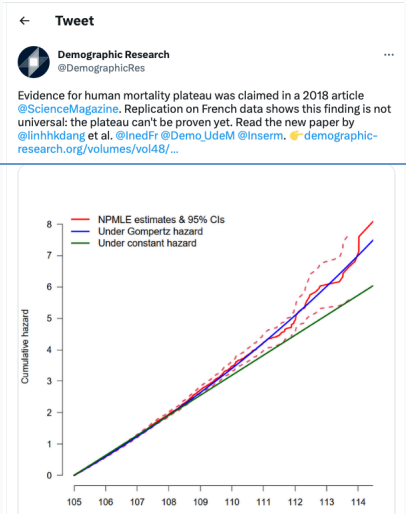
Mathematical Statistics II

STA2212H S LEC0101

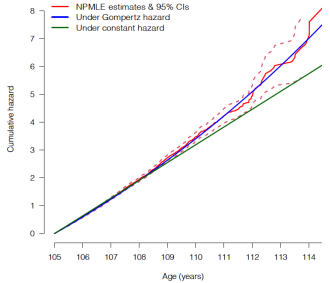
Week 8

March 7 2023



Mortality plateau

Figure 1: **Estimated cumulative hazard using nonparametric and parametric approaches, French females born 1883–1901, ages 105 and above**



Notes: NPMLE is for nonparametric maximum likelihood estimate. This figure is equivalent to Figure 2 in Barbi et al. (2018: p. 1461).
Source: Authors' calculations based on data from the Répertoire national d'identification des personnes physiques (RNIPP).

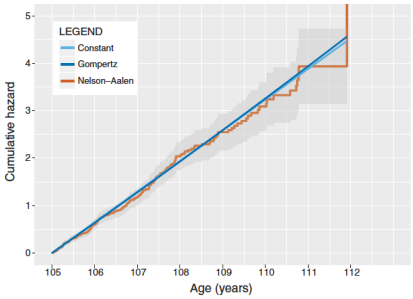


Fig. 2. Cumulative hazard beyond age 105 for the cohort of Italian women born in 1904, as determined by the Nelson-Aalen estimator. Straight lines represent cumulative hazards of the estimated plateau predicted from ISTAT data, under a constant hazard (light blue) and a Gompertz hazard model (darker blue). The shaded area indicates the 95% confidence bands of the Nelson-Aalen estimate.

France 2022 (Dang et al.)

Italy 2017 (Barbi et al.)

$$h(t;x) = \frac{f(t;x)}{1-F(t;x)} = a\exp(bt + \beta_1 x_1 + \beta_2 x_2); \quad H_0 : b = 0$$

Today

1. Next lectures
2. Recap
3. Multiple testing
4. Goodness-of-fit tests

Upcoming

- March 9 3.30 – 4.30 DoSS 9014 Details
  “Valid statistical inference with privacy constraints”
  Aleksandra Slavković, Penn State

- March 13 3.30 –4.30 DoSS 9014 & online Details
  “Training Scientists to Perform Robust Bayesian Inference”
  Justin Bois, CalTech

Next Lectures

- March 14 10.00 – 13.00
- March 21 11.00 – 13.00
- March 28 10.00 – 12.00
- April 4 Project presentations

Next Lectures

- March 14 10.00 – 13.00
- March 21 11.00 – 13.00
- March 28 10.00 – 12.00
- April 4 Project presentations

STA 2212S: Mathematical Statistics II
Syllabus Spring 2023

Week	Date	Methods	References
1	Jan 10	Likelihood inference: review of ML estimation; mis-specified models; computation; nonparametric mle	MS §§5.1–7, SM Ch 4
2	Jan 17	Bayesian estimation; Bayesian inference	MS §5.8; AoS §§ 11.1–4; SM §§11.1,2
3	Jan 24	Optimality in estimation	MS Ch 6; AoS Ch 12; SM §7.1, 11.5.2
4	Jan 31	Interval estimation; Confidence bands	MS §§7.1,2; AoS Ch 7; SM §7.1.4
5	Feb 7	Hypothesis testing; likelihood ratio tests	MS §§7.1–4 AoS Ch 10.6, SM
6	Feb 14	Significance testing	MS §7.5; AoS §10.2,6; SM Ch 4, §7.3.1
	Feb 21	Break	
7	Feb 28	Significance testing	SM 7.3.1
7	Feb 28	Goodness-of-fit testing	MS Ch 9; AoS §§10.3.4,5,8; SM p.327-8 (hard)
8	Mar 7	Multiple testing and FDR	AoS Ch 10.7, EH Ch 15.1,2
9	Mar 14	Empirical Bayes	EH Ch 6, SM Ch 11.5
10	Mar 21	Multivariate Models	AoS Ch 14; SM Ch 6.3
11	Mar 28	Introduction to Causal Inference	AoS Ch 16, 17
12	Apr 4	Recap	

Subject to adjustment as the course progresses.

