Mathematical Statistics II

STA2212H S LEC0101

Week 5

February 7 2023

Science & technology | Archers and heart rates

How to measure how stress affects athletes' performance

Pick a sport where they don't move much, and study skin flushing

Jan 25th 2023

Save Share Give

Science & technology | Archers and heart rates

How to measure how stress affects athletes' performance

Pick a sport where they don't move much, and study skin flushing

Jan 25th 2023

Save Share Give

link

The Economist

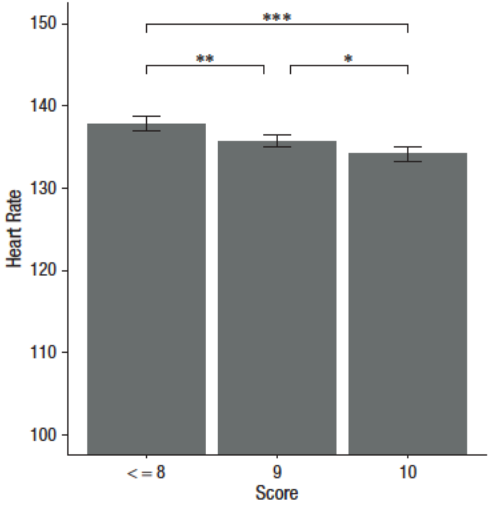


Fig. 1. Heart rate and score. *Heart rate* is the mean of all the available data points before the shot in beats per minute (bpm). We show

Today

1. Recap
2. Confidence bounds, intervals, regions
3. Approximate CIs from likelihood theory
4. Normal approximation to posterior
5. Project

Upcoming seminar of interest

- February 13 3.30 – 4.30 Sandrine Dudoit Details
  "Learning from Data in Single-Cell Transcriptomics"

- February 22-23 Toronto Workshop on Reproducibility

Recap

- Bayesian inference: philosophy, choice of priors
- optimality of Bayes estimators finite sample
- Hierarchical Bayes models, shrinkage estimators

Recap

- Bayesian inference: philosophy, choice of priors
- optimality of Bayes estimators — finite sample
- Hierarchical Bayes models, shrinkage estimators

- Multi-parameter posterior distributions; marginalization

Recap

- Bayesian inference: philosophy, choice of priors
- optimality of Bayes estimators — finite sample
- Hierarchical Bayes models, shrinkage estimators

- Multi-parameter posterior distributions; marginalization

- confidence intervals and bounds
- exact CIs using pivotal inversion
- approximate CIs using limiting distributions

Confidence region

- multiparameter model $f(\boldsymbol{X};\theta)$

$$\mathrm{pr}_\theta\{\boldsymbol{\theta} \in R(\boldsymbol{X})\} \geq 1-\alpha,$$

for all $\boldsymbol{\theta}$, with equality for some $\boldsymbol{\theta}$

- pivotal method:

$$1-\alpha = \mathrm{pr}_\theta\{a \leq g(\boldsymbol{X};\boldsymbol{\theta}) \leq b\} = \mathrm{pr}_\theta\{\boldsymbol{\theta} \in R(\boldsymbol{X})\}$$

- Example: $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ i.i.d. $N_p(\boldsymbol{\mu}, \Sigma)$ MS Ex.7.8
- exact pivot

$$g(\boldsymbol{X};\boldsymbol{\mu}) = \frac{n(n-k)}{k(n-1)}(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu})^T\hat{\Sigma}^{-1}(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu})$$

- multiparameter model $f(\boldsymbol{X}; \theta)$

$$\mathrm{pr}_\theta\{\boldsymbol{\theta} \in R(\boldsymbol{X})\} \geq 1 - \alpha,$$

for all θ, with equality for some θ

- pivotal method:

$$1 - \alpha = \mathrm{pr}_\theta\{a \leq g(\boldsymbol{X}; \theta) \leq b\} = \mathrm{pr}_\theta\{\boldsymbol{\theta} \in R(\boldsymbol{X})\}$$

