Mathematical Statistics II

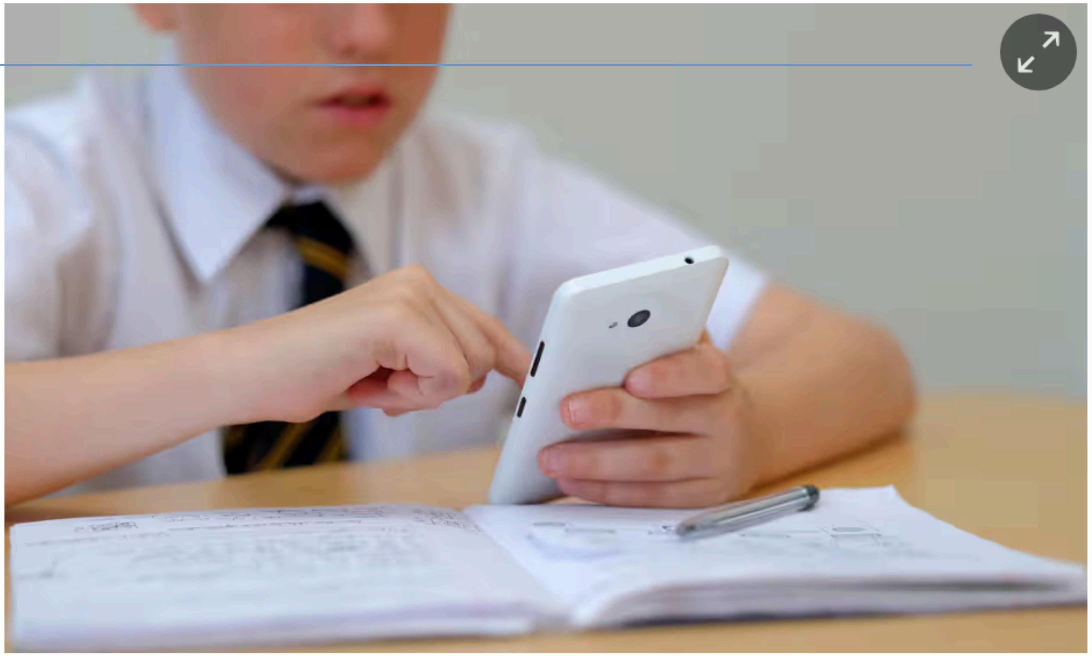
STA2212H S LEC9101

Week 6

February 11 2025

School phone bans alone do not improve grades or wellbeing, says UK study

Researchers say bans need to be part of wider strategy to tackle negative impact of mobile use on children



The negative effects of phone overuse did not differ between schools that banned phones and

School phone policies and their association with mental wellbeing, phone use, and social media use (SMART Schools): a cross-sectional observational study

Victoria A. Goodyear,[a,b,*] *Amie Randhawa,*[a,b] *Péymane Adab,*[c] *Hareth Al-Janabi,*[b,c] *Sally Fenton,*[a,d] *Kirsty Jones,*[e] *Maria Michail,*[b,f] *Breanna Morrison,*[c] *Paul Patterson,*[b,g] *Jonathan Quinlan,*[a,d] *Alice Sitch,*[c,d,h] *Rebecca Twardochleb,*[a,b] *Matthew Wade,*[h,i] and *Miranda Pallan*[c]

[a]School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Birmingham, UK
[b]Institute for Mental Health, University of Birmingham, Birmingham, UK
[c]Department of Applied Health Sciences, University of Birmingham, Birmingham, UK
[d]National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, UK
[e]Services for Education, Birmingham, UK
[f]School of Psychology, University of Birmingham, Birmingham, UK
[g]Forward Thinking Birmingham, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK
[h]ukactive, ukactive Research Institute, London, UK
[i]Advanced Wellbeing Research Centre, College of Health, Wellbeing and Life Sciences, Sheffield Hallam University, Sheffield, UK

Today

1. Recap Feb 4 decision theory, efficiency
2. Theory of interval estimation MS 7.1
3. Approximations to the posterior MS 7.2
4. Project: Choice of papers
5. HW5, Statistics in the News

Department Seminar Thursday February 13 11.00 – 12.00
Hydro Building, Room 9014
Robust nonparametrics and spatial data
Pratheepa Jeganathan, McMaster University

Recap: Decision theory : frequentist

- Loss function $L(\hat\theta, \theta)$ measures "dist" from $\hat\theta$ to θ — squared-error, absolute error, ...
- Risk function expd loss $\int L(\hat\theta(x), \theta) f(x;\theta)\,d\theta = R_\theta(\hat\theta)$ — expected loss
- Admissible point estimators — no other est. with an always better risk fc — not inadmissible — $R_\theta(\hat\theta) \geq R_\theta(\tilde\theta)$ $\forall\theta$, $>$ for some θ
- Bayes risk — average over $\pi(\theta)$ — $R_B(\hat\theta) = \int R_\theta(\hat\theta)\pi(\theta)\,d\theta$
- Bayes estimator — minimize Bayes risk — minimizes Bayes risk

$\pi(\theta|x) = f(x;\theta)\pi(\theta)/m(x)$

- Bayes risk

$$\min_{\hat\theta} \underline{R_B(\hat\theta) = \int R_\theta(\hat\theta)\pi(\theta)d\theta} = \min_{\hat\theta} \int_\Theta \int_{\mathcal{X}} L(\hat\theta(x),\theta)\, \underbrace{f(x;\theta)dx\,\pi(\theta)d\theta}$$

- Optimal Bayes estimators minimize the Bayes risk
- Equivalent to minimizing posterior loss: $\int L\{\hat\theta(x),\theta)\}\pi(\theta \mid x)d\theta$

$$= \min_{\hat\theta} \int_{\mathcal{X}} \underbrace{\int_\Theta L(\hat\theta(x),\theta)\pi(\theta|x)d\theta}\, m(x)\,dx$$

achieved at $\min_{\hat\theta}$ $\int L(\hat\theta(x),\theta)\pi(\theta|x)d\theta$

- Example: absolute-error loss

$L(\hat\theta,\theta) = |\hat\theta - \theta|$

$$\min_{\hat\theta} \int |\hat\theta - \theta|\pi(\theta \mid \boldsymbol{x})d\theta$$

- solution $\hat\theta(\boldsymbol{x}) = \text{median } \pi(\theta \mid \boldsymbol{x})$

$L(\hat\theta,\theta) = (\hat\theta - \theta)^2$

$E(\theta|\underline{x})$ post. mean.

