

# 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

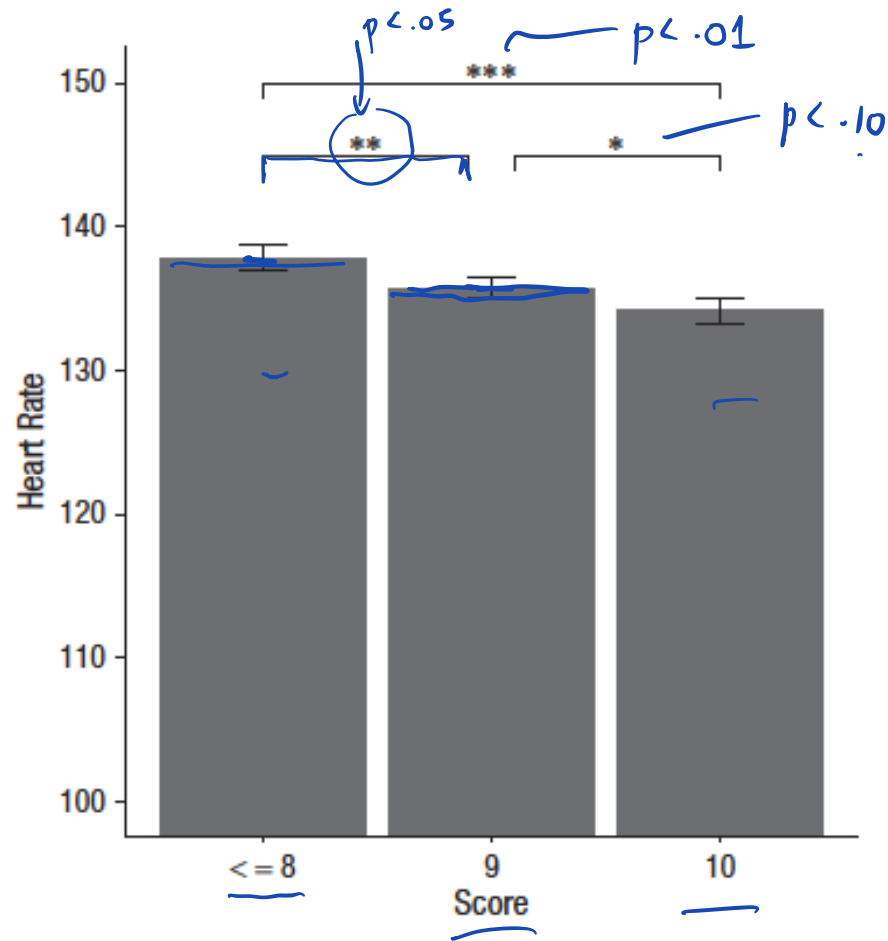
Save

Share

Give

link

The  
Economist



$$I \sqrt{\text{var}(\hat{x}_j)}$$

$$j = 1 \dots 3$$

??

**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 the mean and the standard error of heart rate for each of the three

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

List ready by reading week

5. Testing  
 6. Project

← partner (x)

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  $\rightarrow$  admissible (freq't notion) finite sample
- Hierarchical Bayes models, shrinkage estimators

$$y_i | \theta_i \sim N(\theta_i, 1)$$

model

$$\theta_i | \mu \sim N(\mu, \tau^2)$$

prior

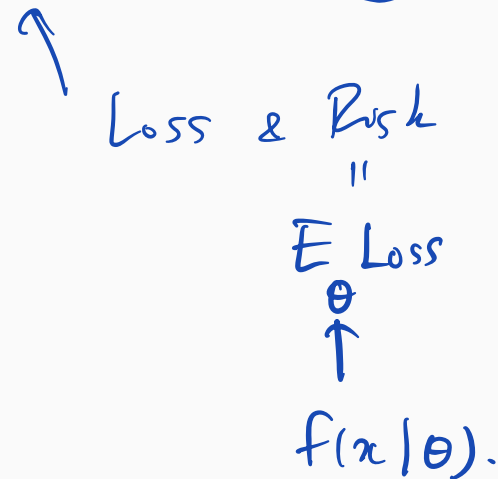
$$\mu \sim N(0, b^2)$$

hyper

$b^2$  large

$$\tau^2 \sim \text{IG}(\nu, \lambda) \leftarrow \text{still need to fix } \nu, \lambda$$

$\uparrow$  conjugate



# Recap

- Bayesian inference: philosophy, choice of priors
- optimality of Bayes estimators
- Hierarchical Bayes models, shrinkage estimators
- Multi-parameter posterior distributions; marginalization

finite sample

# Recap

- Bayesian inference: philosophy, choice of priors
  - optimality of Bayes estimators
  - Hierarchical Bayes models, shrinkage estimators
  - Multi-parameter posterior distributions; marginalization
  - confidence intervals and bounds
  - exact CIs using pivotal inversion ←
  - approximate CIs using limiting distributions ←
- 

finite sample

$$\theta = (\psi, \underline{\lambda})$$

$$\underbrace{\pi_m(\psi | \underline{x})}_{\substack{\uparrow \\ \text{marg'l post.}}} = \int \underbrace{\pi(\theta | \underline{x})}_{\substack{\uparrow \\ \text{full post.}}} d\underline{\lambda}$$

↑  
should check for  
influence of prior

$$\overline{\hat{\theta}} \pm \dots$$

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

$$X_i \sim N(z_i^T \beta, \sigma^2)$$

CR for  $\beta$

- pivotal method:

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

- Example:  $\mathbf{X}_1, \dots, \mathbf{X}_n$  i.i.d.  $N_k(\mu, \Sigma)$

- exact pivot

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

$$\hat{\mu} = \bar{\mathbf{X}}$$

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$$

↑  
kde  
↑  
ntbc

$$P\{R(\mathbf{X}) \text{ includes true } \theta\} \geq 1 - \alpha$$

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

$$\theta \in \mathbb{R}^k$$

for all  $\theta$ , with equality for some  $\theta$

$$X_i \sim N(\mu, \sigma^2)$$

$$i = 1, \dots, n$$

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \approx t_{n-1}$$

MS Ex.7.8

$$g(\mathbf{X}, \mu)$$

$$g(\mathbf{X}; \mu)$$

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

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



for all  $\theta$ , with equality for some  $\theta$

- pivotal method:

