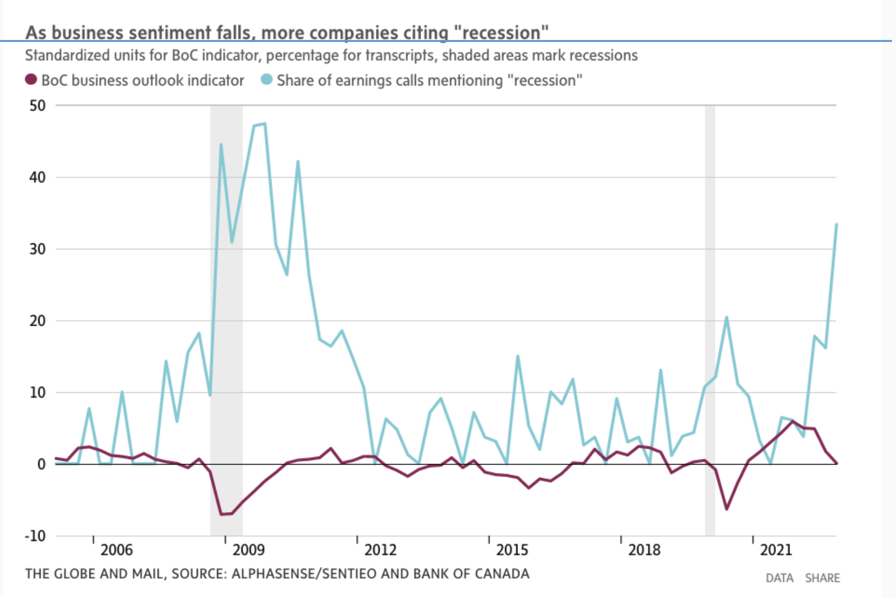
Mathematical Statistics II

STA2212H S LEC0101

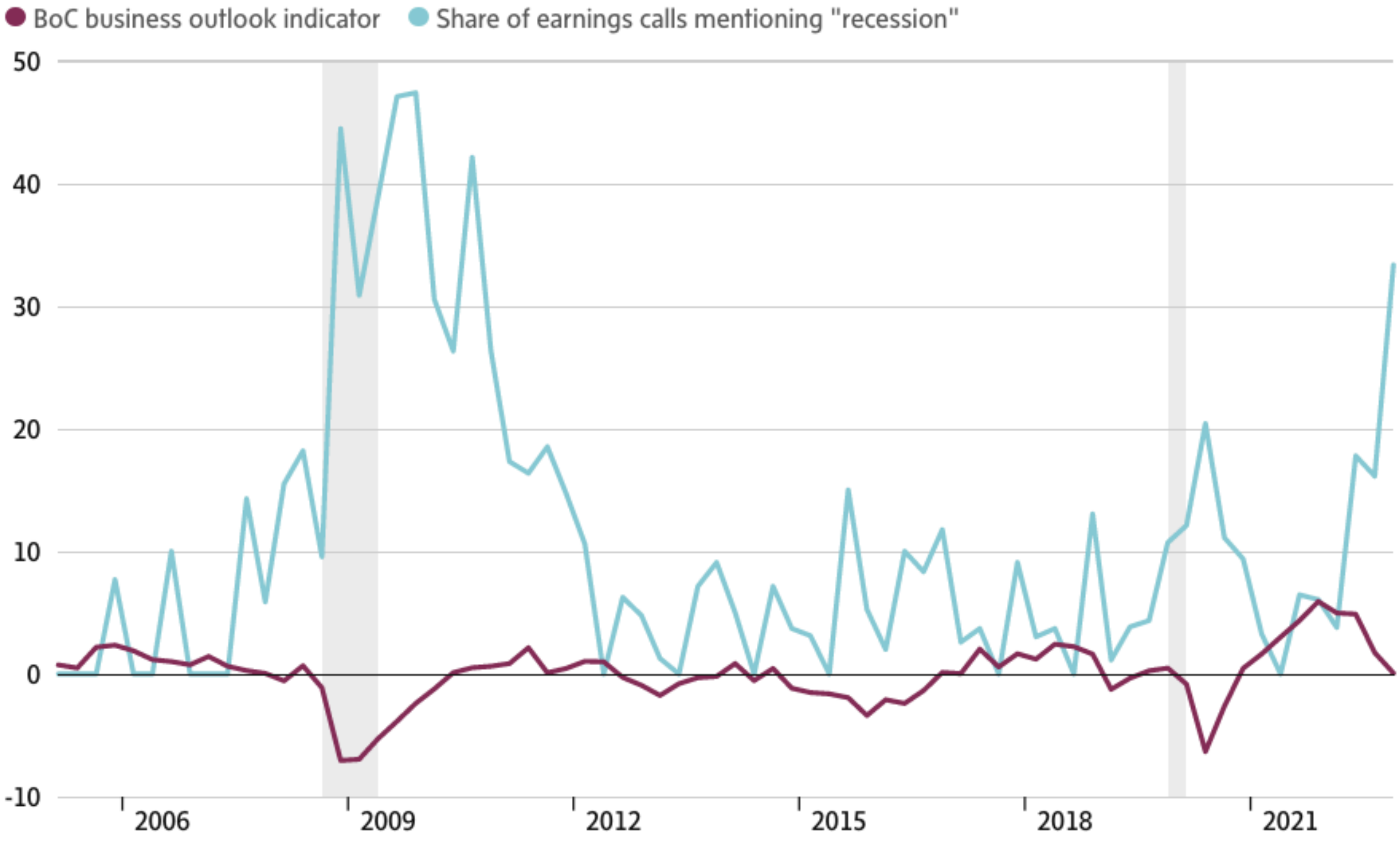
Week 4

January 31 2023



As business sentiment falls, more companies citing "recession"