Next Lectures

- March 14 10.00 – 13.00
- March 21 11.00 – 13.00
- March 28 10.00 – 12.00
- April 4 Project presentations

2	Jan 17	Bayesian estimation; Bayesian inference	MS §9.8; AoS §§ 11.1–4; SM §§11.1,2
3	Jan 24	Optimality in estimation	MS Ch 6; AoS Ch 12; SM §7.1, 11.5.2
4	Jan 31	Interval estimation; Confidence bands	MS §§7.1,2; AoS Ch 7; SM §7.1.4
5	Feb 7	Hypothesis testing; likelihood ratio tests	MS §§7.1–4 AoS Ch 10.6, SM
6	Feb 14	Significance testing	MS §7.5; AoS §10.2,6; SM Ch 4, §7.3.1
	Feb 21	Break	
7	Feb 28	Significance testing	SM 7.3.1
7	Feb 28	Goodness-of-fit testing	MS Ch 9; AoS §§10.3,4,5,8; SM p.327-8 (hard)
8	Mar 7	Multiple testing and FDR	AoS Ch 10.7, EH Ch 15.1,2
9	Mar 14		
10	Mar 21	Likelihood asymptotics; robust estimation (SM 7.2); causal inference (AoS 16); linear and generalized linear models (MS 8); graphical models (AoS 17,18); nonparametric curve estimation (AoS 20); classification (AoS 22); any of 1-8	
11	Mar 28		
12	Apr 4	Course Summary; Presentations	

Subject to adjustment as the course progresses.
References

MS: *Mathematical Statistics* by K. Knight (Chapman & Hall/CRC).

AoS: *All of Statistics* by L. Wasserman (Springer) If your copy has a **Chapter 1. Introduction**, then all Chapter numbers increase by 1.

SM: *Statistical Models* by A.C. Davison (Cambridge University Press)

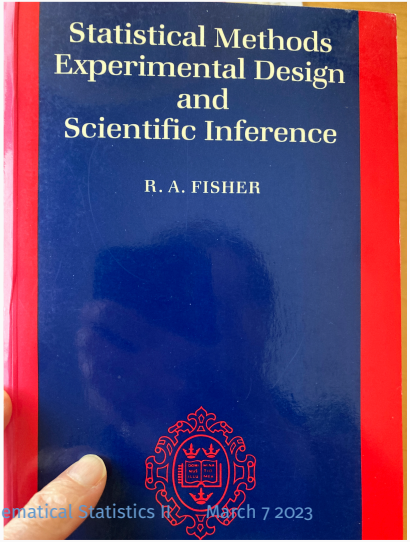
1

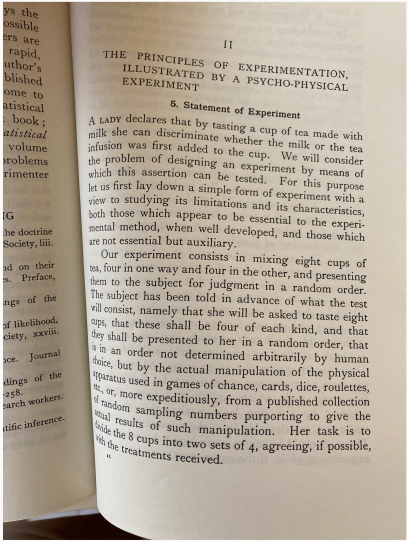
Recap

- Neyman-Pearson lemma; simple and composite hypotheses; power and size

- p-values: definition, interpretation; diagnostic tests
- sign test; permutation test; intro to multiple testing

Fisher 1935





II

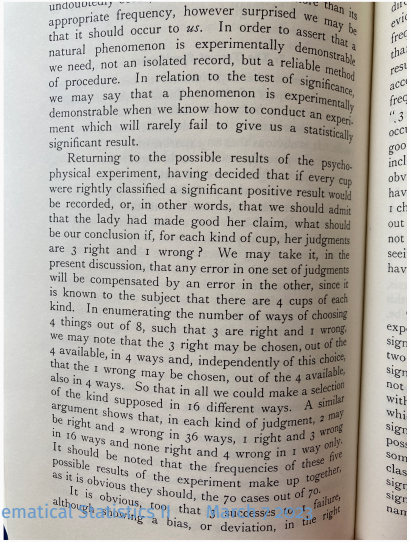
THE PRINCIPLES OF EXPERIMENTATION, ILLUSTRATED BY A PSYCHO-PHYSICAL EXPERIMENT

5. Statement of Experiment

A LADY declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup. We will consider the problem of designing an experiment by means of which this assertion can be tested. For this purpose let us first lay down a simple form of experiment with a view to studying its limitations and its characteristics, both those which appear to be essential to the experimental method, when well developed, and those which are not essential but auxiliary.

Our experiment consists in mixing eight cups of tea, four in one way and four in the other, and presenting them to the subject for judgment in a random order. The subject has been told in advance of what the test will consist, namely that she will be asked to taste eight cups, that these shall be four of each kind, and that they shall be presented to her in a random order, that is in an order not determined arbitrarily by human choice, but by the actual manipulation of the physical apparatus used in games of chance, cards, dice, roulettes, etc., or, more expeditiously, from a published collection of random sampling numbers purporting to give the actual results of such manipulation. Her task is to divide the 8 cups into two sets of 4, agreeing, if possible, with the treatments received.