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

- Example:  $\mathbf{X}_1, \dots, \mathbf{X}_n$  i.i.d.  $N_p(\mu, \Sigma)$
- exact pivot

MS Ex.7.8

$$g(\mathbf{X}; \mu) = \frac{n(n-k)}{k(n-1)} (\hat{\mu} - \mu)^T \hat{\Sigma}^{-1} (\hat{\mu} - \mu) \leftarrow g \sim f_{k, n-k} \text{ d.f.}$$

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

- maximum likelihood estimator is approximately normal

$x_1, \dots, x_n \text{ i.i.d.}$

$f(x; \theta)$

$\theta \in \mathbb{R}^k$

$$\hat{\theta} \sim N\{\theta, I_n^{-1}(\hat{\theta})\} \implies \underbrace{(\hat{\theta} - \theta)^T I_n(\hat{\theta}) (\hat{\theta} - \theta)}_{\sim \chi_k^2}$$

$$1 - \alpha \approx \text{pr}_{\theta}\{\theta \in R(\hat{\theta})\}$$

$$R(\hat{\theta}) = \{\theta : (\hat{\theta} - \theta)^T I_n(\hat{\theta}) (\hat{\theta} - \theta) \leq \chi_{k, 1-\alpha}^2\}$$

regularity  
 $I(\theta)$  is p.d.  
~~not~~  
 (rank  $k$ )

$$I_n(\theta) = E \left( \frac{\partial^2 \ell(\theta; x)}{\partial \theta \partial \theta^T} \right) \Big|_{\theta = \theta}$$

$x_1, \dots, x_n$

$$z \sim N(0, A^{-1}) \implies z^T A z \sim \chi_{\text{rank}(A)}^2$$

- maximum likelihood estimator is approximately normal

•

$$\hat{\theta} \sim N\{\theta, I_n^{-1}(\hat{\theta})\} \implies (\hat{\theta} - \theta)^T I_n(\hat{\theta})(\hat{\theta} - \theta) \sim \chi_k^2$$

•

$$1 - \alpha \approx \text{pr}_{\theta}\{\theta \in R(\hat{\theta})\}$$

•

$$R(\hat{\theta}) = \{\theta : (\hat{\theta} - \theta)^T I_n(\hat{\theta})(\hat{\theta} - \theta) \leq \chi_{k,1-\alpha}^2\}$$

- $k = 1$ :

$$\chi_1^2 \stackrel{d}{=} N(0,1)^2$$

$$\hat{\theta} \pm z_{1-\alpha/2} \widehat{se}(\hat{\theta})$$

$$\hat{\theta} \sim N(\theta, \widehat{se}(\hat{\theta}))$$



$1-\alpha$   
CI for  $\theta$

AoS Thm 6.16

- model  $Y \sim f(y; \psi, \lambda)$ ,  $\psi \in \mathbb{R}$ ,  $\lambda \in \mathbb{R}^{d-1}$ ,  $\theta = (\psi, \lambda)$   $y = (y_1, \dots, y_n)$

- 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  $\lambda$

- maximum likelihood estimate

$$j_p^{(4)} = -\ell_p''(\psi)$$

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

$$(\hat{\psi} - \psi) j_p^{1/2}(\hat{\psi})$$

maximize over  $\lambda$

$$j_p(\psi) = -\ell_p''(\psi)$$

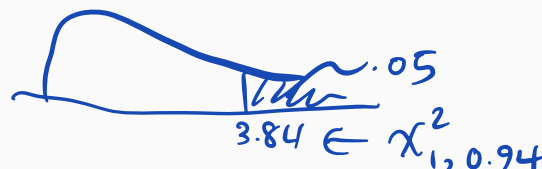
$$\hat{\psi} \pm z_{1-\alpha/2} \hat{j}_p^{1/2}$$

$se(\hat{\psi})$

- likelihood ratio test

$$2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\} \sim \chi_1^2 \Rightarrow 1 - \alpha \text{ CI} \approx \{\psi : 2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\} \geq \chi_{1,1-\alpha}^2\}$$

$\chi_1^2$



↑  
pivot 2

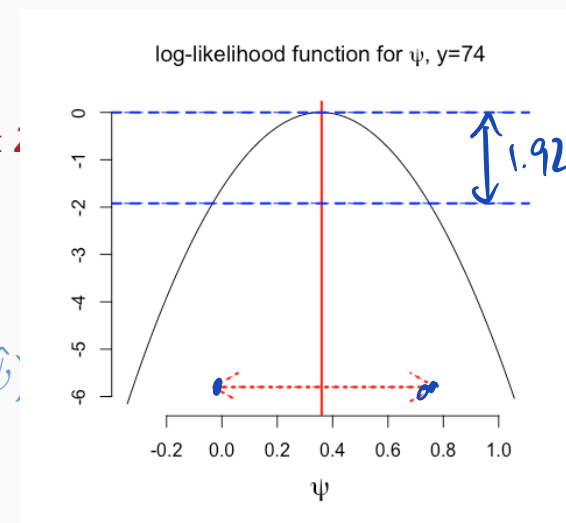
- model  $Y \sim f(y; \psi, \lambda)$ ,  $\psi \in \mathbb{R}, \lambda \in \mathbb{R}^{d-1}$ ,  $\theta = (\psi, \lambda)$   $y = (y_1, \dots, y_n)$
- log-likelihood function  $\ell(\psi, \lambda; y) = \log f(y; \psi, \lambda) = \sum \log f(y_i; \psi, \lambda)$  if independent
- profile log-likelihood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