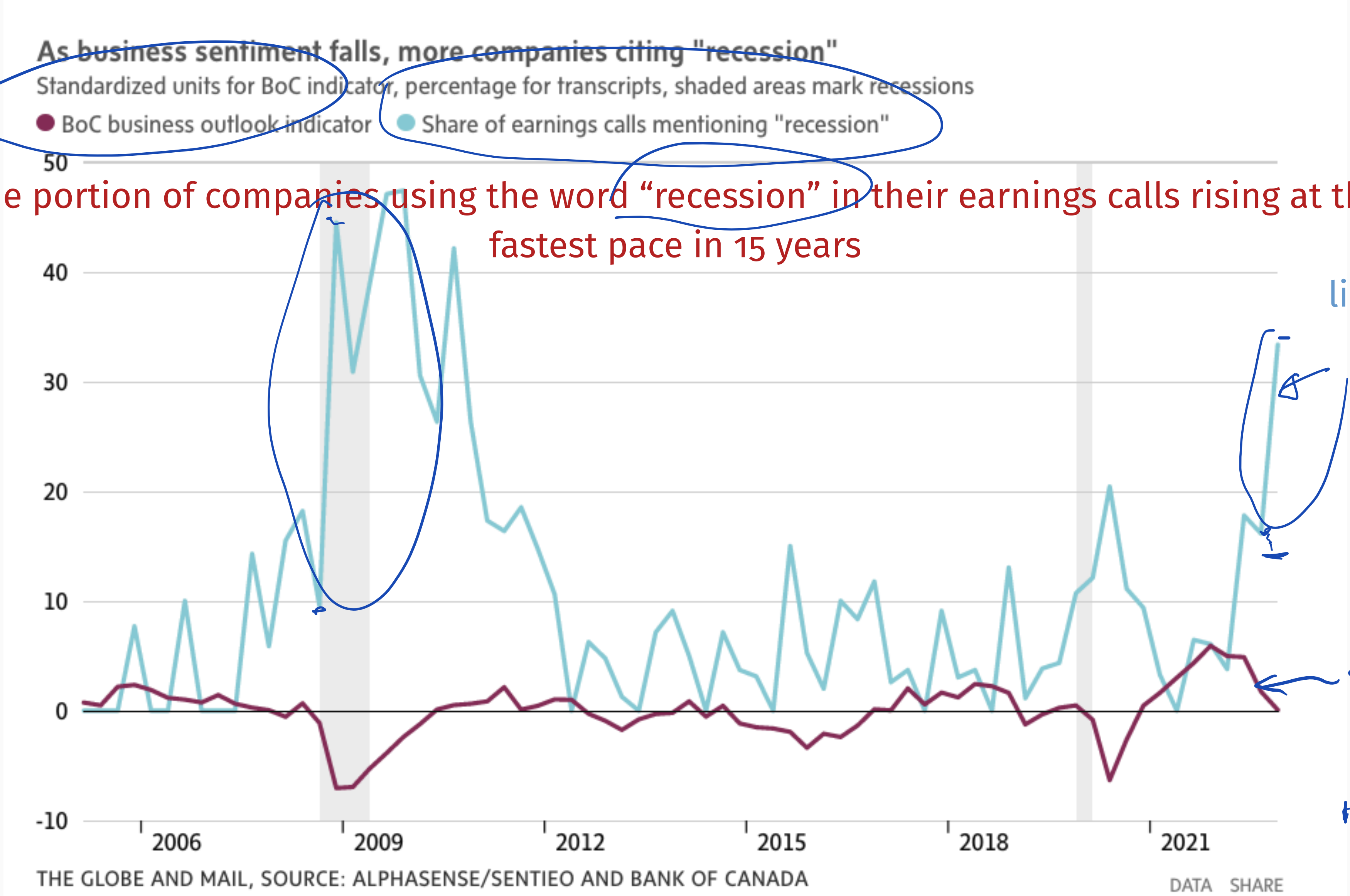
Standardized units for BoC indicator, percentage for transcripts, shaded areas mark recessions



THE GLOBE AND MAIL, SOURCE: ALPHASENSE/SENTIEO AND BANK OF CANADA

DATA SHARE

the portion of companies using the word “recession” in their earnings calls rising at the fastest pace in 15 years



link

survey
St. Canada

1. Recap ← la rev. D. T.
2. Bayesian hierarchical modelling SM 11.4, AoS 24.5
3. Multi-parameter posteriors AoS 11.7, SM, 11.1-3
4. Interval estimation MS 7.1,2
5. H3: project

Recap

- priors for Bayesian inference: conjugate, Jeffreys', flat, convenience, weakly informative, hierarchical, matching

non B

- optimality in estimation: Cramer-Rao lower bound: $\mathrm{Var}_\theta\{S(\boldsymbol{X})\} \geq g'(\theta)^2/\{nI(\theta)\}$
- matrix version, equality

$$\mathrm{var}_\theta\, \underset{\sim}{S}(\underset{\sim}{X}) \geq \left(\frac{\partial g}{\partial \theta}\right)\{nI(\theta)\}^{-1}\left(\frac{\partial g}{\partial \theta}\right)^T$$