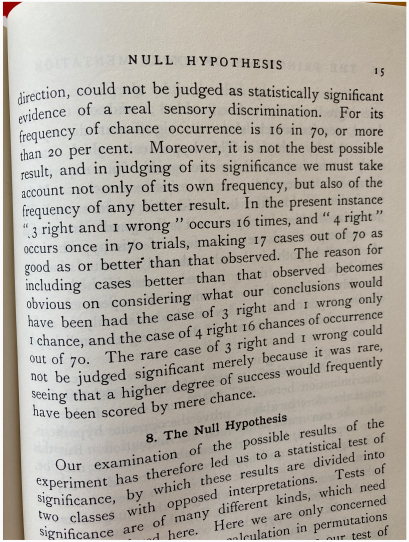
Fisher 1935



undoubtedly ... more than its appropriate frequency, however surprised we may be that it should occur to *us*. In order to assert that a natural phenomenon is experimentally demonstrable we need, not an isolated record, but a reliable method of procedure. In relation to the test of significance, we may say that a phenomenon is experimentally demonstrable when we know how to conduct an experiment which will rarely fail to give us a statistically significant result.

Returning to the possible results of the psycho-physical experiment, having decided that if every cup were rightly classified a significant positive result would be recorded, or, in other words, that we should admit that the lady had made good her claim, what should be our conclusion if, for each kind of cup, her judgments are 3 right and 1 wrong? We may take it, in the present discussion, that any error in one set of judgments will be compensated by an error in the other, since it is known to the subject that there are 4 cups of each kind. In enumerating the number of ways of choosing 4 things out of 8, such that 3 are right and 1 wrong, we may note that the 3 right may be chosen, out of the 4 available, in 4 ways, and, independently of this choice, that the 1 wrong may be chosen, out of the 4 available, also in 4 ways. So that in all we could make a selection of the kind supposed in 16 different ways. A similar argument shows that, in each kind of judgment, 2 may be right and 2 wrong in 36 ways, 1 right and 3 wrong in 16 ways and none right and 4 wrong in 1 way only. It should be noted that the frequencies of these five possible results of the experiment make up together, as it is obvious they should, the 70 cases out of 70.

It is obvious, too, that [illegible] successes to failures, although showing a bias, or deviation, in the right



direction, could not be judged as statistically significant evidence of a real sensory discrimination. For its frequency of chance occurrence is 16 in 70, or more than 20 per cent. Moreover, it is not the best possible result, and in judging of its significance we must take account not only of its own frequency, but also of the frequency of any better result. In the present instance "3 right and 1 wrong" occurs 16 times, and "4 right" occurs once in 70 trials, making 17 cases out of 70 as good as or better than that observed. The reason for including cases better than that observed becomes obvious on considering what our conclusions would have been had the case of 3 right and 1 wrong only 1 chance, and the case of 4 right 16 chances of occurrence out of 70. The rare case of 3 right and 1 wrong could not be judged significant merely because it was rare, seeing that a higher degree of success would frequently have been scored by mere chance.

8. The Null Hypothesis

Our examination of the possible results of the experiment has therefore led us to a statistical test of significance, by which these results are divided into two classes with opposed interpretations. Tests of significance are of many different kinds, which need [illegible] here. Here we are only concerned [illegible] calculation in permutations [illegible] our test of

Recap: Choosing test statistics

1. Context
2. Optimal choice – Neyman-Pearson Lemma — and its extensions; MS 7.3
3. Pragmatic choice – likelihood-based statistics — Wald, score, LRT
4. Pragmatic choice – nonparametric test statistics — sign, permutation

Diagnostic testing

Wikipedia

1. Hypothesis testing

		H_0 not rejected	H_0 rejected
truth	H_0 true		type 1 error
	H_1 true	type 2 error	

2. Diagnostic testing

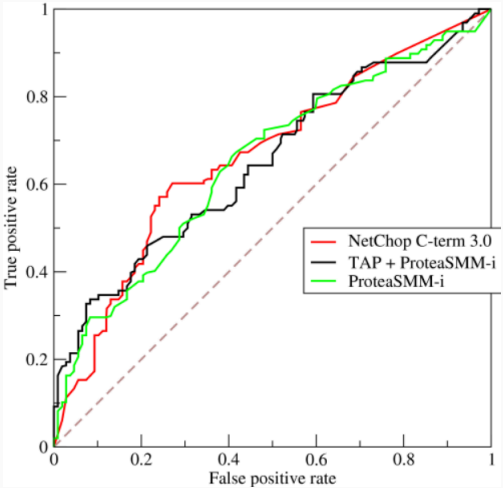
link

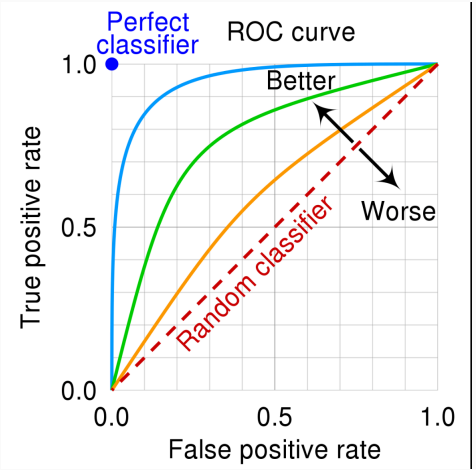
		test negative	test positive	
truth	C19 neg	TN	FP	N
	C19 pos	FN	TP	P