- maximum likelihood estimate

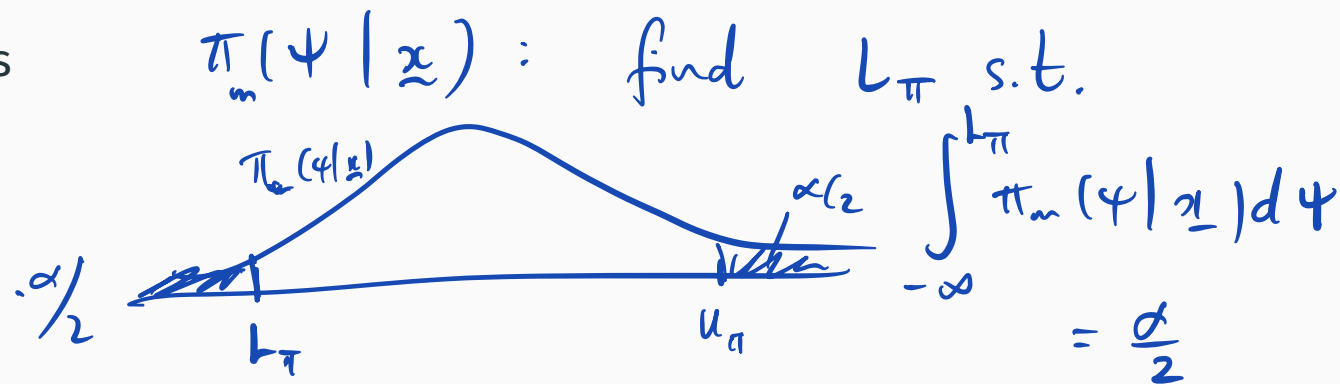
$$\hat{\psi} \sim N\{\psi, j_p^{-1/2}(\psi)\} \implies 1 - \alpha \text{ CI} \approx \hat{\psi} \pm$$

- likelihood ratio test

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



- upper and lower bounds

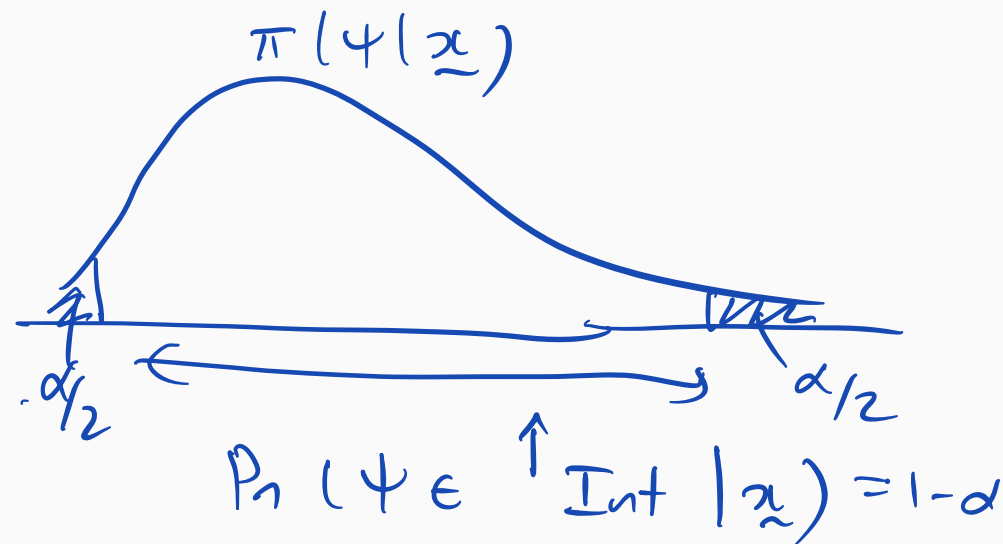
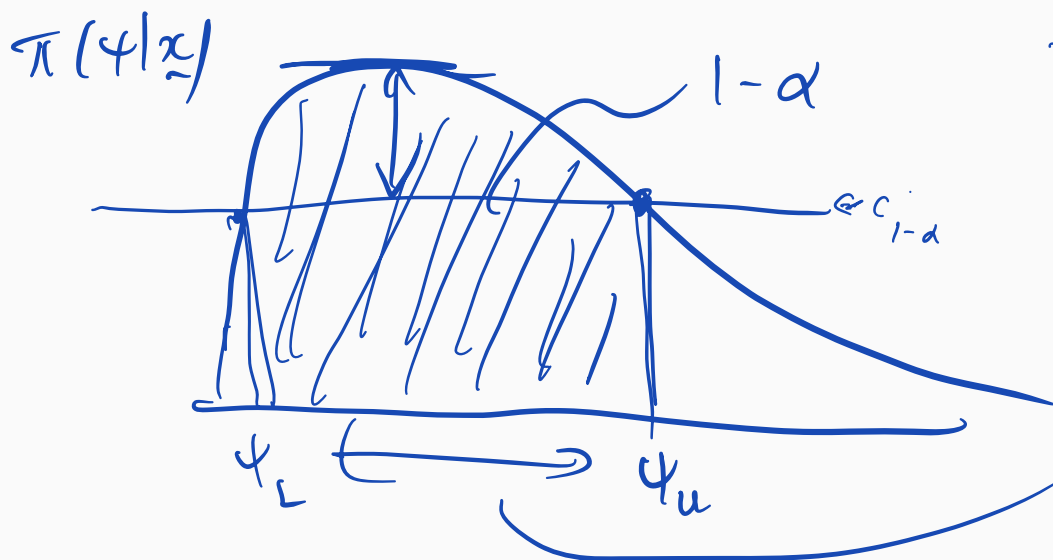


"posterior"