$k\times p \qquad p\times p \qquad p\times k$

if $E_\theta S(\underset{\sim}{x})$

$\parallel \; = g(\theta)$

$= \int \underset{\sim}{S}(\underset{\sim}{x}) f(\underset{\sim}{x};\theta)\,d\underset{\sim}{x}$

$$E_\theta \underset{\sim}{S}(\underset{\sim}{x}) = \underset{\sim}{g}(\underset{\sim}{\theta})$$

$k\times 1 \qquad k\times 1$

$A \geq B \iff A - B$ p.s.d.

In MS, $k \leq p$

A is p.s.d. if $a^T A a \geq 0 \quad \forall$ vectors $a \in \mathbb{R}^k$

$$\mathrm{var}_\theta S(\underset{\sim}{X}) \geq \frac{g'(\theta)^2}{nI(\theta)}$$

$$\mathrm{cov}_\theta(S, U_\theta)^2 \leq \mathrm{var}_\theta S \, \mathrm{var}_\theta U_\theta$$

$$U_\theta(\underset{\sim}{X}) = \left\{\frac{\partial \ell}{\partial\theta}(\theta; \underset{\sim}{X})\right\}$$

$$CS = \iff \mathrm{corr}^2(S, U_\theta) = 1$$

$$E_\theta\{U_\theta(\underset{\sim}{X}) S(\underset{\sim}{X})\}^2 = \mathrm{var}_\theta S \, \underbrace{\mathrm{var}_\theta U_\theta}_{nI(\theta)}$$

$$CS: \mathrm{cov}^2(X, Y) \leq \mathrm{var}X \cdot$$

$$= \iff Y = aX + b$$

$$\mathrm{var}_\theta(S) = \frac{E_\theta\{\ \}^2}{nI(\theta)}$$

$$S(\underset{\sim}{X}) = a(\theta) U_\theta(\underset{\sim}{X}) + b(\theta)$$

$$\frac{\partial \log f(\underset{\sim}{x}; \theta)}{\partial\theta}$$

$$\int U_\theta(\underset{\sim}{x}) S(\underset{\sim}{x}) \, d\underset{\sim}{x}$$

$$U(\underset{\sim}{X}) = A(\theta) S(\underset{\sim}{X}) + B(\theta)$$

$$\frac{\partial \log f(\underset{\sim}{x}; \theta)}{\partial\theta} = A(\theta) S(X) + B(\theta)$$

$$\longrightarrow \log f(\underset{\sim}{x}; \theta) = A^*(\theta) S(X) + B^*(\theta) + C(\underset{\sim}{X})$$

$$f(\underset{\sim}{x}; \theta) = e^{A^*(\theta) S(\underset{\sim}{X}) + B^*(\theta)} h(\underset{\sim}{X})$$

Essentially, CRLB only achieved for ~~the~~ estimate $g(\theta) = E_\theta S(\underset{\sim}{X})$ where $S(\underset{\sim}{X})$ is suff't st. in an exponential family

Recap

- priors for Bayesian inference: conjugate, Jeffreys', flat, convenience, weakly informative, hierarchical, matching

- optimality in estimation: Cramer-Rao lower bound: $\mathrm{Var}\{S(\boldsymbol{X})\} \geq g'(\theta)^2/\{nI(\theta)\}$
- matrix version, equality ✓

- maximum likelihood estimators are "BAN" asymptotic relative efficiencey

- so are other regular estimators with continuous (in θ) variance functions

- Minimum Variance Unbiased Estimators (MVUE) discussed in MS 6.3
- the theory is elegant, but the application is limited

Recap 2

- finite sample optimality: loss function, risk function, admissible estimator
- Bayes estimators are admissible

  proper prior; loss function
- Bayes estimators minimize Bayes risk

$$L_p : \{\hat{\theta}(\underset{\sim}{x}) - \theta\}^p$$

$$L(\hat{\theta}(\underset{\sim}{x}), \theta) = \{\hat{\theta}(\underset{\sim}{x}) - \theta\}^2$$

$$= |\hat{\theta}(\underset{\sim}{x}) - \theta| \quad \text{mean absolute error}$$

$$= \begin{cases} 0 & \hat{\theta}(\underset{\sim}{x}) = \theta \\ 1 & \hat{\theta}(\underset{\sim}{x}) \neq \theta \end{cases}$$

$$\text{Kullback-Leibler} = E_\theta \log\left\{\frac{f(\underset{\sim}{x};\theta)}{f(\underset{\sim}{x};\hat{\theta}(\underset{\sim}{x}))}\right\} = \int \log\left(\frac{f(x;\theta)}{f(x;\hat{\theta}(x))}\right) f(x;\theta)\,dx$$

(min KL Loss $\rightarrow$ MLE)

$$R_\theta(\hat{\theta}(\underline{X})) = E_\theta\, L(\hat{\theta}(X), \theta)$$

$$R(\hat{\theta}; \theta) = \int L(\hat{\theta}(\underline{x}), \theta)\, f(\underline{x}; \theta)\, d\underline{x}$$

