]	2.29	1.38	1.38	0
[2,]	10.24	4.31	4.31	0
[3,]	1.89	13.23	1.89	1
[4,]	3.45	14.61	3.45	1
[5,]	9.77	17.82	9.77	1
[6,]	2.41	16.46	2.41	1
[7,]	3.18	13.83	3.18	1
[8,]	20.50	7.41	7.41	0
[9,]	8.64	8.53	8.53	0
[10,]	2.65	19.61	2.65	1

Likelihood function for censored data

Possibly right censored, with random censoring

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

$$\begin{aligned} \ell(\theta) &= \sum_{i=1}^n (\delta_i \log f(t_i|\theta) + (1 - \delta_i) \log S(t_i|\theta)) \\ &= \sum_{i=1}^n \delta_i \log f(t_i|\theta) + \sum_{i=1}^n (1 - \delta_i) \log S(t_i|\theta) \end{aligned}$$

Example: $T_1, \dots, T_n \stackrel{i.i.d}{\sim} \exp(\lambda)$

$F(t|\lambda) = 1 - e^{-\lambda t}$, Ordinary $\hat{\lambda}_n = 1/\bar{T}_n$

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

As an exercise, you will show that $\hat{\lambda}_n = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n t_i}$.

- So the estimated expected value is $\frac{\sum_{i=1}^n t_i}{\sum_{i=1}^n \delta_i} > \frac{\sum_{i=1}^n t_i}{n} = \bar{T}_n$.
- Ignoring the censoring would cause you to under-estimate average survival time.

Copyright Information

This slide show was prepared by **Jerry Brunner**, Department of Statistics, University of Toronto. It is licensed under a **Creative Commons Attribution - ShareAlike 3.0 Unported License**. Use any part of it as you like and share the result freely. The \LaTeX source code is available from the course website:

<http://www.utstat.toronto.edu/brunner/oldclass/312f23>