specificity = TN/N

sensitivity = TP/P

Diagnostic testing and ROC





2. Diagnostic testing

link

truth	test negative	test positive	
C19 neg	TN	FP	N
C19 pos	FN	TP	P

3. Multiple testing

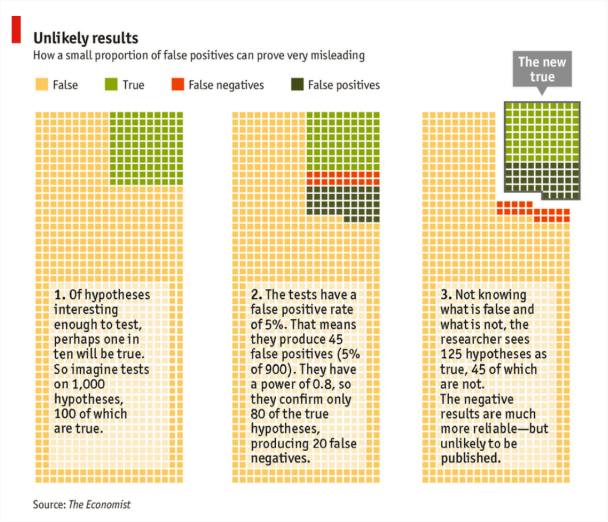
truth	H_0 not rejected	H_0 rejected	
H_0 true	U	V	m_0
H_1 true	T	S	m_1
	$m - R$	R	m

FDP = V/R

$\equiv 0$ if $R = 0$

FDR = $\mathrm{E}(V/R)$

Multiple testing



	H_0 not rejected	H_0 rejected	
H_0 true	U	V	m_0
H_1 true	T	S	m_1
	$m - R$	R	m

	test negative	test positive	
C19 neg	TN	FP	N
C19 pos	FN	TP	P

- H_{0i} versus H_{1i}, $\quad i = 1, \ldots, m$
- p-values $p_1, \ldots, p_m$
- Bonferroni method: reject H_{0i} if $p_i < \alpha/m$
- $\mathrm{pr}(\text{any } H_0 \text{ falsely rejected }) \leq \alpha$ very conservative

- H_{0i} versus $H_{1i}, \quad i = 1, \ldots, m$
- p-values $p_1, \ldots, p_m$
- Bonferroni method: reject H_{0i} if $p_i < \alpha/m$
- $\mathrm{pr}(\text{any } H_0 \text{ falsely rejected }) \leq \alpha$ very conservative

- FDR method controls the number of rejections that are false FDP = V/R

		H_0 not rejected	H_0 rejected	
truth	H_0 true	U	V	m_0
	H_1 true	T	S	m_1
		$m - R$	R	m

- order the p-values $p_{(1)}, \ldots, p_{(m)}$
- find i_{max}, the largest index for which

$$p_{(i)} \leq \frac{i}{m} q$$

- Let BH_q be the rule that rejects H_{oi} for $i \leq i_{max}$, not rejecting otherwise

- order the p-values $p_{(1)}, \ldots, p_{(m)}$
- find i_{max}, the largest index for which

$$p_{(i)} \leq \frac{i}{m} q$$

- Let BH_q be the rule that rejects H_{oi} for $i \leq i_{max}$, not rejecting otherwise
- change the bound under dependence

$$p_{(i)} \leq \frac{i}{m C_m} q \qquad C_m = \sum_{i=1}^{m} \frac{1}{i}$$

- order the p-values $p_{(1)}, \ldots, p_{(m)}$
- find i_{max}, the largest index for which

$$p_{(i)} \leq \frac{i}{m} q$$

- Let BH_q be the rule that rejects H_{oi} for $i \leq i_{max}$, not rejecting otherwise
- change the bound under dependence

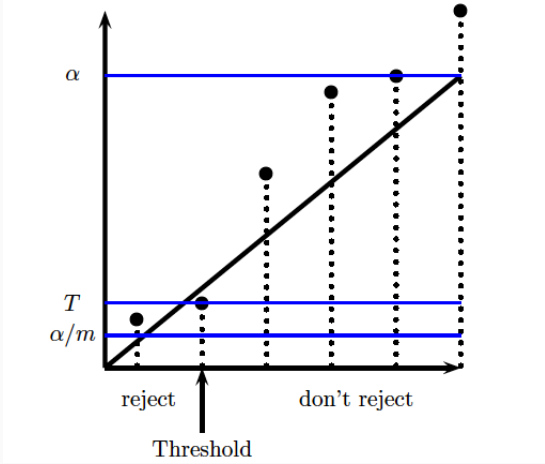
$$p_{(i)} \leq \frac{i}{m C_m} q \qquad C_m = \sum_{i=1}^{m} \frac{1}{i}$$