$$P_{\pi}(\psi \leq L_\pi(\underline{x}) | \underline{x}) = \alpha/2$$

$$P_{\pi}(\psi \geq u_\pi(\underline{x}) | \underline{x}) = \alpha/2$$

- upper and lower bounds
- equi-tailed posterior intervals



$1-\alpha$  credibility interval

highest posterior density

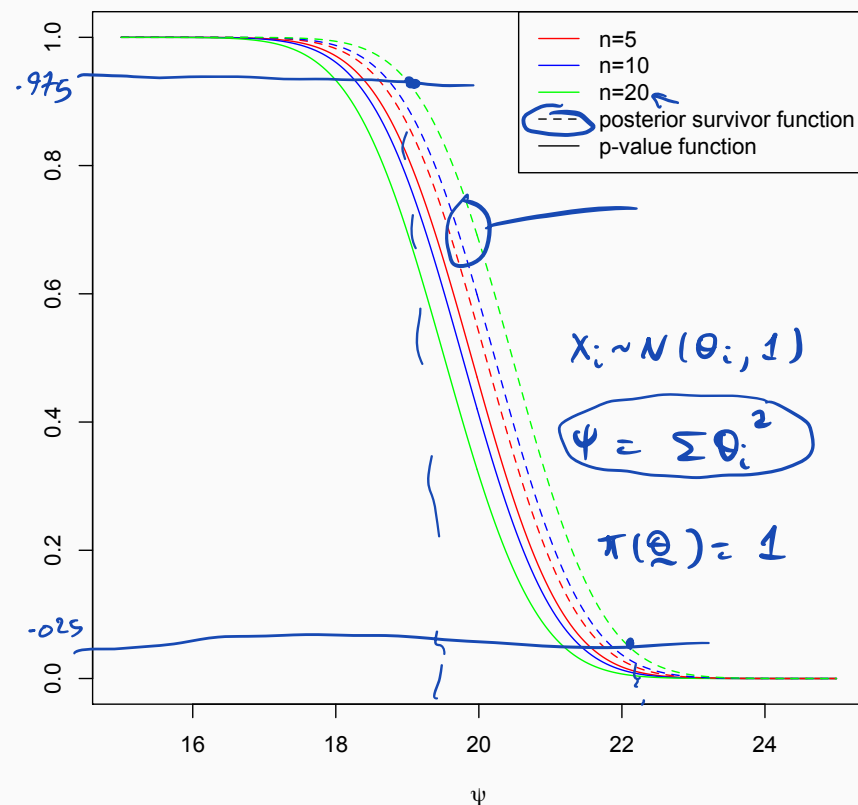
HPD

- upper and lower bounds

- ✓ • equi-tailed posterior intervals  $\{\underline{\theta} : \pi(\underline{\theta} | \underline{x}) \geq c_{1-\alpha}(\underline{x})\}$

- ✓ • highest posterior density

↑  
1-α HPD region  
for  $\underline{\theta}$



$$\frac{Pr(\Psi \geq \psi | \underline{x})}{B}$$

marginal posterior  
for  $\psi$ , given  $\underline{x}$

$$= 1 - \int_{\psi}^{\infty} \pi_{\psi}(\psi | \underline{x}) d\psi$$

post. int'l  
(19.5, 20.1)



$$\bullet X_1, \dots, X_n \sim f(\mathbf{x} \mid \theta), \quad \theta \sim \pi(\theta), \quad \pi(\theta \mid \mathbf{x}^n) = \frac{f(\mathbf{x} \mid \theta)}{f(\mathbf{x})} \quad \mathbf{x} = (x_1, \dots, x_n)$$

in nice (low-d) problems, Bayesian credible interval  
 $\approx$  freq. conf. interval  
 for large  $n$  ; fixed  $\dim(\theta)$

$$\bullet \underbrace{X_1, \dots, X_n \sim f(\mathbf{x} \mid \theta)}, \quad \underbrace{\theta \sim \pi(\theta)}, \quad \underbrace{\pi(\theta \mid \mathbf{x}^n) = \frac{f(\mathbf{x} \mid \theta)}{f(\mathbf{x})}}$$

$$\bullet \underbrace{\pi(\theta \mid \mathbf{x}) \approx N\{\hat{\theta}, j^{-1}(\hat{\theta})\}}_{\substack{\leftarrow x_1, \dots, x_n \\ \leftarrow x_1, \dots, x_n}}; \quad \pi(\theta \mid \mathbf{x}) \approx N\{\tilde{\theta}, \tilde{j}^{-1}(\tilde{\theta})\}$$

$$j(\hat{\theta}) = -l''(\hat{\theta}) \quad \hat{\theta}: l'(\hat{\theta}) = 0$$

obs. Fisher info mle

$$P_B\{\theta \in \hat{\theta} \pm z_{1-\alpha/2} j^{1/2}(\hat{\theta})\} \approx 1-\alpha$$

↑  
1-α credible interval approx

$$\mathbf{x} = (x_1, \dots, x_n)$$

$$\theta \in \mathbb{R}$$

$$\hat{\theta} \pm \overset{j^{1/2}(\hat{\theta})}{\text{se}(\hat{\theta})} \cdot z_{1-\alpha/2}$$

1-α C.I.

if

$$\hat{\theta} \sim N(\theta, j^{-1}(\hat{\theta}))$$

$$\bullet X_1, \dots, X_n \sim f(\mathbf{x} \mid \theta), \quad \theta \sim \pi(\theta), \quad \pi(\theta \mid \mathbf{x}^n) = \frac{f(\mathbf{x} \mid \theta)}{f(\mathbf{x})} \quad \mathbf{x} = (x_1, \dots, x_n)$$

$$\bullet \pi(\theta \mid \mathbf{x}) \approx N\{\hat{\theta}, j^{-1}(\hat{\theta})\}; \quad \pi(\theta \mid \mathbf{x}) \approx N\{\tilde{\theta}, \tilde{j}^{-1}(\tilde{\theta})\}$$

• careful statement

Berger, 1985; Ch.4

- For any  $a, b \in \mathbb{R}, a < b$
- let  $a_n = \tilde{\theta}_n + a j^{-1/2}(\tilde{\theta}_n)$ ,  $b_n = \tilde{\theta}_n + b j^{-1/2}(\tilde{\theta}_n)$
- $\hat{\theta}_n$  is the solution of  $\ell'(\theta; \mathbf{x}) = 0$ , assumed unique, and  $j(\theta) = -\ell''(\theta; \mathbf{x})$

Then

$$\int_{a_n}^{b_n} \pi(\theta \mid \mathbf{x}) d\theta \longrightarrow \Phi(b) - \Phi(a), \quad n \rightarrow \infty. \quad x_1 \dots x_n$$

$\int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$

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

# Approximate normality of posterior

- $X_1, \dots, X_n \sim f(\mathbf{x} \mid \theta), \quad \theta \sim \pi(\theta), \quad \pi(\theta \mid \mathbf{x}) = \frac{f(x^n \mathbf{x} \mid \theta)}{f(\mathbf{x})}$   $\mathbf{v} = (x_1, \dots, x_n)$

- $\pi(\theta \mid \mathbf{x}) \approx N\{\hat{\theta}, j^{-1}(\hat{\theta})\};$

$$\pi(\theta \mid \mathbf{x}) \approx N\{\tilde{\theta}, \tilde{j}(\tilde{\theta})\}$$