- Example: $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ i.i.d. $N_p(\boldsymbol{\mu}, \Sigma)$ MS Ex.7.8
- exact pivot

$$g(\boldsymbol{X}; \boldsymbol{\mu}) = \frac{n(n-k)}{k(n-1)}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \hat{\Sigma}^{-1}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})$$

-

$$R(\boldsymbol{X}) = \{\boldsymbol{\mu} : \frac{n(n-k)}{k(n-1)}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \hat{\Sigma}^{-1}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \leq f_{1-\alpha}\}$$

Approximate confidence intervals

- maximum likelihood estimator is approximately normal
- $$\widehat{\theta} \dot\sim N\{\theta, I_n^{-1}(\widehat{\theta})\} \implies (\widehat{\theta} - \theta)^T I_n(\widehat{\theta})(\widehat{\theta} - \theta) \dot\sim \chi^2_k$$
- $$1 - \alpha \approx \mathrm{pr}_\theta\{\theta \in R(\widehat{\theta})\}$$
- $$R(\widehat{\theta}) = \{\theta : (\widehat{\theta} - \theta)^T I_n(\widehat{\theta})(\widehat{\theta} - \theta) \leq \chi^2_{k,1-\alpha}\}$$

Approximate confidence intervals

- maximum likelihood estimator is approximately normal
- $$\widehat{\boldsymbol{\theta}} \dot{\sim} N\{\boldsymbol{\theta}, I_n^{-1}(\widehat{\boldsymbol{\theta}})\} \implies (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T I_n(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \dot{\sim} \chi^2_k$$
- $$1 - \alpha \approx \mathrm{pr}_\theta\{\theta \in R(\widehat{\boldsymbol{\theta}})\}$$
- $$R(\widehat{\boldsymbol{\theta}}) = \{\boldsymbol{\theta} : (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T I_n(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq \chi^2_{k,1-\alpha}\}$$
- $k = 1$:
  $$\hat{\theta} \pm z_{1-\alpha/2}\widehat{se}(\hat{\theta})$$

AoS Thm 6.16

Profile likelihood intervals

- model $Y \sim f(y;\psi,\lambda), \quad \psi \in \mathbb{R}, \lambda \in \mathbb{R}^{d-1}, \quad \theta = (\psi,\lambda)$ — $y = (y_1,\dots,y_n)$
- log-likelihood function $\ell(\psi,\lambda;y) = \log f(y;\psi,\lambda) = \sum \log f(y_i;\psi,\lambda)$ — if independent
- profile log-likelihood function $\ell_p(\psi) = \ell(\psi,\hat\lambda_\psi)$ — maximize over λ
- maximum likelihood estimate — $j_p(\psi) = -\ell_p''(\psi)$

$$\hat\psi \mathrel{\dot\sim} N\{\psi, j_p^{-1/2}(\psi)\} \implies 1-\alpha \text{ CI } \approx \hat\psi \pm z_{1-\alpha/2}\,\hat{j}_p^{1/2}$$

- likelihood ratio test

$$2\{\ell_p(\hat\psi) - \ell_p(\psi)\} \mathrel{\dot\sim} \chi_1^2 \implies 1-\alpha \text{ CI } \approx \{\psi : 2\{\ell_p(\hat\psi) - \ell_p(\psi)\} \le \chi^2_{1,1-\alpha}\}$$

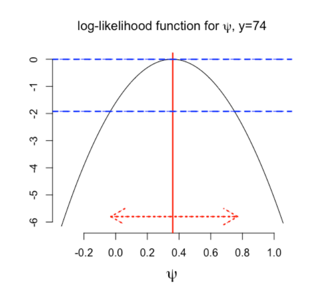
Profile likelihood intervals

- model $Y \sim f(y; \psi, \lambda), \quad \psi \in \mathbb{R}, \lambda \in \mathbb{R}^{d-1}, \quad \theta = (\psi, \lambda)$ $\qquad y = (y_1, \ldots, y_n)$
- log-likelihood function $\ell(\psi, \lambda; y) = \log f(y; \psi, \lambda) = \sum \log f(y_i; \psi, \lambda)$ $\qquad$ if independent
- profile log-likelihood function $\ell_p(\psi) = \ell(\psi, \hat{\lambda}_\psi)$ $\qquad$ maximize over λ
- maximum likelihood estimate

$$\hat{\psi} \dot{\sim} N\{\psi, j_p^{-1/2}(\psi)\} \implies 1 - \alpha \text{ CI } \approx \hat{\psi} \pm z$$