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

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

or

$$\boxed{\min_{\hat{\theta}(\underline{X})} \max_{\theta} R_\theta(\hat{\theta})}$$

minimax

- the Bayes risk of an estimator is the average of the risk function, over a prior distribution

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

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

E_θ

- Example: squared-error loss $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$ need to minimize over $\hat{\theta}$
$$\int (\hat{\theta} - \theta)^2 \pi(\theta \mid \boldsymbol{x})d\theta$$

- solution $\hat{\theta}(\boldsymbol{x}) = \mathrm{E}(\theta \mid \boldsymbol{x})$ posterior mean

$= \int \theta \pi(\theta|x) d\theta$ $\quad \min_a \int (\theta - a)^2 \pi(\theta|x) d\theta$

achieved ~~a = E...~~

Bayes estimators are admissible

- Suppose $\hat{\theta}$ is a Bayes estimator   proper prior $\pi(\theta)$ and is unique
- Suppose we have another estimator $\tilde{\theta}$ with a smaller frequentist risk function:

$$\underline{R_\theta(\tilde{\theta}, \theta)} \leq R_\theta(\hat{\theta}_B, \theta)$$

$\tilde{\theta}$ compete

$\hat{\theta}_B$ Bayes est.

- The Bayes risk of $\tilde{\theta}$ is

$$R_B(\tilde{\theta}) = \int_\Theta R_\theta(\tilde{\theta}, \theta)\,\pi(\theta)\,d\theta$$

$$= \int_\Theta \int_{\mathcal{X}} L(\tilde{\theta}(\underline{x}), \theta) f(\underline{x}; \theta)\, d\underline{x}\, \pi(\theta)\, d\theta$$

$$= \int_\Theta \int_{\mathcal{X}} L(\tilde{\theta}(\underline{x}), \theta)\, \pi(\theta \mid \underline{x}) \cdot f(\underline{x})\, d\underline{x}\, \pi(\theta)\, d\theta$$

$$= \int_{\mathcal{X}} \int_\Theta L(\tilde{\theta}(x), \theta)\, \pi(\theta \mid x)\, d\theta \cdot f(x)\, d\underline{x}$$

$$\pi(\theta \mid x) = \frac{f(x; \theta)\pi(\theta)}{\,}$$

$$f(x) \equiv \int \uparrow d\theta$$

mini'd by Bayes est.

$\leq \int \int_{\Theta} L(\hat{\theta}(x), \theta) \pi(\theta | x) d\theta \cdot f(x) dx$

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

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

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

- The Bayes risk of $\tilde{\theta}$ is

$$R_B(\tilde{\theta}) = \int$$

- instead of minimizing the average (over $\pi(\theta)$) of the risk function we could

$$\min \max R_\theta(\hat{\theta})$$

Definition §6.2

- such estimators are called minimax

Marginalization

$\theta_B(\underset{\sim}{x})$ minimizes $E(L(\hat\theta,\theta) \mid \underset{\sim}{x}) = \int L(\theta,\hat\theta)\pi(\theta|x)d\theta$

- Bayes posterior carries all the information about θ, given $\boldsymbol{x}$ — by definition
- probabilities for any set A computed using the posterior distribution
- post. $\mathrm{pr}(\boldsymbol{\Theta} \in A \mid \boldsymbol{x}) = \int_A \pi(\underset{\sim}{\theta}|\underset{\sim}{x})d\underset{\sim}{\theta}$ $\qquad \pi(\psi,\underset{\sim}{\lambda})$
- if $\boldsymbol{\theta} = (\psi, \boldsymbol{\lambda})$, … $\qquad \pi(\psi|\underset{\sim}{x}) = \int_{\Lambda} \pi(\psi,\underset{\sim}{\lambda}|\underset{\sim}{x})d\underset{\sim}{\lambda}$ $\qquad$ marg'l posterior
- or, if $\psi = \psi(\boldsymbol{\theta})$ $\qquad \pi(\psi|\underset{\sim}{x}) = \int_{\{\theta:\psi(\theta)=\psi\}} \pi(\underset{\sim}{\theta}|\underset{\sim}{x})d\underset{\sim}{\theta}$
- in this context, 'flat' priors can have a large influence on the marginal posterior

Multi-parameter models

- parameter $\theta = (\theta_1, \ldots, \theta_p)$
- model $f(x^n \mid \theta), \quad x^n = (x_1, \ldots, x_n)$
- joint posterior

$$\pi(\theta \mid x^n) \propto f(x^n \mid \theta)\pi(\theta), \quad \theta \in \mathbb{R}^p$$

- parameter $\theta = (\theta_1, \ldots, \theta_p)$
- model $f(\underset{\sim}{x}^n \mid \theta)$, $\quad \underset{\sim}{x}^n = (x_1, \ldots, x_n)$
- joint posterior

$$\pi(\theta \mid \underset{\sim}{x}^n) \propto f(\underset{\sim}{x}^n \mid \theta)\pi(\theta), \quad \theta \in \mathbb{R}^p$$

- marginal posterior — for θ_1

$$\pi_m(\theta_1 \mid \underset{\sim}{x}^n) = \int \pi(\theta \mid \underset{\sim}{x}^n) d\theta_2 \ldots d\theta_p$$