$\tilde{\theta}$  post. mode

- approximate posterior probability intervals

$$\tilde{\theta} \approx \hat{\theta}$$

$$\pi(\theta \mid \mathbf{x})$$

same  
result

$$-\tilde{\ell}''(\tilde{\theta}; \mathbf{x})$$

$$\tilde{j}(\tilde{\theta})$$

$$\tilde{\ell}'(\tilde{\theta}; \mathbf{x}) = 0$$

- exact posterior probability intervals

$$\tilde{\ell}(\theta, \mathbf{x})$$

$$= \tilde{\ell}(\tilde{\theta}; \mathbf{x}) + \dots$$

$$\pi(\theta | \underline{x}) = \frac{f(\underline{x}; \theta) \pi(\theta)}{\int f(\underline{x}; \theta) \pi(\theta) d\theta} \quad \frac{e^{\boxed{\ell(\theta; \underline{x}) + \log \pi(\theta)}}}{\int \dots d\theta}$$

$$= \frac{e^{\ell(\theta; \underline{x})} \pi(\theta)}{\int e^{\ell(\theta; \underline{x})} \pi(\theta) d\theta}$$

$$\ell(\theta; \underline{x}) = \log f(\underline{x}; \theta) \quad \ell(\theta; \underline{x}) = \log f$$

$$= \frac{e^{\cancel{\ell(\hat{\theta})} + (\hat{\theta} - \hat{\theta}) \cancel{\ell'(\hat{\theta})} + \frac{1}{2} (\theta - \hat{\theta})^2 \ell''(\hat{\theta}) + \cancel{R_n}}{\int \dots d\theta} \cdot \pi(\theta)$$

$$\pi(\theta | \underline{x}) = \frac{e^{-\frac{1}{2} (\theta - \hat{\theta})^2 \{ -\ell''(\hat{\theta}) \}} \pi(\hat{\theta})}{\int \dots \uparrow \pi(\theta) d\theta} = \frac{e^{-\frac{1}{2} (\theta - \hat{\theta})^2 j(\hat{\theta})} \cancel{\pi(\hat{\theta}) + R_n}}{\int \cancel{\pi(\hat{\theta}) + R_n} d\theta}$$

$$= N(\hat{\theta}, j^{-1}(\hat{\theta}))$$

Laplace's method

[Link](#) to Guardian

Pfizer-BioNTech vaccine trial:

vaccine: 22000 subjects, 8 cases

placebo: 22000 subjects, 162 cases

$8/162 = 5\% \implies 95\%$  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

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

December 31, 2020

N Engl J Med 2020; 383:2603-2615

DOI: 10.1056/NEJMoa2034577

Chinese Translation 中文翻译



13 References

263 Citing Articles

Letters

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).

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

| Efficacy End Point                                                                                                     | BNT162b2     |                              | Placebo      |                              | Vaccine Efficacy, %<br>(95% Credible Interval)‡ | Posterior Probability<br>(Vaccine Efficacy >30%)§ |
|------------------------------------------------------------------------------------------------------------------------|--------------|------------------------------|--------------|------------------------------|-------------------------------------------------|---------------------------------------------------|
|                                                                                                                        | No. of Cases | Surveillance Time (n)†       | No. of Cases | Surveillance Time (n)†       |                                                 |                                                   |
| Covid-19 occurrence at least 7 days after the second dose in participants without evidence of infection                | 8            | (N=18,198)<br>2.214 (17,411) | 162          | (N=18,325)<br>2.222 (17,511) | 95.0 (90.3–97.6)                                | >0.9999                                           |
| Covid-19 occurrence at least 7 days after the second dose in participants with and those without evidence of infection | 9            | (N=19,965)<br>2.332 (18,559) | 169          | (N=20,172)<br>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  $X_1 \sim \text{Bin}(18000, p_1)$
  - placebo group 18000 participants; 162 cases  $X_2 \sim \text{Bin}(18000, p_2)$
  - $0.05 = 8/162 \rightarrow 95\%$  efficacy
  - model  
interest:  $\psi = 1$  (vacc. doesn't work)  
 $\underbrace{X_1 \sim \text{Poisson}(\lambda\psi)}$ ,  $\underbrace{X_2 \sim \text{Poisson}(\lambda)}$   $\underbrace{X_1 | S = X_1 + X_2 \sim \text{Binom}(S, \psi/(1+\psi))}$   
• prior  $\text{Beta}(a, b) \rightarrow$  posterior  $\text{Beta}(\underbrace{x_1 + a}, \underbrace{s - x_1 + b})$
- $(170, \frac{\psi}{1+\psi})$   
 $S=170$   
 $162+8$

# Credible intervals

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

$$X_1 \sim \text{Poisson}(\lambda\psi), \quad X_2 \sim \text{Poisson}(\lambda)$$

$$X_1 | S = X_1 + X_2 \sim \text{Binom}(S, \psi/(1+\psi))$$

- prior  $\text{Beta}(a, b) \rightarrow$  posterior  $\text{Beta}(x_1 + a, s - x_1 + b)$

0.7 / 1

0.7      1

- `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

prior is  
washed out by  
the data

c.i. based  
on this pivot

$$VE = 1 - \psi$$

converts to  $\psi$

# Highest posterior density (HPD) regions

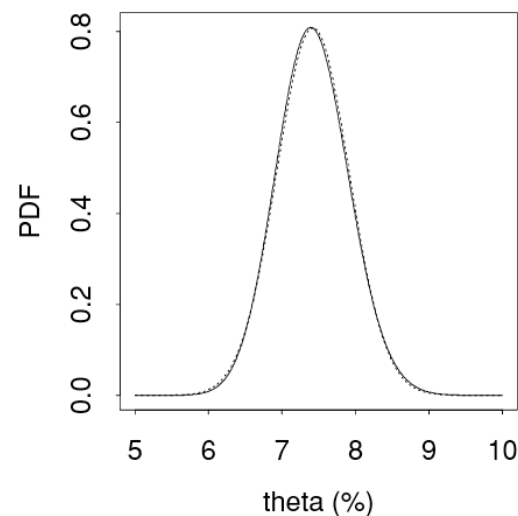
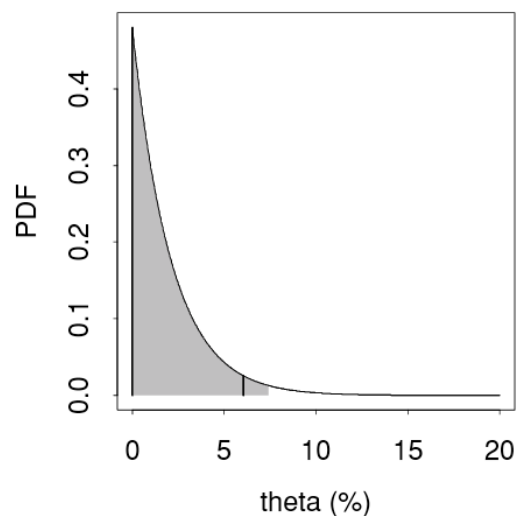
- HPD region C for  $\theta$ :