- likelihood ratio test

$$2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\} \dot{\sim} \chi_1^2 \implies 1 - \alpha \text{ CI } \approx \{\psi : 2\{\ell_p(\hat{\psi})$$



Bayesian interval estimation

- upper and lower bounds

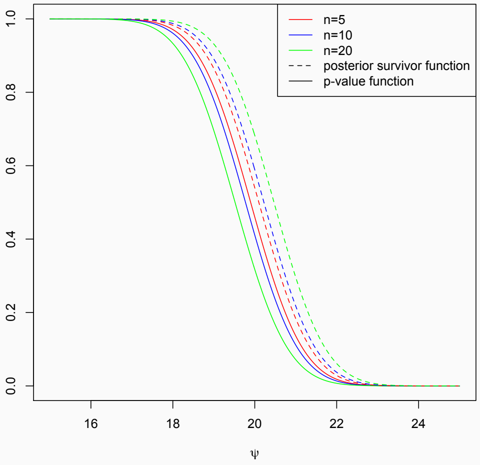
Bayesian interval estimation

- upper and lower bounds
- equi-tailed posterior intervals

Bayesian interval estimation

- upper and lower bounds
- equi-tailed posterior intervals
- highest posterior density

Upper and lower bounds



Approximate normality of posterior

- $X_1, \ldots, X_n \sim f(\boldsymbol{x} \mid \theta), \qquad \theta \sim \pi(\theta), \qquad \pi(\theta \mid x^n) = \dfrac{f(\boldsymbol{x} \mid \theta)}{f(\boldsymbol{x})}$ $\qquad \boldsymbol{x} = (x_1, \ldots, x_n)$

Approximate normality of posterior

- $X_1, \ldots, X_n \sim f(\boldsymbol{x} \mid \theta), \qquad \theta \sim \pi(\theta), \qquad \pi(\theta \mid x^n) = \dfrac{f(\boldsymbol{x} \mid \theta)}{f(\boldsymbol{x})}$ $\qquad \boldsymbol{x} = (x_1, \ldots, x_n)$
- $\pi(\theta \mid \boldsymbol{x}) \approx N\{\hat{\theta}, j^{-1}(\hat{\theta})\}; \qquad \pi(\theta \mid \boldsymbol{x}) \approx N\{\tilde{\theta}, \tilde{\jmath}^{-1}(\tilde{\theta})\}$

Approximate normality of posterior

- $X_1, \ldots, X_n \sim f(\boldsymbol{x} \mid \theta), \qquad \theta \sim \pi(\theta), \qquad \pi(\theta \mid x^n) = \dfrac{f(\boldsymbol{x} \mid \theta)}{f(\boldsymbol{x})}$ $\qquad \boldsymbol{x} = (x_1, \ldots, x_n)$
- $\pi(\theta \mid \boldsymbol{x}) \approx N\{\hat\theta, j^{-1}(\hat\theta)\}; \qquad \pi(\theta \mid \boldsymbol{x}) \approx N\{\tilde\theta, \tilde\jmath^{-1}(\tilde\theta)\}$
- careful statement $\qquad$ Berger, 1985; Ch.4
  - For any $a, b \in \mathbb{R}, a < b$
  - let $a_n = \hat\theta_n + a j^{-1/2}(\hat\theta_n)$, $b_n = \hat\theta_n + b j^{-1/2}(\hat\theta_n)$
  - $\hat\theta_n$ is the solution of $\ell'(\theta; \boldsymbol{x}) = 0$, assumed unique, and $j(\theta) = -\ell''(\theta; \boldsymbol{x})$