$\boxed{\text{posterior median}}$

Bayes estimators are admissible

- Suppose $\hat{\theta}$ is a Bayes estimator and is unique

requires proper prior

- Suppose we have another estimator $\tilde{\theta}$ with a smaller frequentist risk function:

$$R_\theta(\tilde{\theta}) \leq R_\theta(\hat{\theta})$$

- The Bayes risk of $\tilde{\theta}$ is $R_B(\tilde{\theta}) = \int R_\theta(\tilde{\theta})\pi(\theta)\,d\theta$

$$\leq \int R_\theta(\hat{\theta})\pi(\theta)\,d\theta = R_B(\hat{\theta})$$

contradiction

- Suppose $\hat{\theta}$ is a Bayes estimator and is unique

- Suppose we have another estimator $\tilde{\theta}$ with a smaller frequentist risk function:

$$R_\theta(\tilde{\theta}) \leq R_\theta(\hat{\theta})$$

- The Bayes risk of $\tilde{\theta}$ is $R_B(\tilde{\theta}) = \int$

- instead of finding estimator to minimize the weighted average of the risk function we could

$$\min_{\hat{\theta}} \max_{\theta} R_\theta(\hat{\theta})$$

Definition §6.2

admissible

- such estimators are called minimax (not quite admissible)
- Bayes estimators with constant risk are minimax

$X \sim Bin(n, \theta)$ $Beta(\alpha, \beta)$ $\alpha = \beta = \sqrt{n}$

Decision theory (under sq'd error loss)

- finding the 'best' point estimator $\hat{\theta}$
- best = smallest expected loss
- no asymptotic theory involved
- can find these using a Bayesian argument
- but the justification is not Bayesian
- another non-asymptotic approach to 'best' estimators: UMVUE MS 6.3

unbiased

Interval estimation

- $X_1, \ldots, X_n$ i.i.d. $f(x;\theta)$, $\theta \in \mathbb{R}$
- a 100$(1-\alpha)$% confidence interval for θ is a random interval $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ with

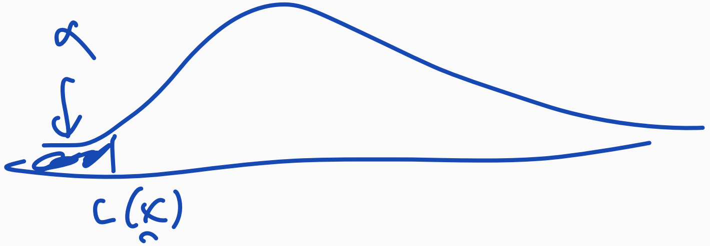
$$\Pr_\theta\{L(\underset{\sim}{X}) \le \theta \le U(\underset{\sim}{X})\} \ge 1-\alpha$$

under $f(x;\theta)$

$\underset{\sim}{X}$ continuous
can replace $\ge$ w/ $=$

- similarly, upper and lower $(1-\alpha)$-confidence bounds:

$$\underbrace{\mathrm{pr}\{\theta \ge L(\boldsymbol{X})\} = 1-\alpha}; \quad \underbrace{\mathrm{pr}\{\theta \le U(\boldsymbol{X})\} = 1-\alpha}$$



recall confidence distn

- $X_1, \ldots, X_n$ i.i.d. $f(x; \theta), \theta \in \mathbb{R}$
- a $100(1-\alpha)\%$ confidence interval for θ is a random interval $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ with

continuous

- similarly, upper and lower $(1-\alpha)$-confidence bounds:

$$\mathrm{pr}\{\theta \geq L(\boldsymbol{X})\} = 1-\alpha; \quad \mathrm{pr}\{\theta \leq U(\boldsymbol{X})\} = 1-\alpha$$

- exact limits if we have exact distribution of $\boldsymbol{X}$
- approximate limits if $\mathrm{pr}_\theta\{L(\boldsymbol{X}) \leq \theta \leq U(\boldsymbol{X})\} \approx 1-\alpha$

- Example: $X_1, \ldots, X_n$ i.i.d. $N(\mu, 1)$

pivotal quantity = $g(\underset{\sim}{x}; \theta)$ w/ property that $g(\underline{X}; \theta)$ is fixed $\forall\, \theta$

$\bar{X} \sim N(\mu, \frac{1}{n})$ $\quad (\bar{X} - \mu) \sim \underbrace{N(0, \tfrac{1}{n})}$

$\underbrace{\sqrt{n}(\bar{x} - \mu)} \sim \underbrace{N(0,1)}$

$\Pr\{-1.96 \leq Z \leq 1.96\} = 0.95 \quad Z \sim N(0,1)$

$= \Pr\{-1.96 \leq \sqrt{n}(\bar{X} - \mu) \leq 1.96\}$

$= \Pr\{\bar{X} - \frac{1.96}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{1.96}{\sqrt{n}}\}$

Exact confidence intervals

- Example: $X_1, \ldots, X_n$ i.i.d. $N(\mu, 1)$
- Example $X_1, \ldots, X_n$ i.i.d. $U(0, \theta)$