- marginal posterior — for $\psi(\theta)$

$$\pi_m(\psi \mid x^n) = \int_{\{\theta:\psi(\theta)=\psi\}} \pi(\theta \mid x^n) d\theta$$

$$\hat{\theta} \,\dot{\sim}\, N\left(\theta, \{nI(\theta)\}^{-1}\right)$$

$$\hat{se}(\hat{\theta}) \,\dot{\sim}\, \{nI(\hat{\theta})\}^{-1/2}$$

95% CI $= \hat{\theta} \pm 1.96 \cdot \hat{se}(\hat{\theta})$

approx $= \hat{\theta} \pm 1.96 \sqrt{n}\{I(\hat{\theta})\}^{-1/2}$

Bayesian inference: Multi-parameter models

- model: $x_i \sim N(\mu_i, 1), i = 1, \ldots, n$
- prior: $\pi(\mu)d\mu \propto d\mu$

$\pi(\mu_i)d\mu_i = d\mu_i$ $\quad \pi(\mu_i) = 1$ $\quad \mu_i$'s ind't

- posterior $\pi(\mu \mid x^n) \propto \prod_{i=1}^n \pi(\mu_i \mid x_i) = \prod_{i=1}^n \phi(x_i, 1/\cancel{n})$

$$\pi(\underset{\sim}{\mu} \mid \underset{\sim}{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\mu_i - x_i)^2}$$

$$\hat{\underset{\sim}{\mu}}_B = \underset{\sim}{x} \quad \text{under squared error loss}$$

$$\pi(\mu_1 \mid \underset{\sim}{x}) = N(x_1, 1) \quad \ldots$$

- model: $x_i \sim N(\mu_i, 1), i = 1, \ldots, n$

find me an est. of $\Sigma \mu_i^2$

$\Sigma x_i^2 \sim \chi_n(\Sigma \mu_i^2)$

$E(\downarrow) = n + \Sigma \mu_i^2$

$\Sigma x_i^2 - n$ unbiased for ψ

- prior: $\pi(\mu)d\mu \propto d\mu$
- posterior $\pi(\mu \mid x^n) \propto \prod_{i=1}^n \pi(\mu_i \mid x_i) = \prod_{i=1}^n \phi(x_i, 1/n)$
- $\psi = \sum_{i=1}^n \mu_i^2$

squared length of mean vector

$$\pi(\psi \mid x^n) = \int_A \pi(\mu \mid x^n) d\mu$$

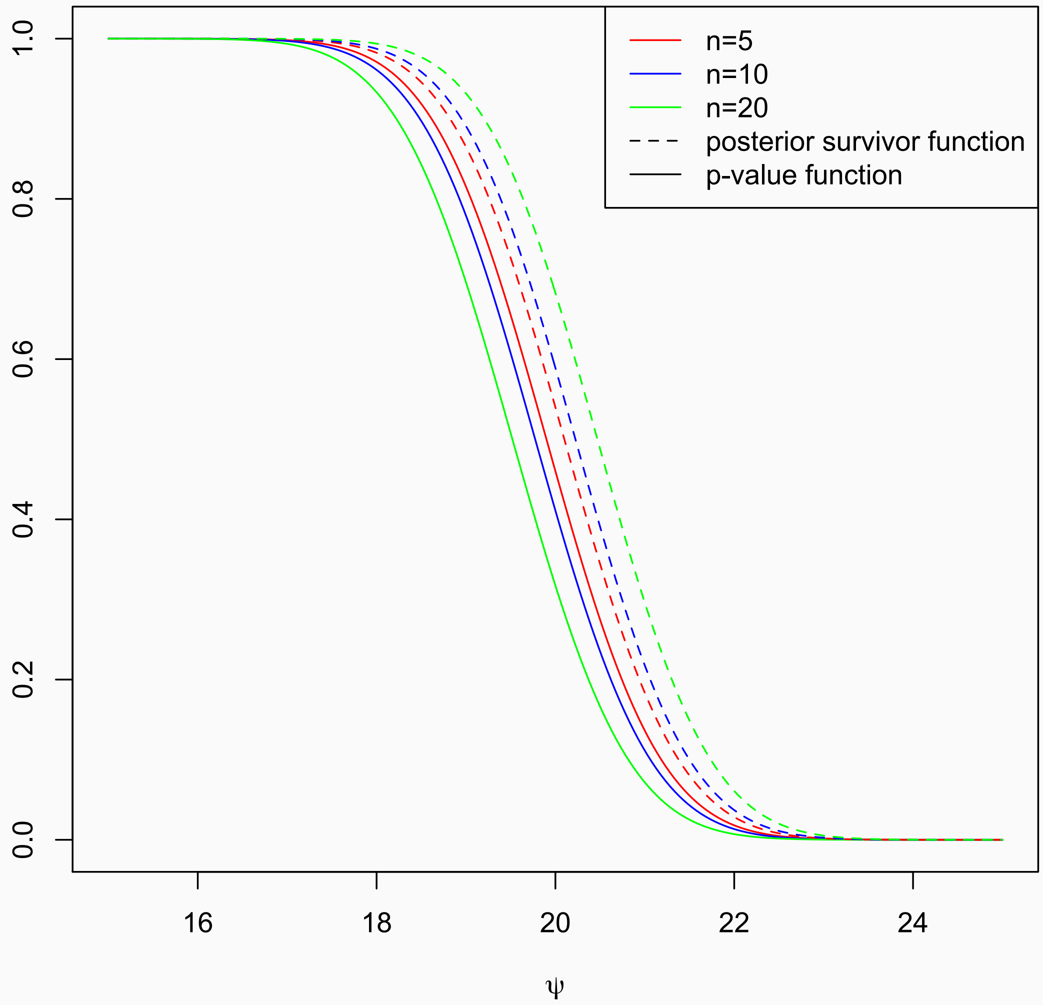
$A = \{(\mu_1, \ldots, \mu_n) : ||\mu||^2 = \psi\}$