  Then
$$\int_{a_n}^{b_n} \pi(\theta \mid \boldsymbol{x}) d\theta \longrightarrow \Phi(b) - \Phi(a), \quad n \to \infty.$$

need $\pi(\theta) > 0, \pi'(\theta)$ continuous

Approximate normality of posterior

- $X_1, \ldots, X_n \sim f(\boldsymbol{x} \mid \theta), \qquad \theta \sim \pi(\theta), \qquad \pi(\theta \mid \boldsymbol{x}) = \dfrac{f(x^n\boldsymbol{x} \mid \theta)}{f(\boldsymbol{x})}$ $\quad v = (x_1, \ldots, x_n)$

- $\pi(\theta \mid \boldsymbol{x}) \approx N\{\hat{\theta}, j^{-1}(\hat{\theta})\}; \qquad \pi(\theta \mid \boldsymbol{x}) \approx N\{\tilde{\theta}, \tilde{\jmath}(\tilde{\theta})\}$

- approximate posterior probability intervals $\quad \tilde{\theta} \approx \hat{\theta}$

- exact posterior probability intervals

Link to Guardian

Pfizer-BioNTech vaccine trial:

vaccine: 22000 subjects, 8 cases

placebo: 22000 subjects, 162 cases

$8/162 = 5\% \Longrightarrow 95\%$ efficacy

data released November 18 2020 link

published December 31 2020 in NEJM link

Behind the numbers: what does it mean if a Covid vaccine has '90% efficacy'?

David Spiegelhalter and Anthony Masters

Confusion surrounds the vaccines' effectiveness. The leading Cambridge professor clarifies the data behind the trials

▲ People rest in Salisbury Cathedral, England, after receiving the Pfizer/BioNTech vaccine. Photograph: Neil Hall/EPA

Pfizer publication

NEJM January 2021

Editor's Note: This article was published on December 10, 2020, at NEJM.org.

ORIGINAL ARTICLE

Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine

Fernando P. Polack, M.D., Stephen J. Thomas, M.D., Nicholas Kitchin, M.D., Judith Absalon, M.D., Alejandra Gurtman, M.D., Stephen Lockhart, D.M., John L. Perez, M.D., Gonzalo Pérez Marc, M.D., Edson D. Moreira, M.D., Cristiano Zerbini, M.D., Ruth Bailey, B.Sc., Kena A. Swanson, Ph.D., et al., for the C4591001 Clinical Trial Group*

Article Figures/Media Metrics

13 References 263 Citing Articles Letters

December 31, 2020
N Engl J Med 2020; 383:2603-2615
DOI: 10.1056/NEJMoa2034577
Chinese Translation 中文翻译

Results: A total of 43,548 participants underwent randomization, of whom 43,448 received injections: 21,720 with BNT162b2 and 21,728 with placebo. There were 8 cases of Covid-19 with onset at least 7 days after the second dose among participants assigned to receive BNT162b2 and 162 cases among those assigned to placebo; BNT162b2 was 95% effective in preventing Covid-19 (95% credible interval, 90.3 to 97.6).

Pfizer publication

NEJM

Table 2. Vaccine Efficacy against Covid-19 at Least 7 days after the Second Dose.*

Efficacy End Point	BNT162b2		Placebo		Vaccine Efficacy, % (95% Credible Interval)‡	Posterior Probability (Vaccine Efficacy >30%)§
	No. of Cases	Surveillance Time (n)†	No. of Cases	Surveillance Time (n)†		
	(N=18,198)		(N=18,325)			
Covid-19 occurrence at least 7 days after the second dose in participants without evidence of infection	8	2.214 (17,411)	162	2.222 (17,511)	95.0 (90.3–97.6)	>0.9999
	(N=19,965)		(N=20,172)			
Covid-19 occurrence at least 7 days after the second dose in participants with and those without evidence of infection	9	2.332 (18,559)	169	2.345 (18,708)	94.6 (89.9–97.3)	>0.9999

* The total population without baseline infection was 36,523; total population including those with and those without prior evidence of infection was 40,137.

† The surveillance time is the total time in 1000 person-years for the given end point across all participants within each group at risk for the end point. The time period for Covid-19 case accrual is from 7 days after the second dose to the end of the surveillance period.

‡ The credible interval for vaccine efficacy was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.

§ Posterior probability was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.