$L(\underline{X})$ $\quad$ $U(\underline{X})$

$\hat{\theta} = X_{(n)}$ $\quad$ $n(X_{(n)} - \theta) \xrightarrow{d} \exp(1)$

$\frac{X_{(n)}}{\theta}$ has known dist^{n} $\quad$ $\mathrm{pr}(a \leq \frac{X_{(n)}}{\theta} \leq b)$

$f(x_i) = \frac{1}{\theta} \quad 0 \leq x_i \leq \theta$ $\qquad$ $F(x_i) = \frac{x_i}{\theta}, \quad 0 \leq x_i \leq \theta$

$F_{X_{(n)}}(x) = \Pr(X_{(n)} \leq x) = \prod_{i=1}^{n} \left(\frac{x}{\theta}\right) = \left(\frac{x}{\theta}\right)^n$

$Z = \frac{X_{(n)}}{\theta}$

$F_Z(z) = z^n \quad 0 \leq z \leq 1$

$\Pr\left(a \leq \frac{X_{(n)}}{\theta} \leq b\right) = b^n - a^n$

$= P(bX_{(n)} \leq \theta \leq aX_{(n)})$

choose a, b s.t. $b^n - a^n = 1 - \alpha$

$\underset{a,b}{\text{minimize}}\ (b-a)\,X_{(n)}$ s.t.)

"It can be shown that
$b=1$, $a=(1-\alpha)^{1/n}$ "

- Example: $X_1, \ldots, X_n$ i.i.d. $N(\mu, 1)$
- Example $X_1, \ldots, X_n$ i.i.d. $U(0, \theta)$ $\mathrm{pr}(a \leq \frac{X_{(n)}}{\theta} \leq b)$
- Example $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$

$$\Pr\left\{\frac{(\bar{X}-\mu)}{S/\sqrt{n}} \le t\right\} = \Pr\{T_{n-1} \le t\}$$

student's t-distr.

$ES^2 = \sigma^2$ $\quad S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2 \quad \bar{X} = \frac{1}{n}\sum X_i$

$$(*) \longrightarrow \Pr\left\{-t_{n-1}^{\alpha/2} \le \frac{\sqrt{n}(\bar{X}-\mu)}{S} \le t_{n-1}^{\alpha/2}\right\} = 1-\alpha$$

by def.

$$= \Pr\left(\bar{X} - t_{n-1}^{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{n-1}^{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

95% value > 2

- Example: $X \sim Binom(n, \theta)$, $\hat{\theta} \dot{\sim} N(\theta, \theta(1-\theta)/n)$ MS Ex.7.6

$$\mathrm{pr}_\theta\left[-1.96 \le \frac{\sqrt{n}(\hat{\theta}-\theta)}{\{\theta(1-\theta)\}^{1/2}} \le 1.96\right] \approx 0.95$$

approximate pivot

$I_n^{-1}(\theta)$

$$\hat{\theta} - \theta \dot{\sim} N\left(0, \frac{\theta(1-\theta)}{n}\right)$$

$I_n^{-1}(\theta)$ as approx variance of $\hat{\theta}$

$$\frac{\sqrt{n}(\hat{\theta}-\theta)}{\sqrt{\theta(1-\theta)}} \dot{\sim} N(0,1)$$

solve for θ means solve a quadratic eq$^{=}$

$$\frac{n(\hat{\theta}-\theta)^2}{\theta(1-\theta)} \le 1.96$$

$$n(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2) \le 1.96\,\theta - 1.96\,\theta^2$$

- Example: $X \sim Binom(n, \theta)$, $\hat\theta \dot\sim N(\theta, \theta(1-\theta)/n)$ MS Ex.7.6

$$\mathrm{pr}_\theta\left[-1.96 \le \frac{\sqrt{n}(\hat\theta - \theta)}{\{\theta(1-\theta)\}^{1/2}} \le 1.96\right] \approx 0.95$$

$$\mathrm{pr}_\theta\left[-1.96 \le \frac{\sqrt{n}(\hat\theta - \theta)}{\underbrace{\{\hat\theta(1-\hat\theta)\}}_{I_n(\hat\theta)}{}^{1/2}} \le 1.96\right] \approx 0.95$$

pivot

dividing by $\hat{se}$ is called studentizing

$$\hat\theta \pm 1.96 \frac{\hat\theta(1-\hat\theta)}{\sqrt{n}}$$

Approximate confidence intervals

- Example: $X \sim Binom(n, \theta)$, $\hat{\theta} \dot{\sim} N(\theta, \theta(1-\theta)/n)$ MS Ex.7.6

$$\mathrm{pr}_\theta\left[-1.96 \le \frac{\sqrt{n}(\hat{\theta}-\theta)}{\{\theta(1-\theta)\}^{1/2}} \le 1.96\right] \approx 0.95$$