- $\mu_i \mid x_i \sim N(x_i, 1) \implies \sum \mu_i^2 \mid x^n \sim \chi_n^2(\sum x_i^2)$

noncentral χ_n^2, noncent $||x||^2$

$\hat{\psi}_B = E \uparrow = n + \sum_{i=1}^n x_i^2$

$$E_{\mu}(\hat{\psi}_B) = n + E\sum x_i^2 = n + n + \|\mu\|^2$$
$$= 2n + \psi$$



Bayesian philosophy

F1 Probability refers to limiting relative frequencies. Probabilities are objective properties of the real world.

F2 Parameters are fixed, unknown constants. Because they are not fluctuating, no useful probability statements can be made about parameters.

F3 Statistical procedures should be designed to have well-defined long run frequency properties. For example, a 95 percent confidence interval should trap the true value of the parameter with limiting frequency at least 95 percent.

- $x_i \mid \theta_i \sim N(\theta_i, v_i)$
- $\theta_i \mid \mu \sim N(\mu, \sigma^2)$
- $\mu \sim N(\mu_0, \tau^2)$

v_i known

σ^2 known

hyperparameters

$\theta_i \mid \underset{\sim}{x} \sim N(- , -)$

$$f(\underset{\sim}{\theta}, \mu \mid \underset{\sim}{x}) = f(\underset{\sim}{x}; \underset{\sim}{\theta}) \times \pi(\underset{\sim}{\theta} \mid \mu) \times \pi(\mu; \mu_0)$$

$E(\theta_i \mid \underset{\sim}{x}) = \hat{\theta}_{i,B}$ squared error loss

$E(\mu \mid \underset{\sim}{x}) = \hat{\mu}_B$ " " "

- $x_i \mid \theta_i \sim N(\theta_i, v_i)$
- $\theta_i \mid \mu \sim N(\mu, \sigma^2)$
- $\mu \sim N(\mu_0, \tau^2)$
- $\pi(\theta, \mu \mid x)$ m.variate normal

$$\pi(\theta_i \mid \underset{\sim}{x}) = N\left(x_i\left(\frac{\sigma^2}{\sigma^2+v_i}\right) + \mu\left(\frac{v_i}{\sigma^2+v_i}\right),\ ?\right)$$

hyperparameters

$$\tilde{\theta}_{i,B} = x_i\left(\frac{\sigma^2}{\sigma^2+v_i}\right) + \underbrace{E(\mu \mid \underset{\sim}{x})}_{\tilde{\mu}_B}\left(\frac{v_i}{\sigma^2+v_i}\right)$$

$$E(\mu \mid x) =$$

$$\mathrm{var}(\mu \mid x) =$$

$$E(\theta_i \mid x) =$$

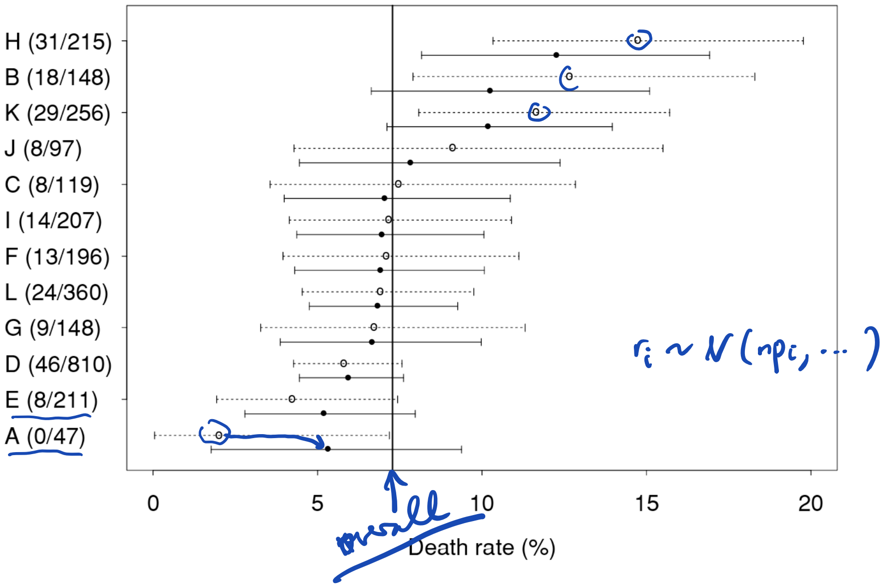


Figure 11.11 Posterior summaries for mortality rates for cardiac surgery data. Posterior means and 0.95 equitailed credible intervals for separate analyses for each hospital are shown by hollow circles and dotted lines, while blobs and solid lines show the corresponding quantities for a hierarchical model. Note the shrinkage of the estimates for the hierarchical model towards the overall posterior mean rate, shown as the solid vertical line; the hierarchical intervals are slightly shorter than those for the simpler model.