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

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

580

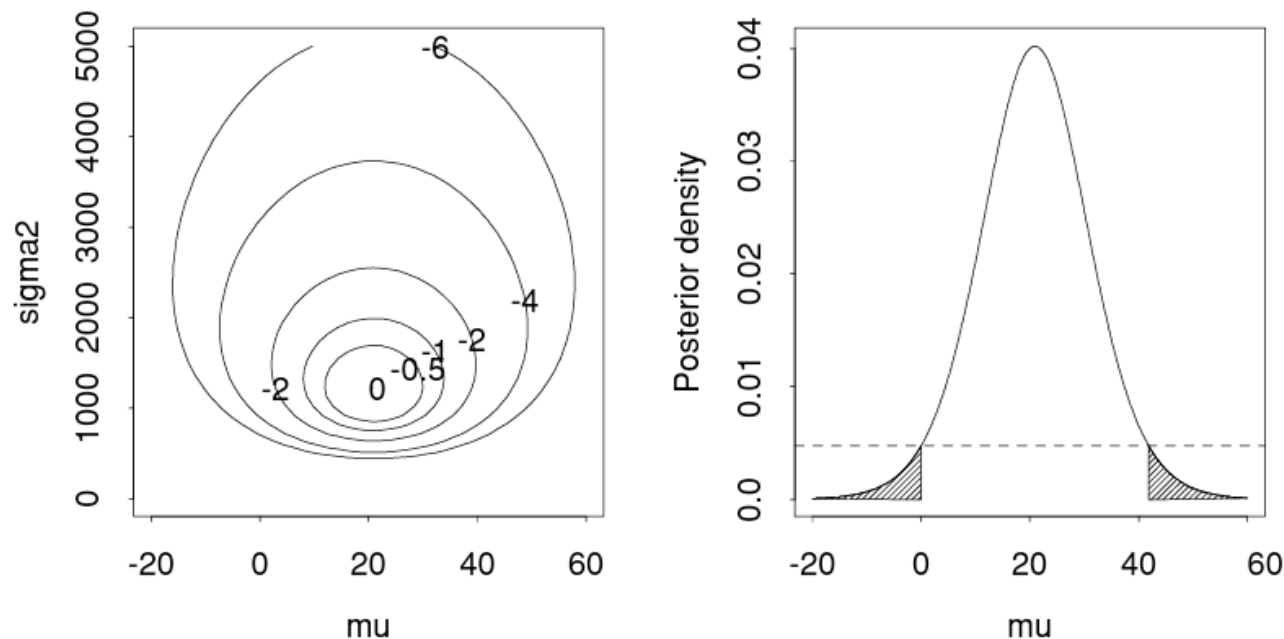
11 · Bayesian Models



**Figure 11.1** Cardiac surgery data. Left panel: posterior density for  $\theta_A$ , 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  $\theta$  (solid) and normal approximation (dots).

582

11 · Bayesian Models



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

# Likelihood ratio based approximate confidence regions

- $X_1, \dots, X_n \sim f(\mathbf{x}; \theta)$
- $L(\theta; \mathbf{x}) = f(\mathbf{x}; \theta), \quad \ell(\theta) = \log L(\theta; \mathbf{x})$

- 

$$w(\theta) = 2\{\ell(\hat{\theta}) - \ell(\theta)\} \xrightarrow{d} \chi_p^2, \quad n \rightarrow \infty$$

# Likelihood ratio based approximate confidence regions

- $X_1, \dots, X_n \sim f(\mathbf{x}; \boldsymbol{\theta})$
- $L(\boldsymbol{\theta}; \mathbf{x}) = f(\mathbf{x}; \boldsymbol{\theta}), \quad \ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}; \mathbf{x})$

- 

$$w(\boldsymbol{\theta}) = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta})\} \xrightarrow{d} \chi_p^2, \quad n \rightarrow \infty$$

- approximation:

$$w(\boldsymbol{\theta}) \sim \chi_p^2$$

- approximate confidence region

$$\{\boldsymbol{\theta} : w(\boldsymbol{\theta}) \leq \chi_{p,1-\alpha}^2\}$$

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

- empirical cdf

- properties:

empirical cdf      mle of  $F(t)$   
"mle" of  $F(\cdot)$

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

likelihood for  
∞-d. par is  
hard to define

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

any fixed  $t$

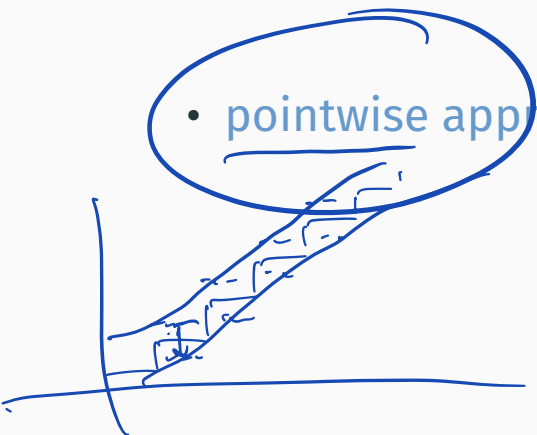
- pointwise approximate confidence limits  $\hat{F}_n(t) \pm z_{1-\alpha/2} [\hat{F}_n(t)\{1 - \hat{F}_n(t)\}]^{1/2}$

↑  $1-\alpha$  CI

↑ binomial

$$P\{F(t) \in \hat{F}_n(t) \pm \dots\} \approx 1-\alpha$$

(approx to binomial)



- 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:

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

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 :  $\text{pr}\{L(t) \leq F(t) \leq U(t) \text{ for all } t\} \geq 1 - \alpha$ :