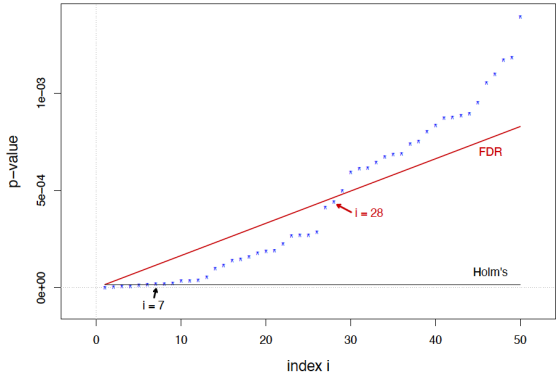
- Theorem: If the p-values corresponding to valid null hypotheses are independent of each other, then

$$FDR(BH_q) = \pi_0 q \leq q, \qquad \text{where } \pi_0 = m_0/m$$

π_0 unknown but close to 1

Benjamini-Hochberg





Example

AoS Ex.10.28

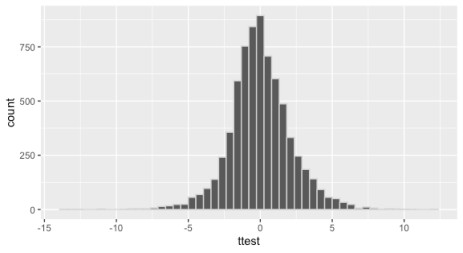
index	1	2	3	4	5	6	7	8	9	10
pval	0.00017	0.00448	0.00671	0.00907	0.01220	0.33626	0.3934	0.5388	0.5813	0.9862
cut1	0.00500	0.01000	0.01500	0.02000	0.02500	0.03000	0.0350	0.0400	0.0450	0.0500
cut2	0.00171	0.00341	0.00512	0.00683	0.00854	0.01024	0.0119	0.0137	0.0154	0.0171

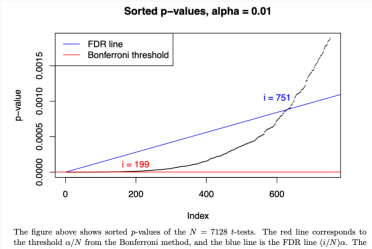
```
leukemia_big <- read.csv
 ("http://web.stanford.edu/~hastie/CASI_files/DATA/leukemia_big.csv")
 dim(leukemia_big)
  [1] 7128   72
df <- t(leukemia_big)
df_all <- df[startsWith(rownames(df), "ALL"), ]
df_aml <- df[startsWith(rownames(df), "AML"), ]
n1 <- nrow(df_all); n2 <- nrow(df_aml)
m1 <- apply(df_all, 2, mean); m2 <- apply(df_aml, 2, mean)
s1 <- apply(df_all, 2, sd); s2 <- apply(df_aml, 2, sd)

pooled <- sqrt(((n1 - 1) * s1^2 + (n2 - 1) * s2^2) / (n1 + n2 - 2))
ttest <- (m1 - m2) / pooled / sqrt(1 / n1 + 1 / n2)
pvalues <- 2 * pt(abs(ttest), df = n1 + n2 - 2, lower.tail = F)
```

Multiple testing

EH 1.2, 15.2



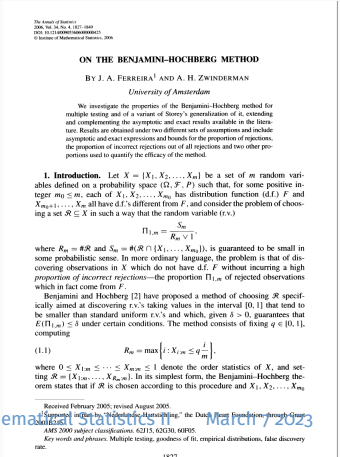


The figure above shows sorted p-values of the $N = 7128$ t-tests. The red line corresponds to the threshold α/N from the Bonferroni method, and the blue line is the FDR line $(i/N)\alpha$. The

```
> summary(ttest)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-13.52611  -1.20672  -0.08406   0.02308   1.20886  12.26065
```

Theorem: If the p-values corresponding to valid null hypotheses are independent of each other, then

$$FDR(BH_q) = \pi_0 q \leq q, \qquad \text{where } \pi_0 = m_0/m$$



The Annals of Statistics
2006, Vol. 34, No. 4, 1827–1849
DOI: 10.1214/009053606000000623
© Institute of Mathematical Statistics, 2006

ON THE BENJAMINI–HOCHBERG METHOD

BY J. A. FERREIRA[1] AND A. H. ZWINDERMAN

University of Amsterdam

We investigate the properties of the Benjamini–Hochberg method for multiple testing and of a variant of Storey's generalization of it, extending and complementing the asymptotic and exact results available in the literature. Results are obtained under two different sets of assumptions and include asymptotic and exact expressions and bounds for the proportion of rejections, the proportion of incorrect rejections out of all rejections and two other proportions used to quantify the efficacy of the method.

1. Introduction. Let $X = \{X_1, X_2, \ldots, X_m\}$ be a set of m random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ such that, for some positive integer $m_0 \leq m$, each of $X_1, X_2, \ldots, X_{m_0}$ has distribution function $\mathrm{d.f.}\ F$ and $X_{m_0+1}, \ldots, X_m$ all have d.f.'s different from F, and consider the problem of choosing a set $\mathcal{R} \subseteq X$ in such a way that the random variable (r.v.)

$$\Pi_{1,m} = \frac{S_m}{R_m \vee 1},$$

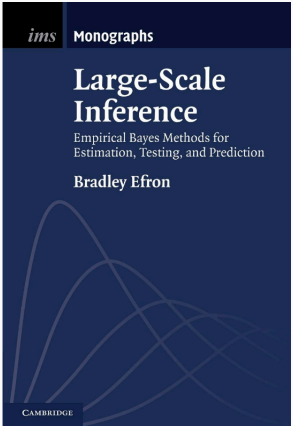
where $R_m = \#\mathcal{R}$ and $S_m = \#(\mathcal{R} \cap \{X_1, \ldots, X_{m_0}\})$, is guaranteed to be small in some probabilistic sense. In more ordinary language, the problem is that of discovering observations in X which do not have d.f. F without incurring a high *proportion of incorrect rejections*—the proportion $\Pi_{1,m}$ of rejected observations which in fact come from F.