Credible intervals

- vaccine group 18000 participants; 8 cases
- placebo group 18000 participants; 162 cases
- $0.05 = 8/162 \longrightarrow$ 95% efficacy
- model

  $$X_1 \sim Poisson(\lambda\psi), \quad X_2 \sim Poisson(\lambda) \qquad X_1 \mid S = X_1 + X_2 \sim Binom(S, \psi/(1+\psi))$$
- prior $Beta(a, b) \longrightarrow$ posterior $Beta(x_1 + a, s - x_1 + b)$

Credible intervals

- vaccine group 18000 participants; 8 cases
- placebo group 18000 participants; 162 cases
- $0.05 = 8/162 \longrightarrow$ 95% efficacy
- model
  $X_1 \sim Poisson(\lambda\psi), \quad X_2 \sim Poisson(\lambda) \qquad X_1 \mid S = X_1 + X_2 \sim Binom(S, \psi/(1+\psi))$
- prior $Beta(a, b) \longrightarrow$ posterior $Beta(x_1 + a, s - x_1 + b)$

- `qbeta(c(0.025,0.975), shape1 = 8.7, shape2 = 163)`
- `> [1] 0.02319 0.08799`
- `1- .Last.value/(1-.Last.value)`
- `> [1] 0.97625 0.90352`

$VE = 1 - \psi$

Highest posterior density (HPD) regions

- HPD region C for θ:

$$(1) \qquad \int_C \pi(\theta \mid \boldsymbol{x}) = 1 - \alpha$$

$$(2) \qquad \pi(\theta \mid \boldsymbol{x}) \geq \pi(\theta^* \mid \boldsymbol{x})$$

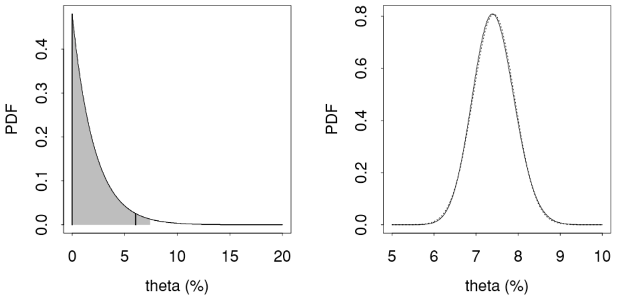


Figure 11.1 Cardiac surgery data. Left panel: posterior density for θ_A, showing boundaries of 0.95 highest posterior credible interval (vertical lines) and region between posterior 0.025 and 0.975 quantiles of $\pi(\theta_A \mid y)$ (shaded). Right panel: exact posterior beta density for overall mortality rate θ (solid) and normal approximation (dots).

Highest posterior density (HPD) regions

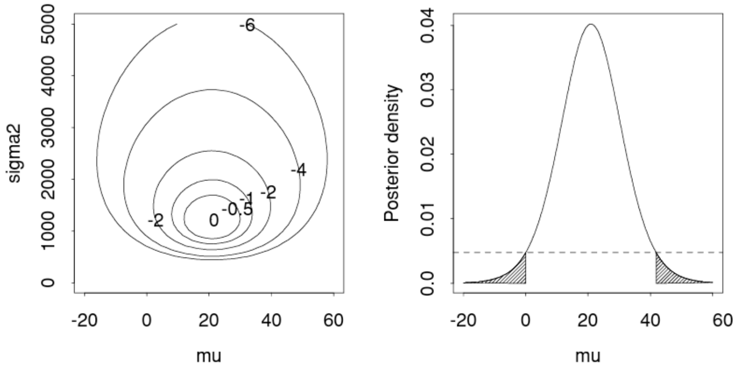


Figure 11.2 Posterior densities of (μ, σ^2) of normal model for maize data. Left: contours of the normalized log joint posterior density. Right: marginal posterior density for μ, showing 95% HPD credible set, which is the set of values of μ whose values of the posterior density $\pi(\mu \mid y)$ lie above the dashed line. The shaded region has area 0.05.