- $$E(\theta_i \mid x) = x_i \frac{\sigma^2}{\sigma^2 + v_i} + E(\mu \mid x)(1 - \frac{\sigma^2}{\sigma^2 + v_i})$$
- $$\tilde{\mu}_B = E(\mu \mid x) = \frac{\mu_0/\tau^2 + \sum x_i/(\sigma^2 + v_i)}{1/\tau^2 + \sum 1/(\sigma^2 + v_i)}$$

  $$= \frac{\mu_0/\tau^2}{\cdots} + \frac{\sum x_i/(\sigma^2+v_i)}{}$$
- If σ^2 unknown, then need to sample from the posterior, no closed form available
- Figure 11.11 applies similar ideas, plus sampling from the posterior, in logistic regression

$$\tilde{\theta}_{i,B} = x_i \left(\frac{\sigma^2}{\sigma^2+v_i}\right) + \tilde{\mu}_B \left(\frac{v_i}{\sigma^2+v_i}\right)$$

JS $\hat{\theta}_{i,JS}$

$$\left(1 - \frac{n-2}{\sum x_i^2}\right) \cdot x_i$$

if $\mu_0 = 0$

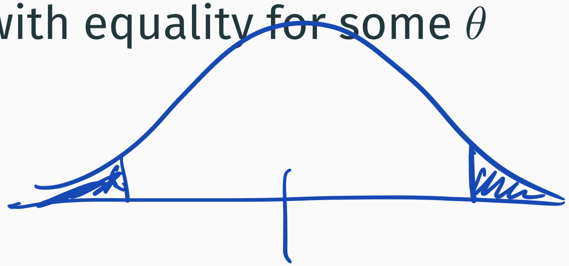
many steps

- $X_1, \ldots, X_n$ 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

$$\mathrm{pr}_\theta\{L(\boldsymbol{X}) \le \theta \le U(\boldsymbol{X})\} \ge 1-\alpha,$$

MS uses p

for all θ, with equality for some θ



$f(x;\theta)$

- in continuous cases usually can get $= 1-\alpha$ for all θ
- similarly, upper and lower $(1-\alpha)$-confidence bounds:

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

- exact limits if we have exact distribution of $\boldsymbol{X}$

$X_1, \ldots, X_n$

- approximate limits if $\mathrm{pr}_\theta\{L(\boldsymbol{X}) \le \theta \le U(\boldsymbol{X})\} \approx 1-\alpha$

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

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

"pivot on μ"

$$pr\{a \le \sqrt{n}(\bar{X} - \mu) \le b\} = \Phi(b) - \Phi(a)$$

$b > a$

$$pr\{\underbrace{\bar{X} - \frac{1.96}{\sqrt{n}}}_{L(\underline{X})} \le \mu \le \underbrace{\bar{X} + \frac{1.96}{\sqrt{n}}}_{U(\underline{X})}\} = \Phi(b) - \Phi(a) = 1 - \alpha$$

$\sqrt{n}(\bar{X} - \mu) \le b$

$\bar{X} - \mu \le \frac{b}{\sqrt{n}}$

$\bar{X} \le \frac{b}{\sqrt{n}} + \mu$

$\bar{X} - \frac{b}{\sqrt{n}} \le \mu$

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

$X_i \sim U(0,\theta)$

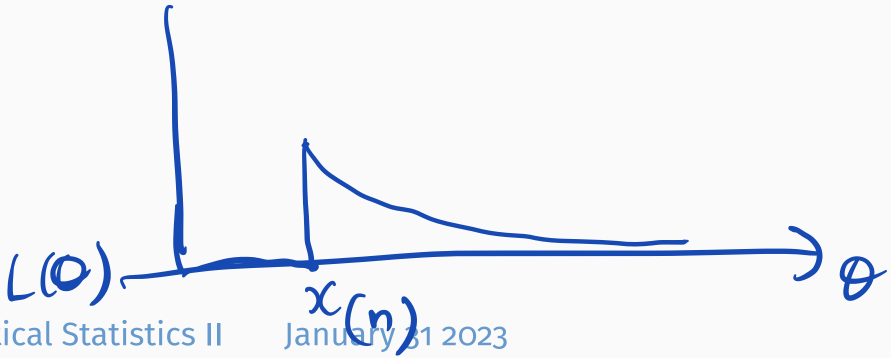
$X_i/\theta \sim U(0,1)$

$$L(\theta; \underline{x}) = \prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} \prod_{i=1}^{n} \cdot \left(\frac{1}{\theta}\right) & 0 \le x_i \le \theta, \quad i=1,\ldots,n \\ 0 & \end{cases}$$