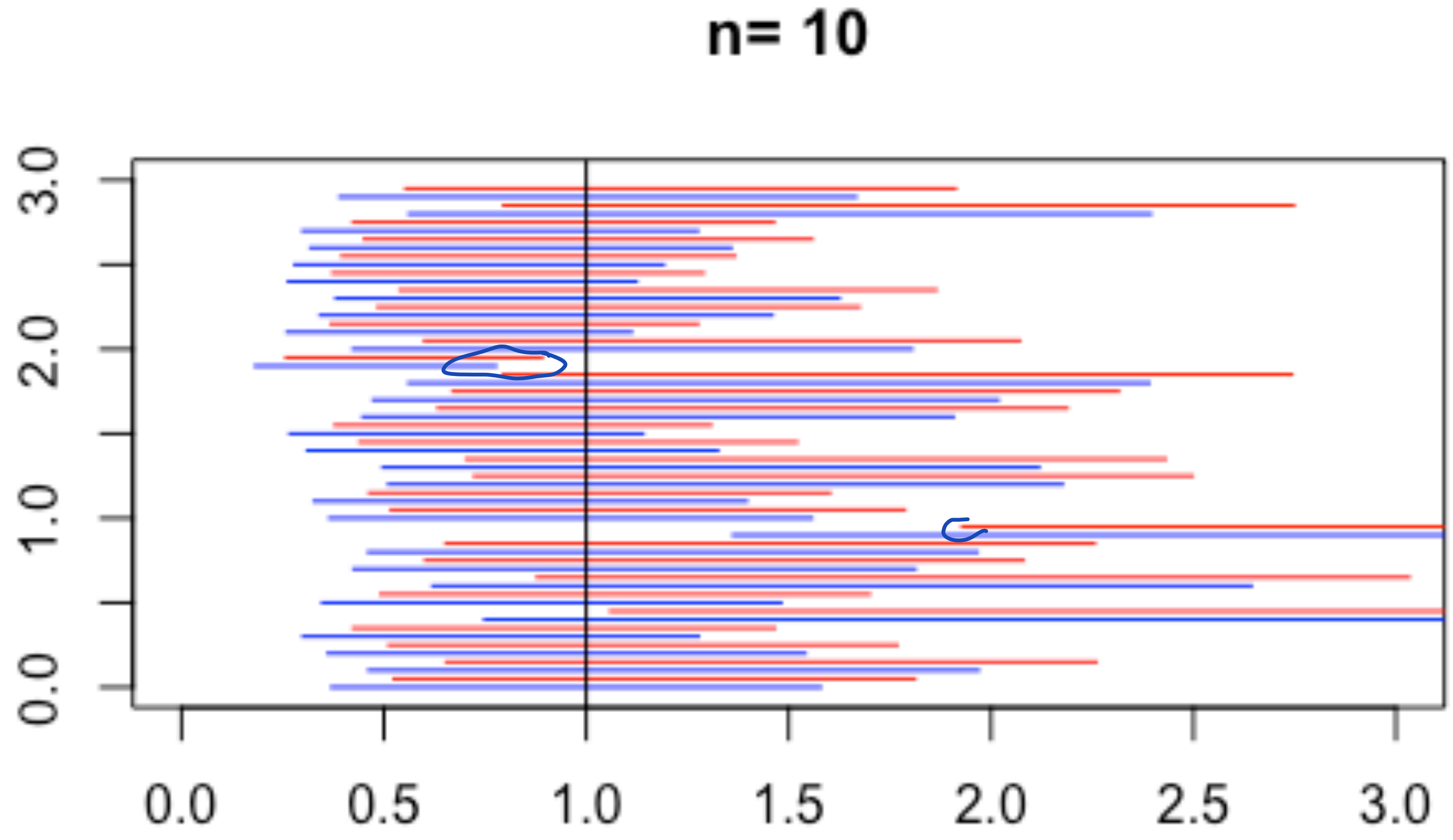
$$\mathrm{pr}_\theta\left[-1.96 \le \frac{\sqrt{n}(\hat{\theta}-\theta)}{\{\hat{\theta}(1-\hat{\theta})\}^{1/2}} \le 1.96\right] \approx 0.95$$

- $\hat{\theta}_n$ maximum likelihood estimate $\hat{\theta} \dot{\sim} N[\theta, \{nI(\theta)\}^{-1}]$
- approximate 95% confidence interval AoS Thm 6.16

$\hat{\theta} \pm z_{\alpha/2}\,\hat{se}$

- $X_1, \ldots, X_n$ i.i.d. $Exp(\lambda)$ $\quad f(x_i; \lambda) = \lambda e^{-\lambda x_i}; \quad x_i > 0, \ \lambda > 0$
- $\hat{\lambda} = \quad \frac{1}{\bar{X}}$

$\phi =$
- $g(\lambda) = \log \lambda$ $\quad \cancel{g(\lambda)} \quad \hat{\phi} = g(\hat{\lambda}) = -\log \bar{X}$


n= 10
0.0
1.0
2.0
3.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0

- upper and lower bounds $\theta \in \mathbb{R}$

- upper and lower bounds

$\theta \in \mathbb{R}$

- equi-tailed posterior intervals

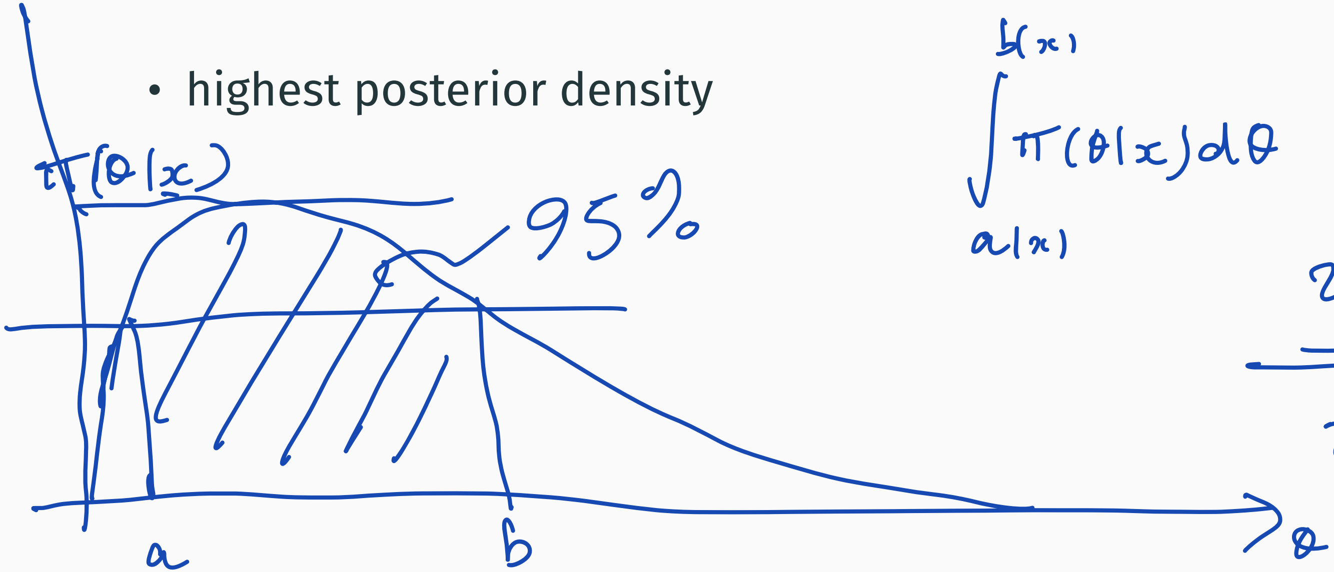
- upper and lower bounds $\theta \in \mathbb{R}$
- equi-tailed posterior intervals
- highest posterior density $\theta \in \mathbb{R}^p, p \geq 1$