Likelihood ratio based approximate confidence regions

- $X_1, \ldots, X_n \sim f(\boldsymbol{x}; \theta)$
- $L(\theta; \boldsymbol{x}) = f(\boldsymbol{x}; \theta), \quad \ell(\theta) = \log L(\theta; \boldsymbol{x})$
- $$w(\theta) = 2\{\ell(\hat{\theta}) - \ell(\theta)\} \xrightarrow{d} \chi^2_p, \quad n \to \infty$$

Likelihood ratio based approximate confidence regions

- $X_1, \dots, X_n \sim f(\boldsymbol{x}; \boldsymbol{\theta})$
- $L(\boldsymbol{\theta}; \boldsymbol{x}) = f(\boldsymbol{x}; \boldsymbol{\theta}), \quad \ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}; \boldsymbol{x})$
- $$w(\boldsymbol{\theta}) = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta})\} \xrightarrow{d} \chi^2_p, \quad n \to \infty$$
- approximation:
$$w(\boldsymbol{\theta}) \dot{\sim} \chi^2_p$$
- approximate confidence region
$$\{\boldsymbol{\theta} : w(\boldsymbol{\theta}) \le \chi^2_{p,1-\alpha}\}$$

Nonparametric confidence bands

- recall $X_1, \dots, X_n, i.i.d. F(\cdot)$
- empirical cdf

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} 1\{X_{(i)} \leq t\}$$

- properties:

$$\mathrm{E}\{\hat{F}_n(t)\} = F(t), \quad \mathrm{var}\{\hat{F}_n(t)\} = \frac{1}{n} F(t)\{1 - F(t)\}$$

any fixed t

- pointwise approximate confidence limits $\hat{F}_n(t) \pm z_{1-\alpha/2}[\hat{F}_n(t)\{1 - \hat{F}_n(t)\}]^{1/2}$

Nonparametric confidence bands

- recall $X_1, \ldots, X_n, i.i.d. F(\cdot)$
- empirical cdf

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_{(i)} \leq t\}$$

- properties:

$$\mathrm{E}\{\hat{F}_n(t)\} = F(t), \quad \mathrm{var}\{\hat{F}_n(t)\} = \frac{1}{n}F(t)\{1 - F(t)\}$$

any fixed t

- pointwise approximate confidence limits $\hat{F}_n(t) \pm z_{1-\alpha/2}[\hat{F}_n(t)\{1 - \hat{F}_n(t)\}]^{1/2}$
- simultaneous confidence band : $\mathrm{pr}\{L(t) \leq F(t) \leq U(t) \text{ for all } t\} \geq 1 - \alpha$:

$$L(t) = \max\{\hat{F}_n(t) - \epsilon_n, 0\}, \quad U(t) = \min\{\hat{F}_n(t) - \epsilon_n, 1\}, \quad \epsilon_n = \left\{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)\right\}^{1/2}$$

… Nonparametric confidence bands

AoS 7.4

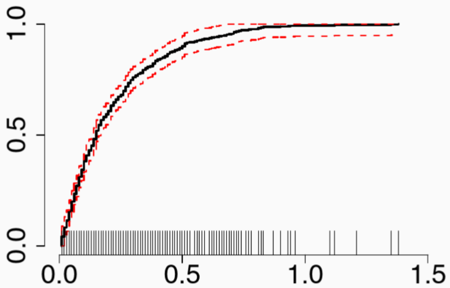


FIGURE 7.1. Nerve data. Each vertical line represents one data point. The solid line is the empirical distribution function. The lines above and below the middle line are a 95 percent confidence band.

7.2 Example (Nerve Data). Cox and Lewis (1966) reported 799 waiting times between successive pulses along a nerve fiber. Figure 7.1 shows the empirical CDF $\widehat{F}_n$. The data points are shown as small vertical lines at the bottom of the plot. Suppose we want to estimate the fraction of waiting times between .4 and .6 seconds. The estimate is $\widehat{F}_n(.6) - \widehat{F}_n(.4) = .93 - .84 = .09$. ∎