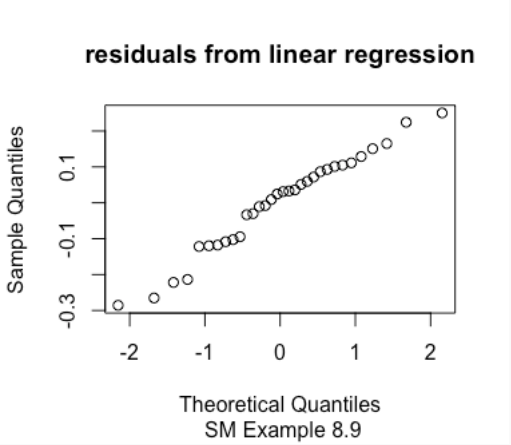
Benjamini and Hochberg [2] have proposed a method of choosing $\mathcal{R}$ specifically aimed at discovering r.v.'s taking values in the interval $[0, 1]$ that tend to be smaller than standard uniform r.v.'s and which, given $\delta > 0$, guarantees that $E(\Pi_{1,m}) \leq \delta$ under certain conditions. The method consists of fixing $q \in [0, 1]$, computing

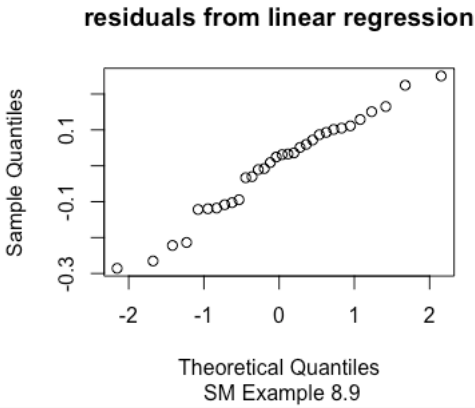
$$R_m = \max\left\{i : X_{i:m} \leq q\frac{i}{m}\right\}, \tag{1.1}$$

where $0 \leq X_{1:m} \leq \cdots \leq X_{m:m} \leq 1$ denote the order statistics of X, and setting $\mathcal{R} = \{X_{1:m}, \ldots, X_{R_m:m}\}$. In its simplest form, the Benjamini–Hochberg theorem states that if $\mathcal{R}$ is chosen according to this procedure and $X_1, X_2, \ldots, X_{m_0}$

Received February 2005; revised August 2005.
[1]Supported in part by "Stichting Mathematisch Gezelschap," the Dutch Heart Foundation through Grant 2000B061.
AMS 2000 subject classifications. 62J15, 62G30, 60F05.
Key words and phrases. Multiple testing, goodness of fit, empirical distributions, false discovery rate.




residuals from linear regression
Sample Quantiles
-0.3
-0.1
0.1
-2
-1
0
1
2
Theoretical Quantiles
SM Example 8.9



SM Example 8.9

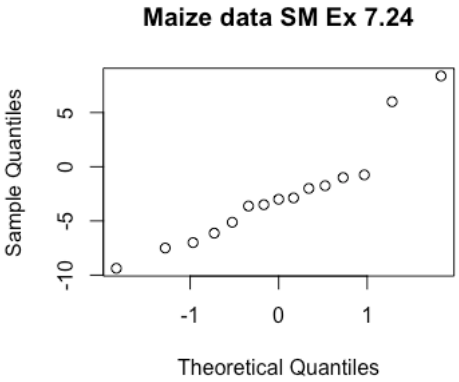
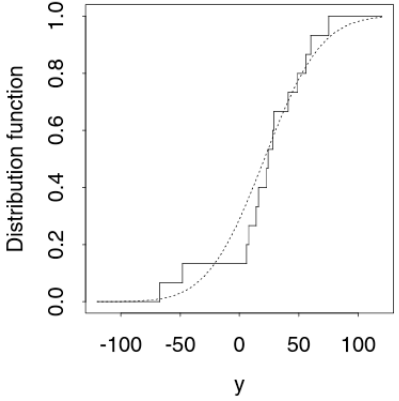
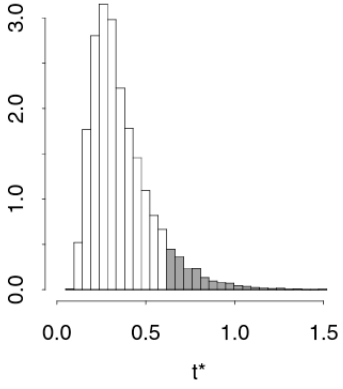


Figure 7.5 Analysis of maize data. Left: empirical distribution function for height differences, with fitted normal distribution (dots). Right: null density of Anderson–Darling statistic T for normal samples of size $n = 15$ with location and scale estimated. The shaded part of the histogram shows values of T^* in excess of the observed value t_{obs}.