… Bayesian credible intervals

- upper and lower bounds — $\theta \in \mathbb{R}$
- equi-tailed posterior intervals
- highest posterior density — $\theta \in \mathbb{R}^p, p \geq 1$

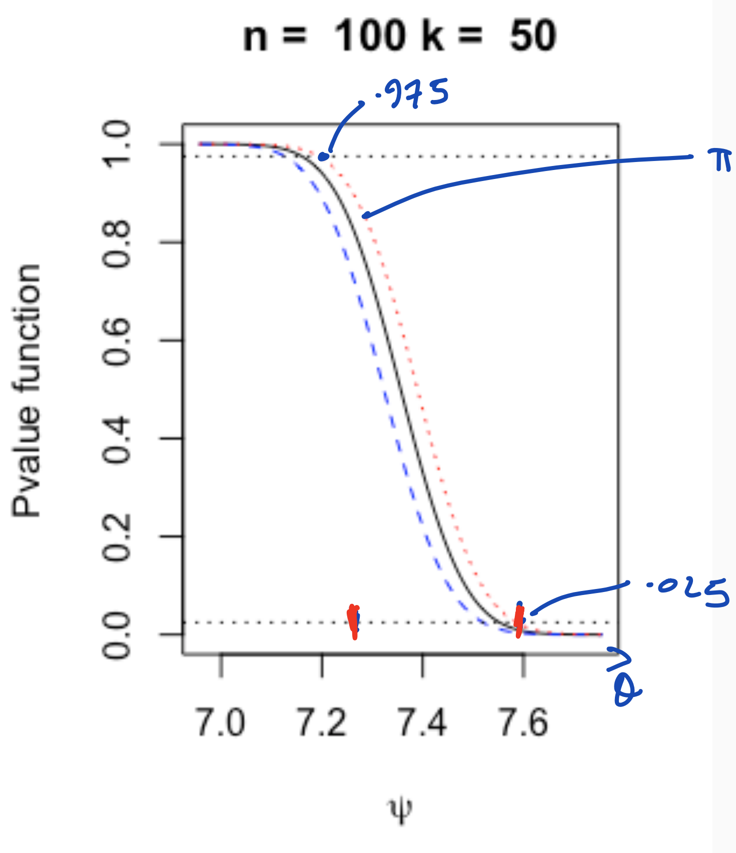
$\pi(\theta | \underset{\sim}{x})$ gives all our probabilities

$$\int_{a(x)}^{b(x)} \pi(\theta | x)\, d\theta = 1 - \alpha$$



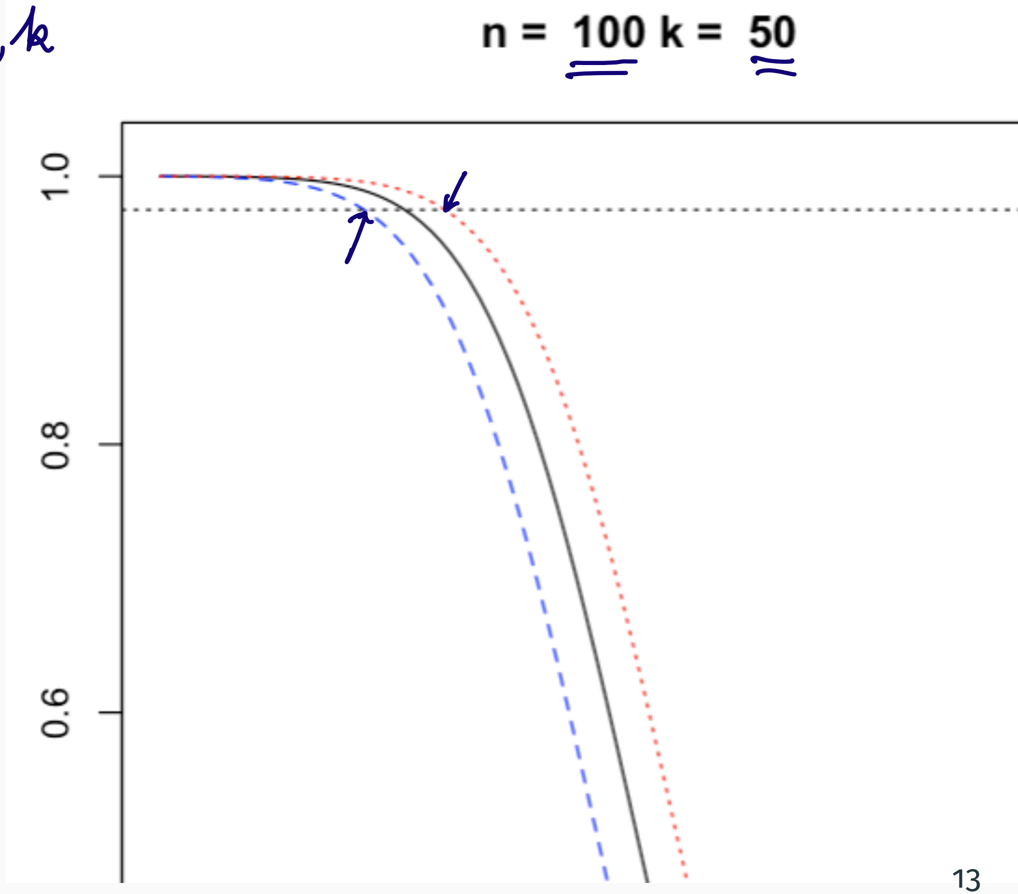
$X_i \sim N(\theta_i, \frac{1}{n})$ $\quad \psi = \sum_{i=1}^{k} \theta_i^2$