Kernel regression

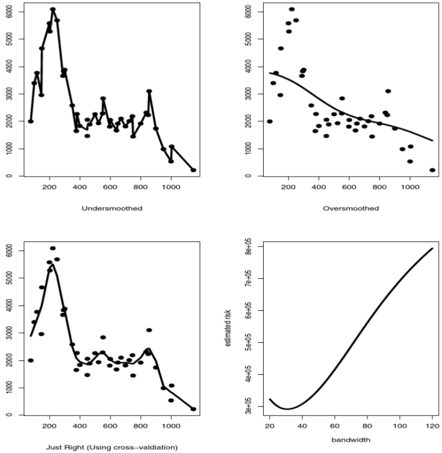


FIGURE 20.8. Regression analysis of the CMB data. The first fit is undersmoothed, the second is oversmoothed, and the third is based on cross-validation. The last panel shows the estimated risk versus the bandwidth of the smoother. The data are from BOOMERaNG, Maxima, and DASI.

Kernel regression

Confidence Bands for Kernel Regression

An approximate $1-\alpha$ confidence band for $\overline{r}_n(x)$ is

$$\ell_n(x) = \widehat{r}_n(x) - q\ \widehat{\mathsf{se}}(x), \quad u_n(x) = \widehat{r}_n(x) + q\ \widehat{\mathsf{se}}(x) \tag{20.37}$$

where

$$\begin{aligned} \widehat{\mathsf{se}}(x) &= \widehat{\sigma}\sqrt{\sum_{i=1}^{n} w_i^2(x)}, \\ q &= \Phi^{-1}\left(\frac{1+(1-\alpha)^{1/m}}{2}\right), \\ m &= \frac{b-a}{\omega}, \end{aligned}$$

$\widehat{\sigma}$ is defined in (20.36) and ω is the width of the kernel. In case the kernel does not have finite width then we take ω to be the effective width, that is, the range over which the kernel is non-negligible. In particular, we take $\omega = 3h$ for the Normal kernel.

Kernel regression

AoS 20.4

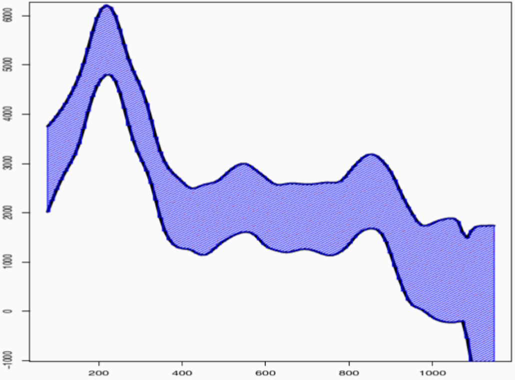


FIGURE 20.9. 95 percent confidence envelope for the CMB data.

$X_1, \ldots, X_n \sim f(\boldsymbol{x}; \theta)$

- Null and alternative hypothesis
- Test function
- Rejection region
- Type I and Type II error
- Power and Size

… Hypothesis testing: intro

$X_1, \ldots, X_n \sim f(\boldsymbol{x}; \theta)$

- Null and alternative hypothesis: $H_0 : \theta \in \Theta_0$; $H_1 : \theta \in \Theta_1$, $\quad \Theta_0 \cup \Theta_1 = \Theta$
- Test (decision) function: $\phi : \mathcal{X} \to \{0, 1\}$
  $\phi(\boldsymbol{X}) = 1$ decide $\theta \in \Theta_1$, else decide $\theta \in \Theta_0$
- Rejection region: $R \subset \mathcal{X}$; if $\boldsymbol{x} \in R$ "reject" H_0 $\quad R = \{\boldsymbol{x} : \phi(\boldsymbol{x}) = 1\}$
- Type I and Type II error: $\Pr\{\boldsymbol{X} \in R \mid \theta \in \Theta_0\}$, $\quad \Pr\{\boldsymbol{X} \notin R \mid \theta \in \Theta_1\}$
- Power and Size: $\beta(\theta) = \Pr_\theta(X \in R)$ $\quad \alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$
- Optimal tests: among all level-α tests, find that with the highest power under H_1

level-α means size $\leq \alpha$

... Hypothesis testing: intro

- goal is to identify R, or $\phi(\cdot)$ with small Type I and Type II errors
- can't reduce both errors at once