$\hat{\theta}_{mle} = X_{(n)}$

$= \theta^{-n}, \quad \theta \geq x_1, x_2, \ldots, x_n$

$x_{(n)} = \max$



Exact confidence intervals

- Example: $X_1, \ldots, X_n$ i.i.d. $N(\mu, 1)$
- Example $X_1, \ldots, X_n$ i.i.d. $U(0, \theta)$
- Example $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$

$$P_\theta(a \le X_{(n)} \le b)$$

$$b = 1, \quad a = (1-\alpha)^{1/n}$$

$$= P_r\left(a' \le \frac{X_{(n)}}{\theta} \le b'\right) = G(b') - G(a')$$

$$= b'^n - a'^n$$

$$L(\underset{\sim}{X}) = \frac{X_{(n)}}{b'} \qquad U(\underset{\sim}{X}) = \frac{X_{(n)}}{a'}$$

$$\cancel{P_r\left(\frac{X_{(n)}}{\theta} \le b'\right) = P_r(X_{(n)} \le b'\theta)}$$

$$P_r\left(\frac{X_i}{\theta} \le x\right) = x$$

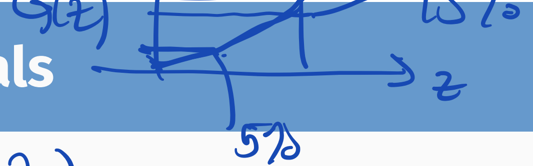
$$= P_r(a'\theta \le X_{(n)} \le b'\theta)$$

$$\boxed{= P_r(X_{(n)}/b' \le \theta \le X_{(n)}/a')}$$

$P_r(Z_{(n)} \le z)$, if $Z_{(n)}$ = largest of $Z_1, \ldots, Z_n \sim \text{iid } U(0,1)$

$= z^n$, $0 < z \le 1$

$1-\alpha$



$\rightarrow$ Corr = (J.17)

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

$\hat{\theta}$ = mle $= \dfrac{X}{n}$

$$\boxed{f(\hat{\theta}; \theta) = \dots}$$

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

pivot thus ~~ut~~ x/n

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

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

$\theta \in (0,1)$

$$\left[\bar{x} \pm t_{n-1}^{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right]$$

$$a\theta^2 + b\theta + c \le 0$$

$$\theta = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

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

$\hat\theta \pm 1.96\ \widehat{se}(\hat\theta)$

approx 95% CI

$\hat\theta_{mle} \dot\sim N(\theta, \{nI(\theta)\}^{-1})$

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

$\hat\theta \pm \frac{I_1(\hat\theta)^{+1/2}}{\sqrt n}$

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

$= \hat\theta \pm \{nI(\hat\theta)\}^{+1/2}$

$= \mathrm{pr}\left(-\frac{1.96\{\hat\theta(1-\hat\theta)\}^{1/2}}{\sqrt n} \le \hat\theta - \theta \le \frac{1.96\{\hat\theta(1-\hat\theta)\}^{1/2}}{\sqrt n}\right)$

$\mathrm{pr}\{\hat\theta - \quad \le \theta \le \hat\theta + \ldots\}$

Approximate confidence intervals

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

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

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

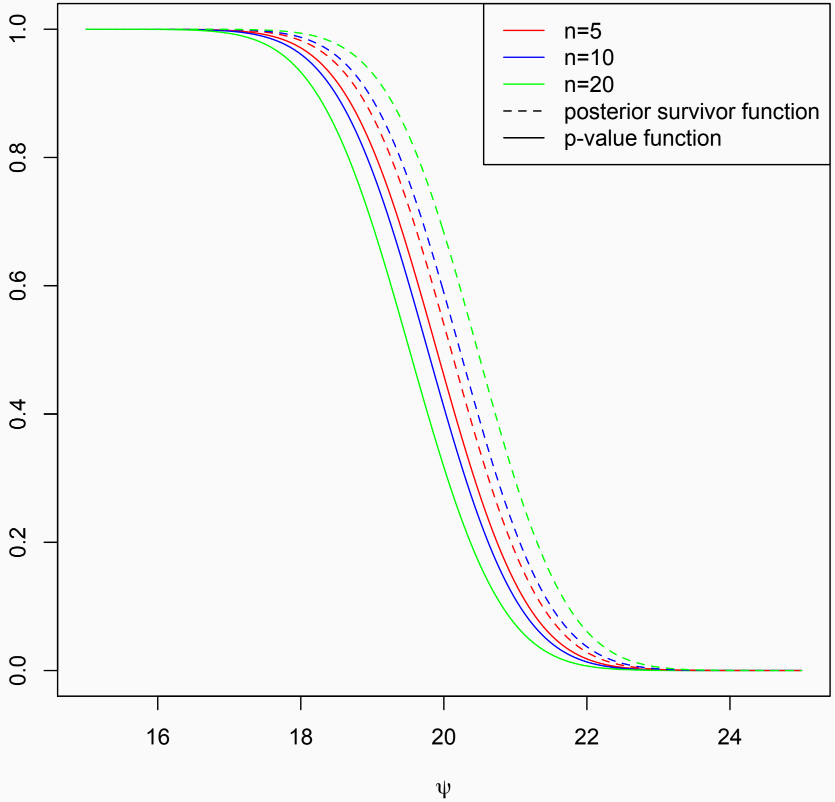
- $\hat{\theta}_n$ maximum likelihood estimate $\hat{\theta} \dot\sim N[0, \{nI(\theta)\}^{-1}]$
- approximate 95% confidence interval AoS Thm 6.16

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

Bayesian credible intervals



- upper and lower bounds

- upper and lower bounds
- highest posterior density

… Bayesian credible intervals

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

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

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

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

Approximate normality of posterior

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

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

- careful statement $\qquad$ 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 $\ell'(\theta; x^n) = 0$, 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

Approximate normality of posterior

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

$x^n = (x_1, \ldots, x_n)$

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

- approximate posterior probability intervals

- exact posterior probability intervals

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