$i = 1, \dots, k$



$\pi(\theta \geq \theta \mid \underset{\sim}{x})$

$1 - \text{cdf}$



Approximate normality of posterior

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

Approximate normality of posterior

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

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

$\tilde\theta = \arg\max_\theta \pi(\theta \mid x)$

$\tilde{j}(\tilde\theta) = -\dfrac{\partial^2}{\partial\theta^2} \log \pi(\theta \mid x) \Big|_{\theta=\tilde\theta}$

prior washed out by data

for large n

$e^{-\frac{j(\hat\theta)}{2}(\theta-\hat\theta)^2} \dfrac{1}{\sqrt{2\pi}} (j(\hat\theta))^{1/2}$

Approximate normality of posterior

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

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

- careful statement $\qquad \underbrace{\hat{\theta} + a\,\hat{se}}_{a_n} \le \theta \le \underbrace{\hat{\theta}_n + b\,\hat{se}}_{b_n}$ $\qquad$ Berger, 1985; Ch.4
  - For any $a, b \in \mathbb{R}, a < b$
  - let $a_n = \hat{\theta}_n + aj^{-1/2}(\hat{\theta}_n)$, $b_n = \hat{\theta}_n + bj^{-1/2}(\hat{\theta}_n)$
  - $\hat{\theta}_n$ is the solution of $\ell'(\theta; x^n) = 0$, assumed unique, and $j(\theta) = -\ell''(\theta; x^n)$

  Then

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

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

normal cdf $\qquad$ usual cond's

Approximate normality of posterior

on $f(x;\theta)$

- $X_1, \ldots, X_n \sim f(x^n \mid \theta), \qquad \theta \sim \pi(\theta), \qquad \pi(\theta \mid x^n) = \dfrac{f(x^n \mid \theta)}{f(x^n)}$ $\qquad x^n = (x_1, \ldots, x_n)$
- $\boxed{\pi(\theta \mid x^n) \approx N\{\hat\theta, j^{-1}(\hat\theta)\};}$ $\qquad \underline{\underline{\pi(\theta \mid x^n)} \approx N\{\tilde\theta, \tilde{j}(\tilde\theta)\}}$ used N approx to Bin.
- approximate posterior probability intervals

$\hat\theta \pm j^{1/2}(\hat\theta)\, 1.96$ — freq approx 95% CI

$\hat\theta \pm j^{1/2}(\hat\theta)\, 1.96$ — approx Bayes credible interval

$$\int_{a_n}^{b_n} \pi(\theta|\underset{\sim}{x})\, d\theta \doteq \int_{a_n}^{b_n} e^{\ell(\theta;\underset{\sim}{x})}\pi(\theta)\, d\theta \Big/ \iint_{x}\int_{a_n}^{b_n} e^{\ell(\theta;\underset{\sim}{x})}\pi(\theta)\, d\theta\, d\underset{\sim}{x}$$

i.e.

$$\underline{\underline{\propto}} \int_{a_n}^{b_n} e^{\ell(\theta;\underset{\sim}{x})}\pi(\theta)\, d\theta = \int_{a_n}^{b_n} e^{\frac{\ell(\theta;\underset{\sim}{x}) + \log\pi(\theta)}{g(\theta)}}\, d\theta$$

v.2 $\quad g'(\tilde{\theta}) = 0$

$$= \int_{a_n}^{b_n} e^{\ell(\hat{\theta};\underset{\sim}{x}) + \cancel{(\theta-\hat{\theta})\ell'(\hat{\theta};\underset{\sim}{x})} + \frac{1}{2}(\theta-\hat{\theta})^2\ell''(\hat{\theta}) + R_n}\pi(\theta)\, d\theta$$

$$\doteq e^{\ell(\hat{\theta};\underset{\sim}{x})}\int_{a_n}^{b_n} e^{-\frac{1}{2}(\theta-\hat{\theta})^2(-\ell''(\hat{\theta}))}\{\pi(\hat{\theta}) + (\theta-\hat{\theta})\pi'(\hat{\theta}) + r\}\, d\theta$$

$$\doteq \pi(\hat{\theta})e^{\ell(\hat{\theta};x)} \cdot \int_{a_n}^{b_n} e^{-\frac{1}{2}(\theta-\hat{\theta})^2 J(\hat{\theta})}\left\{1 + (\theta-\hat{\theta})\frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})}\right\} d\theta$$

$$\int_{a_n}^{b_n}\pi(\theta|\underset{\sim}{x})\, d\theta = \frac{\sqrt{2\pi}}{|J(\hat{\theta})|^{1/2}}\pi(\hat{\theta})e^{\ell(\hat{\theta};x)} \cdot \int_{a_n}^{b_n}\frac{|J(\hat{\theta})|^{1/2}}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta-\hat{\theta})^2 J(\hat{\theta})}\{1 + \dots\}\, d\theta$$

$$z = (\theta-\hat{\theta})J^{1/2}(\hat{\theta})$$

$$\theta = a_n = \hat{\theta} + a\hat{J}^{-1/2}$$

$$z = a$$

$$= \text{stuff} \times \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz\ \{1 + \text{other stuff}\}$$

constant in θ

??