SM Example 7.24 testing $N(\mu, \sigma^2)$ distribution

Smooth goodness-of-fit tests

- $X_1, \ldots, X_n \quad i.i.d. \quad F(\cdot); \quad H_0 : F = F_0$ cumulative d.f.
- $\widehat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1\{X_i \leq t\}$

- three test statistics:
  1. $\sup_t |\widehat{F}_n(t) - F_0(t)|$
  2. $\int \{\widehat{F}_n(t) - F_0(t)\}^2 dF_0(t)$
  3. $\int \frac{\{\widehat{F}_n(t) - F_0(t)\}^2}{F_0(t)\{1 - F_0(t)\}} dF_0(t)$

- SM Example 7.24 testing $N(\mu, \sigma^2)$ distribution
- SM Example 7.23; 6.14 testing $U(0, 1)$ distribution

Smooth goodness-of-fit tests

- Special case $H_0 : F(t) = F_0(t) = t$ $\quad X_i \sim U(0,1)$
- Recall

$$E_0\{\widehat{F}_n(t)\} = F_0(t) = t, \quad \text{var}\{\widehat{F}_n(t)\} = t(1-t)/n$$

- What about distribution of

$\sup_t |\widehat{F}_n(t) - t|$ $\qquad$ $\int \{\widehat{F}_n(t) - t\}^2 dt$ $\qquad$ $\int \dfrac{\{\widehat{F}_n(t) - t\}^2}{F_0(t)\{1-t\}} dt$

- need joint density of $\widehat{F}_n(t)\ \forall\ t$

- Special case $H_0 : F(t) = F_0(t) = t$ $X_i \sim U(0,1)$
- Recall

$$E_0\{\widehat{F}_n(t)\} = F_0(t) = t, \quad \text{var}\{\widehat{F}_n(t)\} = t(1-t)/n$$

- What about distribution of

$\sup_t |\widehat{F}_n(t) - t|$ $\int \{\widehat{F}_n(t) - t\}^2 dt$ $\int \frac{\{\widehat{F}_n(t) - t\}^2}{F_0(t)\{1-t\}} dt$

- need joint density of $\widehat{F}_n(t)\ \forall\ t$
- define Brownian bridge $B_n(t) = \sqrt{n}(\widehat{F}_n(t) - t)$
- vector $(B_n(t_1), \ldots, B_n(t_k)) \xrightarrow{d} N_k(0, C), \quad C_{ij} = \min(t_i, t_j) - t_i t_j$ MS 9.3
- a Brownian bridge is a continuous function on $(0,1)$ with all finite-dimensional distributions as above

Smooth goodness-of-fit tests

- Kolmogorov-Smirnov test

$$K_n = \sup_{0 \leq t \leq 1} |B_n(t)|$$

- Cramer-vonMises test

$$W_n^2 = \int_0^1 B_n^2(t) dt$$

- Anderson-Darling test

$$A_n^2 = \int_0^1 \frac{B_n^2(t)}{t(1-t)} dt$$

Smooth goodness-of-fit tests

- Kolmogorov-Smirnov test

$$K_n = \sup_{0 \le t \le 1} |B_n(t)|$$

- Cramer-vonMises test

$$W_n^2 = \int_0^1 B_n^2(t)dt$$

- Anderson-Darling test

$$A_n^2 = \int_0^1 \frac{B_n^2(t)}{t(1-t)}dt$$

- limit theorems

$$K_n \xrightarrow{d} K, \qquad W_n^2 \xrightarrow{d} \sum_{j=1}^{\infty} \frac{Z_j^2}{j^2\pi^2}, \qquad A_n^2 \xrightarrow{d} \sum_{j=1}^{\infty} \frac{Z_j^2}{j(j+1)}$$

$$\mathrm{pr}(K > x) = 2\sum_{j=1}^{\infty}(-1)^{j+1}\exp(-2j^2x^2)$$

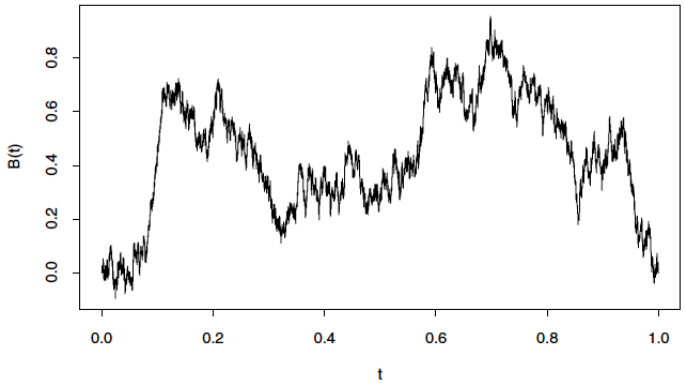


Figure 9.1 *A simulated realization of a Brownian bridge process.*

Multinomial distribution

MS 9.2; AoS 14.4; SM Ex.2.36

- $X \sim Mult_k(n; p)$

$X_j =$ number of obs in category j

- $\mathrm{pr}(X_1 = x_1, \ldots, X_k = x_k; p) =$
- $\mathrm{E}(X) =$
- $\mathrm{cov}(X) =$