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

AoS



probability bound on  $P_n\{\hat{F}_n(t) \leq \dots\} \leq \dots$

98 7. Estimating the CDF and Statistical Functionals

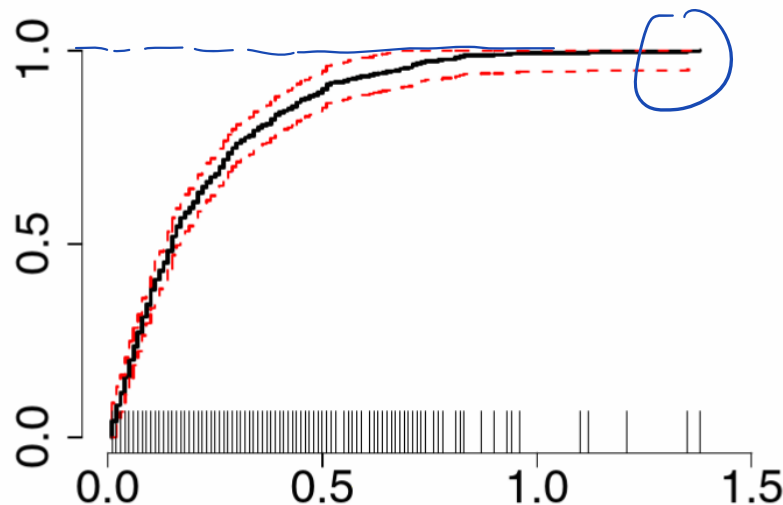


FIGURE 7.1. Nerve data. Each vertical line represents one data point. The solid line is the empirical distribution function. The lines above and below the middle line are a 95 percent confidence band.

**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  $\hat{F}_n$ . 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  $\hat{F}_n(.6) - \hat{F}_n(.4) = .93 - .84 = .09$ . ■

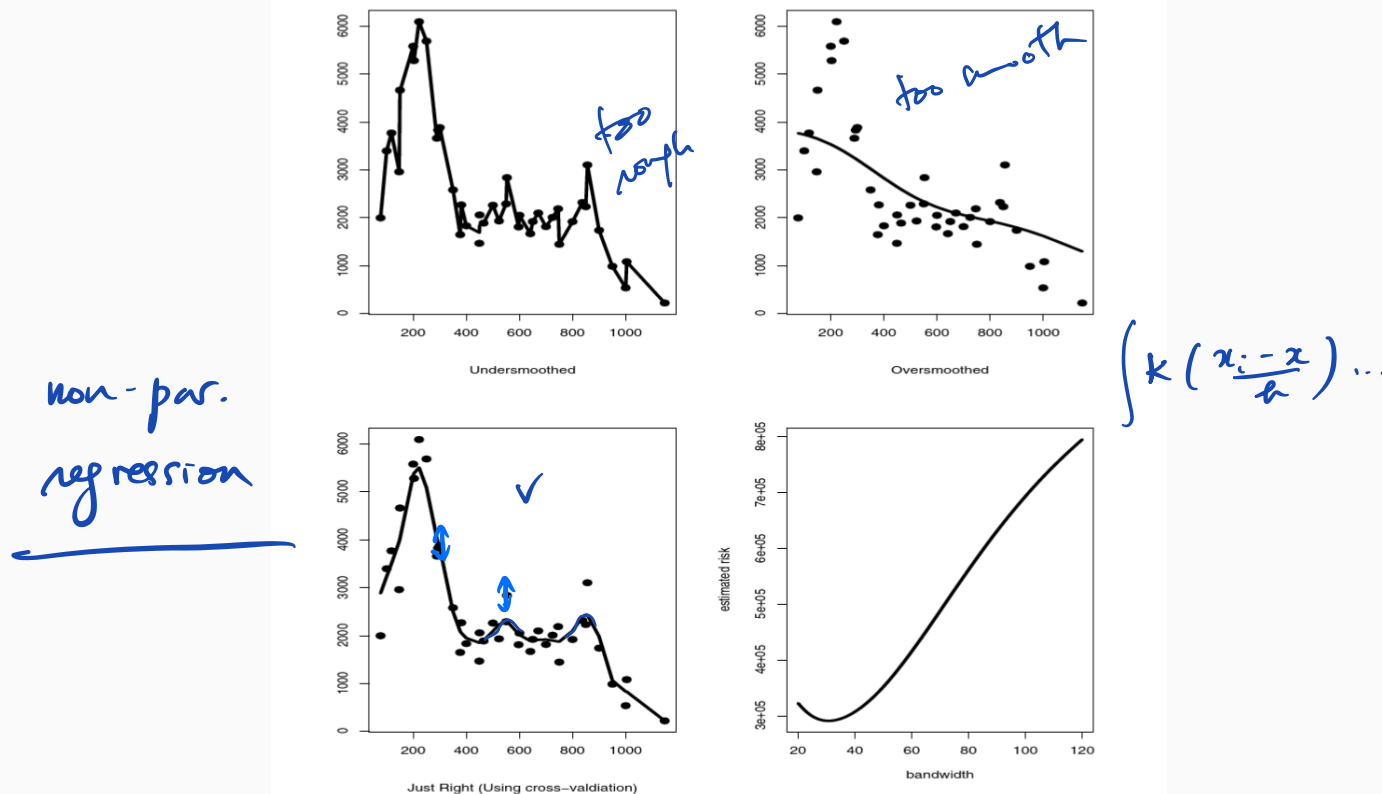


FIGURE 20.8. Regression analysis of the CMB data. The first fit is undersmoothed, the second is oversmoothed, and the third is based on cross-validation. The last panel shows the estimated risk versus the bandwidth of the smoother. The data are from BOOMERaNG, Maxima, and DASL.