must be to this order of approx constant $\doteq \theta$

ntbc

$\exists$ a better version of this argument $\longrightarrow$ Laplace approx $\doteq$

$$\hat{\pi}(\theta|\underset{\sim}{x}) \doteq \frac{1}{\sqrt{2\pi}}|j(\hat{\theta})|^{1/2} \cdot e^{\ell(\theta)-\ell(\hat{\theta})}\pi(\hat{\theta})$$

$\{1 + O(n^{-1})\}$

Example: vaccine efficacy

$\pi(\theta|\underline{x}) = \tilde{\pi}(\theta|x)\{1+O(n^{-1})\}$

Guardian, Jan 24 2021

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

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

  $X_1 \sim Poisson(\lambda\psi), \quad X_2 \sim Poisson(\lambda) \qquad X_1 \mid S = X_1 + X_2 \sim Binom(S, \psi/(1+\psi))$

  170 ill
- prior $Beta(a, b) \longrightarrow$ posterior $Beta(x_1 + a, s - x_1 + b)$

  8 + 0.7 162 + 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`

credible interval

$VE = 1 - \psi$

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

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

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. $N_p(\boldsymbol{\mu},\Sigma)$ $\theta \in \mathbb{R}^p$ MS Ex.7.8
- exact pivot

$$g(\boldsymbol{X};\boldsymbol{\mu}) = \underbrace{\frac{n(n-p)}{p(n-1)}(\underset{\bar{X}}{\widehat{\boldsymbol{\mu}}}-\boldsymbol{\mu})^T\hat{\Sigma}^{-1}(\underset{\bar{X}-\mu}{\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}})}_{\sim F_{p,n-p}\text{ dist}^{=}}$$

"Student's T"

Hotelling's T^2

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

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

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

conf. region

exact

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

$F_{p,n-p}$

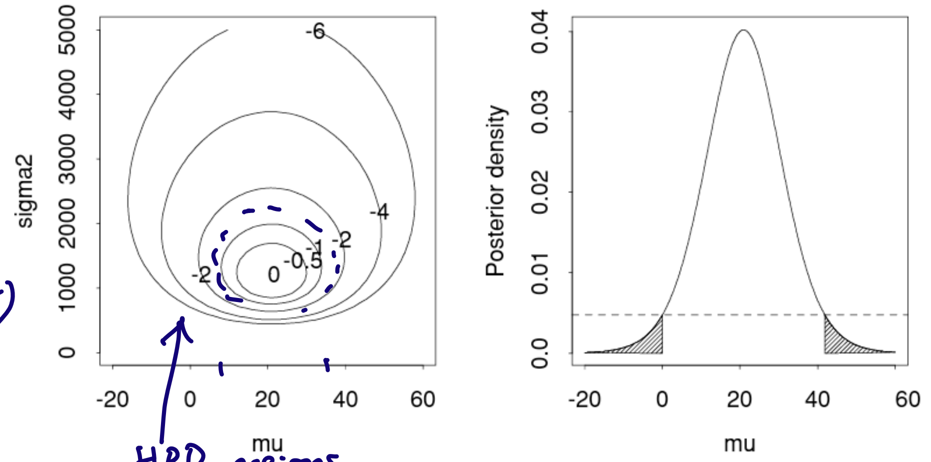
- HPD region C for θ:

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

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

582   *11 · Bayesian Models*

$\theta = (\mu, \sigma^2)$



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.

Approximate confidence regions

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

$\underline{\theta} \in \Theta = \mathbb{R}^p$

Approximate confidence regions

- 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

Likelihood ratio based approximate confidence regions

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

$$\Pr\{w(\theta) \le \chi^2_p(1-\alpha)\} \doteq 1-\alpha$$

freq. v. of HPD interval

defines conf. region for $\underline{\theta} \in \mathbb{R}^p$

Likelihood ratio based approximate confidence regions

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

Profile likelihood intervals

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

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

- likelihood ratio test

$$2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\} \dot{\sim} \chi^2_1 \implies 1 - \alpha \text{ CI } \approx \{\psi : 2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\} \leq \chi^2_{1,1-\alpha}\}$$

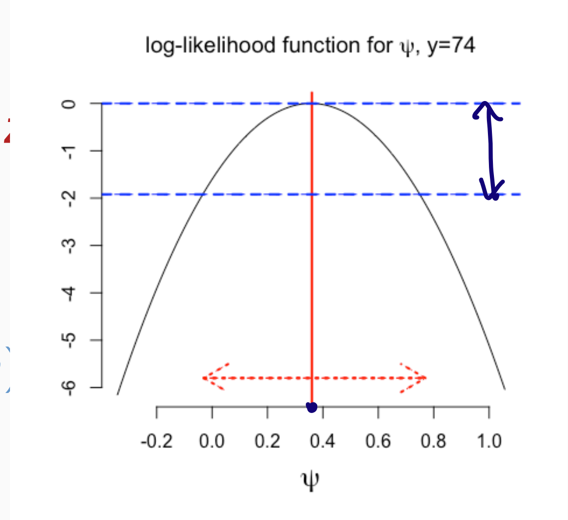
Profile likelihood intervals

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

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

- likelihood ratio test

$$2\{\ell_p(\hat\psi) - \ell_p(\psi)\} \mathrel{\dot\sim} \chi^2_1 \implies 1-\alpha \text{ CI } \approx \{\psi : 2\{\ell_p(\hat\psi)$$



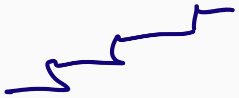
unk.

- recall $X_1, \ldots, X_n, i.i.d. F(\cdot)$ ← nonp.

- empirical cdf

est.