AoS Thm 14.4

- $\hat{p} =$
- $\mathrm{cov}(\hat{p}) =$

Multinomial distribution

MS 9.2; AoS 14.4; SM Ex.2.36

- $X \sim Mult_k(n; p)$ — X_j = number of obs in category j
- $\mathrm{pr}(X_1 = x_1, \ldots, X_k = x_k; p) =$
- $\mathrm{E}(X) =$
- $\mathrm{cov}(X) =$ — AoS Thm 14.4
- $\hat{p} =$
- $\mathrm{cov}(\hat{p}) =$

Multinomial distribution

AoS §14.4; SM Ex.2.36

- log-likelihood function
- Fisher information

Multinomial goodness-of-fit tests

AoS 10.8; MS 9.2

- $H_0 : X_1, \ldots, X_n$ i.i.d. $f(x; \theta), \quad x \in \mathbb{R}; \quad \theta \in \mathbb{R}^s$
- Let $I_1, \ldots, I_k$ be disjoint intervals on $\mathbb{R}$
- Define $N_j = \sum_{i=1}^{n} 1\{X_i \in I_j\}$
- $(N_1, \ldots, N_k) \sim$

- $H_0 : X_1, \ldots, X_n$ i.i.d. $f(x; \theta), \quad x \in \mathbb{R}; \quad \theta \in \mathbb{R}^s$
- Let $I_1, \ldots, I_k$ be disjoint intervals on $\mathbb{R}$
- Define $N_j = \sum_{i=1}^n 1\{X_i \in I_j\}$
- $(N_1, \ldots, N_k) \sim$

- $L(\theta) = \prod_{j=1}^k p_j(\theta)^{N_j}; \quad \ell(\theta) = \sum_{j=1}^k N_j \log\{p_j(\theta)\}; \quad \tilde{\theta} = \arg\max_\theta \ell(\theta)$

Multinomial goodness-of-fit tests

AoS 10.8; MS 9.2

- $H_0 : X_1, \ldots, X_n$ i.i.d. $f(x;\theta), \quad x \in \mathbb{R}; \quad \theta \in \mathbb{R}^s$
- Let $I_1, \ldots, I_k$ be disjoint intervals on $\mathbb{R}$
- Define $N_j = \sum_{i=1}^n 1\{X_i \in I_j\}$
- $(N_1, \ldots, N_k) \sim$

- $L(\theta) = \prod_{j=1}^k p_j(\theta)^{N_j}; \quad \ell(\theta) = \sum_{j=1}^k N_j \log\{p_j(\theta)\}; \quad \tilde{\theta} = \arg\max_\theta \ell(\theta)$

- Theorem 10.29 (AoS): Under H_0, MS Thm 9.2

$$Q = \sum_{j=1}^k \frac{\{N_j - np_j(\tilde{\theta})\}^2}{np_j(\tilde{\theta})} \xrightarrow{d} \chi^2_{k-1-s}$$

- Theorem 9.1 (MS): Under H_0

$$W = 2\sum_{j=1}^k N_j \log\left(\frac{N_j}{np_j(\tilde{\theta})}\right) \xrightarrow{d} \chi^2_{k-1-s}$$

Table 9.1 *Frequency of goals in First Division matches and "expected" frequency under Poisson model in Example 9.2*

Goals	0	1	2	3	4	≥ 5
Frequency	252	344	180	104	28	16
Expected	248.9	326.5	214.1	93.6	30.7	10.2

$$p_0(\lambda) = 1 - \sum_{j=0}^{4} p_j(\lambda); \quad p_j(\lambda) = e^{-\lambda}\lambda^j/j!, \quad \tilde{\lambda} = 1.3118$$

$$Q = 11.09; \quad W = 10.87; \quad \mathrm{pr}(\chi^2_4 > [11.09, 10.87]) = [0.026, 0.028]$$

		Antigen 'B'		
		Absent	Present	Total
Antigen 'A'	Absent	'O': 202	'B': 35	237
	Present	'A': 179	'AB': 6	185
Total		381	41	422

Table 4.3 Blood groups in England (Taylor and Prior, 1938). The upper part of the table shows a cross-classification of 422 persons by presence or absence of antigens 'A' and 'B', giving the groups 'A', 'B', 'AB', 'O' of the human blood group system. The lower part shows genotypes and corresponding probabilities under one- and two-locus models. See Example 4.38 for details.

	Two-locus model		One-locus model	
Group	Genotype	Probability	Genotype	Probability
'A'	$(AA;bb), (Aa;bb)$	$\alpha(1-\beta)$	$(AA), (AO)$	$\lambda_A^2 + 2\lambda_A\lambda_O$
'B'	$(aa;BB), (aa;Bb)$	$(1-\alpha)\beta$	$(BB), (BO)$	$\lambda_B^2 + 2\lambda_B\lambda_O$
'AB'	$(AA;BB), (Aa;BB), (AA;Bb), (Aa;Bb)$	$\alpha\beta$	(AB)	$2\lambda_A\lambda_B$
'O'	$(aa;bb)$	$(1-\alpha)(1-\beta)$	(OO)	λ_O^2

$Q = 15.73; W = 17.66$ (two-locus)

$p < 10^{-5}$

$Q = 2.82; W = 3.17$ (single locus)

$p = 0.09; 0.07$