## 20.4 Nonparametric Regression 323

## Confidence Bands for Kernel Regression

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

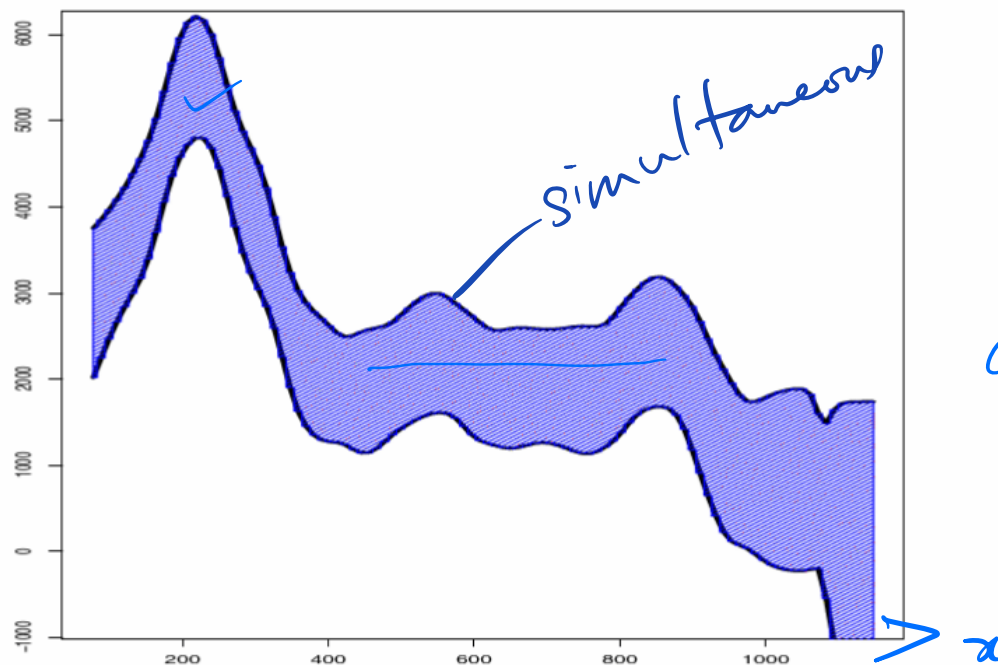
$$\ell_n(x) = \hat{r}_n(x) - q \hat{\text{se}}(x), \quad u_n(x) = \hat{r}_n(x) + q \hat{\text{se}}(x) \quad (20.37)$$

where

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

$\hat{\sigma}$  is defined in (20.36) and  $\omega$  is the width of the kernel. In case the kernel does not have finite width then we take  $\omega$  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.

324 20. Nonparametric Curve Estimation



just  $\sim$  hC

coverage  $\geq 1-\alpha$

$\forall x$  in range

$$X_1, \dots, X_n \sim f(\mathbf{x}; \theta)$$

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

$$X_1, \dots, X_n \sim f(\mathbf{x}; \theta)$$

- 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:  $\phi : \mathcal{X} \rightarrow \{0, 1\}$

$\phi(\mathbf{X}) = 1$  decide  $\theta \in \Theta_1$ , else decide  $\theta \in \Theta_0$

- Rejection region:  $R \subset \mathcal{X}$ ; if  $\mathbf{x} \in R$  “reject”  $H_0$

$$R = \{\mathbf{x} : \phi(\mathbf{x}) = 1\}$$

- Type I and Type II error:  $\Pr\{\mathbf{X} \in R \mid \theta \in \Theta_0\}, \quad \Pr\{\mathbf{X} \notin R \mid \theta \in \Theta_1\}$

- Power and Size:  $\beta(\theta) = \Pr_{\theta}(X \in R) \quad \alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$

- Optimal tests: among all level- $\alpha$  tests, find that with the highest power under  $H_1$

level- $\alpha$  means size  $\leq \alpha$

- 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

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

- subject to this constraint, minimize

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

- 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

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

- subject to this constraint, minimize

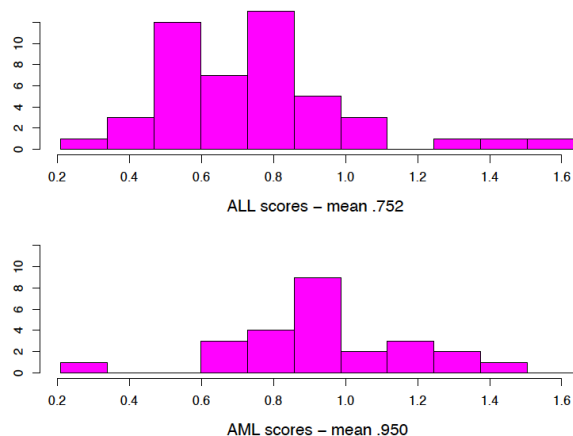
$$E_{\theta}\{\phi(\mathbf{X})\}, \quad \theta \in \Theta_1$$

- find a **test statistic**,  $T = t(\mathbf{X})$ , and  $\phi(\mathbf{X}) = 1\{T \geq t_{crit}\}$

$t_{crit}$  to be determined

## 1.2 Hypothesis Testing

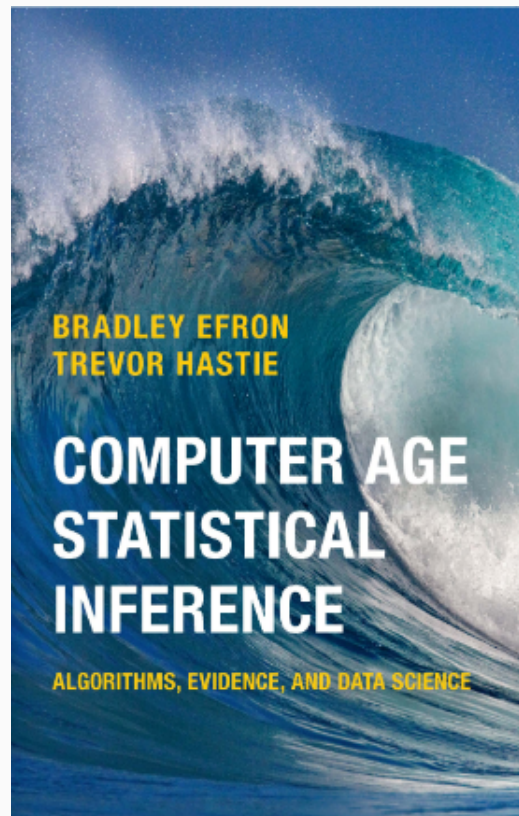
Our second example concerns the march of methodology and inference for *hypothesis testing* rather than estimation: 72 leukemia patients, 47 with **ALL** (acute lymphoblastic leukemia) and 25 with **AML** (acute myeloid leukemia, a worse prognosis) have each had genetic activity measured for a panel 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

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



```
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

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