see text following Ex. 7.10

- classical solution: require

$$E_\theta\{\phi(\boldsymbol{X})\} \leq \alpha, \quad \theta \in \Theta_0$$

- subject to this constraint, minimize

$$E_\theta\{\phi(\boldsymbol{X})\}, \quad \theta \in \Theta_1$$

- goal is to identify R, or $\phi(\cdot)$ with small Type I and Type II errors
- can't reduce both errors at once — see text following Ex. 7.10
- classical solution: require
$$E_\theta\{\phi(\boldsymbol{X})\} \leq \alpha, \quad \theta \in \Theta_0$$
- subject to this constraint, minimize
$$E_\theta\{\phi(\boldsymbol{X})\}, \quad \theta \in \Theta_1$$
- find a test statistic, $T = t(\boldsymbol{X})$, and $\phi(\boldsymbol{X}) = 1\{T \geq t_{crit}\}$ — t_{crit} to be determined

Example: Two-sample t-test

EH §1.2

1.2 Hypothesis Testing

Our second example concerns the march of methodology and inference for *hypothesis testing* rather than estimation: 72 leukemia patients, 47 with **ALL** (acute lymphoblastic leukemia) and 25 with **AML** (acute myeloid leukemia, a worse prognosis) have each had genetic activity measured for a panel of 7,128 genes. The histograms in Figure 1.4 compare the genetic activities in the two groups for gene 136.

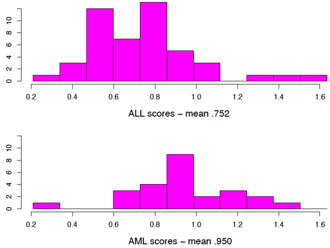
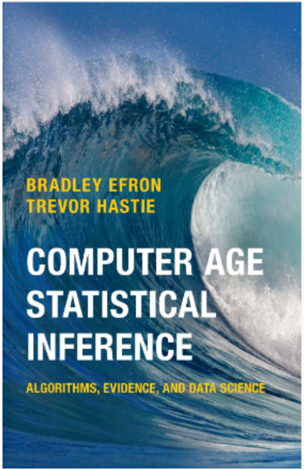


Figure 1.4 Scores for gene 136, leukemia data. Top **ALL** ($n = 47$), bottom **AML** ($n = 25$). A two-sample t-statistic $= 3.01$ with p-value $= .0036$.

The **AML** group appears to show greater activity, the mean values being

$$\overline{\texttt{ALL}} = 0.752 \quad \text{and} \quad \overline{\texttt{AML}} = 0.950. \qquad (1.5)$$



```
leukemia_big <- read.csv
  ("http://web.stanford.edu/~hastie/CASI_files/DATA/leukemia_big.csv")
oneline <- leukemia_big[136,]
one <- c(1:20, 35:61) #  I had to extract these manually,
two <- c(21:34, 62:72) #  couldn't figure out the data frame
n1 <- length(one); n2 <- length(two)
mean_one <- sum(oneline[1,one])/n1.  ##[1] 0.7524794
mean_two <- sum(oneline[1,two])/n2.  ##[1] 0.9499731
var_one <-  sum((oneline[1,one]-mean_one)^2)/(n1-1)
var_two <- sum((oneline[1,two]-mean_two)^2)/(n2-1)
pooled <- ((n1-1)*var_one + (n2-1)*var_two)/(n1+n2-1)
taos <- (mean_one-mean_two)/sqrt((var_one/n1)+(var_two/n2))
##[1] -3.132304
tbe <- (mean_one-mean_two)/sqrt(pooled*((1/n1)+(1/n2)))
##[1] -3.035455
```

- model
- null and alternative hypothesis
- rejection region
- test statistics and critical value
- type I and type II error

Example: comparing two proportions

- $X \sim Binom(m, p_1), Y \sim Binom(n, p_2)$
- $\delta = p_1 - p_2; H_0 : \delta = 0$
- maximum likelihood estimate of δ
- estimated standard error

- same test set: $D_i = X_i - Y_i$