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



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

binomial

approx. pivot

- 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:

Bonferroni

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

any fixed t

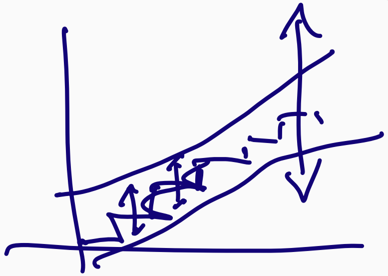
- pointwise approximate confidence limits $\hat{F}_n(t) \pm z_{1-\alpha/2}[\hat{F}_n(t)\{1 - \hat{F}_n(t)\}]^{1/2}$
- simultaneous confidence band : $\mathrm{pr}\{L_n(t) \leq F(t) \leq U_n(t) \text{ for all } t\} \geq 1 - \alpha$:

$\simeq$ $=$

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

best L, U

95% conf.



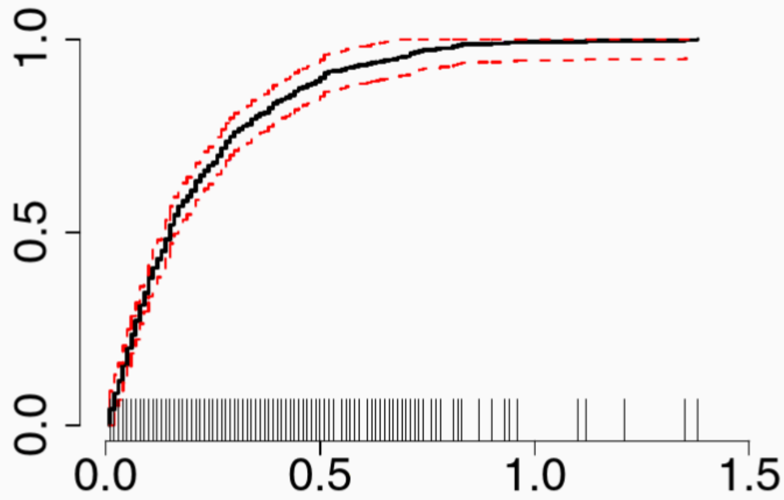
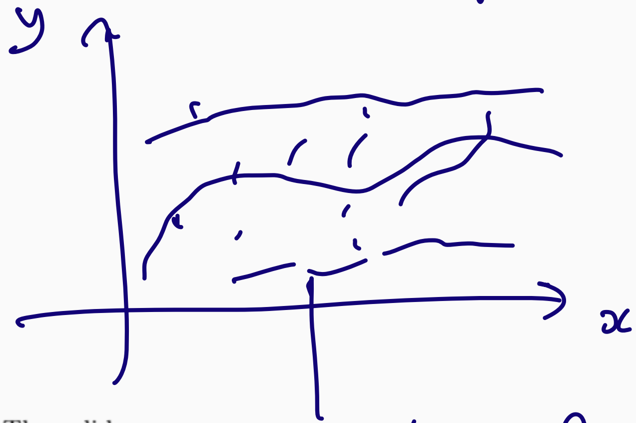


FIGURE 7.1. Nerve data. Each vertical line represents one data point. The solid line is the empirical distribution function. The lines above and below the middle line are a 95 percent confidence band.

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

$y_i = \underbrace{m(x_i)}_{\text{smooth fn}} + \varepsilon_i$



sim. band for $m(\cdot)$

Example: Bootstrap confidence intervals

Calculate in R

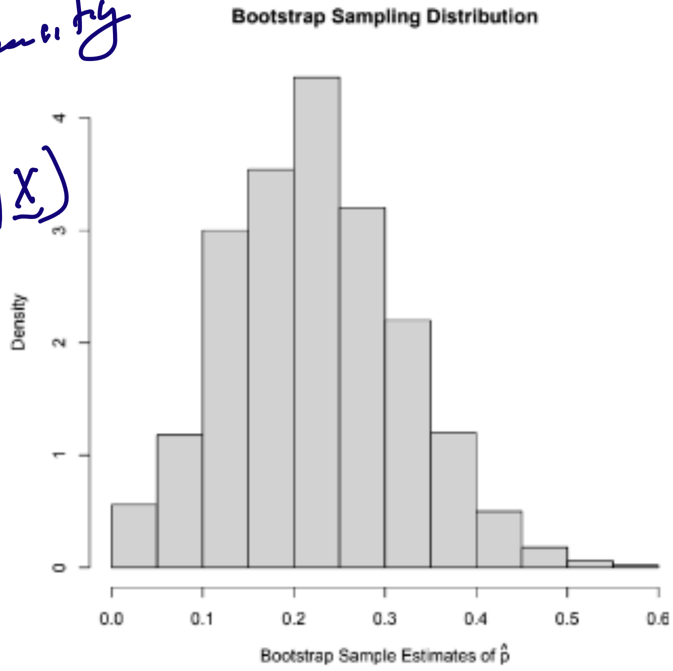
```
> alpha = 0.05
>
> # Normal-based CI
> c(phat - qnorm(1-alpha/2)*sd(bs_est),
+   phat + qnorm(1-alpha/2)*sd(bs_est))
[1] 0.05709686 0.44290314
>
> # Percentile CI
> c(quantile(bs_est, alpha/2),
+   quantile(bs_est, 1-alpha/2))
 2.5% 97.5%
 0.10  0.45
```

$\Rightarrow n = 20 = \begin{cases} 5 \text{ successes} \\ 15 \text{ failures} \end{cases}$

Example: Bootstrap confidence intervals

B.S. density

$p^*(\hat{\psi}^* \mid \underset{\sim}{X})$



$\underset{\sim}{X}^*$ iid $f(x; \hat{\underset{\sim}{\theta}})$

or

$\hat{F}_n(\cdot)$

$\psi = g(\theta)$