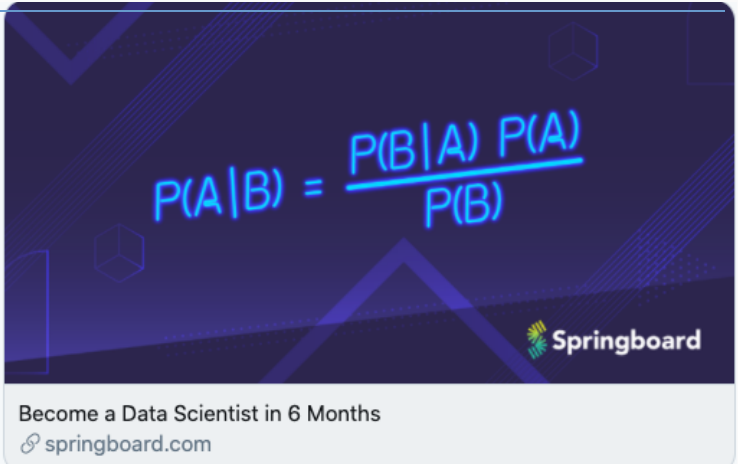
Mathematical Statistics II

STA2212H S LEC9101

Week 6

February 24 2021

Start recording!



STA 2212S: Mathematical Statistics II
Syllabus Spring 2021

Updated Feb 3

Week	Date	Methods	References
1	Jan 13/15	Review of parametric inference	AoS Ch 9
2	Jan 20/22	~~Significance testing~~Hypothesis testing	AoS Ch 10.1,2,~~6~~,7; SM Ch 7.3.2, 4
3	Jan 27/29	Significance testing	AoS Ch 10.2, 6; SM Ch 7.3.1, Ch 4
4	Feb 3/5	Goodness of fit testing, Intro to multiple testing	AoS Ch 10.3,4,5,8; SM p.327-8 (hard)
5	Feb 10/12	Multiple testing and FDR	AoS Ch 10.7, EH Ch 15.1,2
6	Feb 17/19	Break	
7	Feb 24/26	Bayesian Inference	AoS Ch 11.1-4; SM Ch 11.1,2; EH Ch 3, 13
8	Mar 3/5	Bayesian Inference	AoS Ch 11.5-9; SM Ch 11.4
9	Mar 10/12	Empirical Bayes	EH Ch 6
10	Mar 17/19	Statistical Decision theory	AoS Ch 12
11	Mar 24/26	Multivariate Models	AoS Ch 14; SM Ch 6.3
12	Mar 31	Causal Inference	AoS Ch 16
13	Apr 7	Recap	

1. HW 5 updated
2. Bayesian inference Part I
3. Friday: Solutions re HW 4, especially careful analysis re bonus question; if time, Proof of B-H FDR control

- Toronto Workshop on Reproducibility Feb 25-26
- Mar 1 3.00 pm EST Rebecca Barter
  "Teaching Data Science in the Real World" Link
  Data Science ARES
- Feb 25 1.00 pm EST Dylan Small
  CANSSI National Seminar Series (Journal Club; Slack)

version 1. $P(A \mid B) = \dfrac{P(\cancel{B}A, B)}{P(B)} = \dfrac{P(B|A)\,P(A)}{P(B)}$ $P(B) > 0$

cond'l prob.

version 2. $\underbrace{A_1, \ldots, A_k}$ partition the sample space

$P(A_j \mid B) = \dfrac{P(B|A_j)P(A_j)}{P(B)} = \dfrac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$ "discrete"

version 3. random variables Y, Z,

$f_{Z|Y}(z \mid y) = \dfrac{f_{Y|Z}(y|z)f_Z(z)}{\int f_{Y|Z}(y|z)f_Z(z)} = \dfrac{f_{Y|Z}(y|z)f_Z(z)}{f_Y(y)}$ if $f_Y(y) > 0$ on support $\mathcal{Y}$

Schervish text

Bayes' theorem: Example

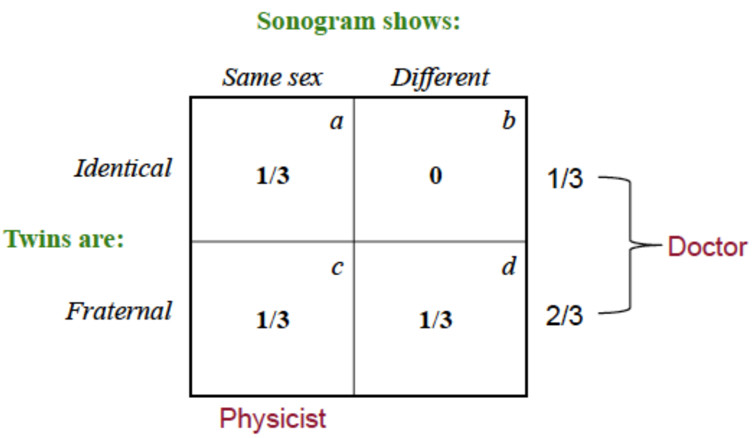


Figure 3.1 Analyzing the twins problem.

$$P(\text{ident} \mid s.s.)$$

$$= \frac{P(s.s. \mid id) \cdot P(id)}{P(ss \mid id)P(id) + P(ss \mid \overline{id})P(\overline{id})}$$

$$= \frac{1 \cdot \frac{1}{3}}{+\frac{1}{3} \cdot 1 + \frac{1}{2} \cdot \frac{2}{3}} = \frac{1}{2}$$

Sonogram shows twin boys. What is the probability they are identical twins?

link

(numbers)

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

TP/P

prevalence "in pop"

$$\text{PPV} = \Pr(\text{C19 pos} \mid \text{test pos}) = \frac{P(\text{+ve} \mid c19)\cdot \underbrace{P(c19)}_{\text{prevalence}}}{P(\text{+ve} \mid c19)\, P(c19) + P(\text{+ve} \mid \overline{c19})\, P(\overline{c19})}$$

TP/P

FP/N

positive predictive value

model $X \sim f_{X|\theta}(x|\theta)$ $\quad \theta \in \Theta \subseteq \mathbb{R}^k$ $\quad$ st. model

prior $\theta \sim \pi_{\Theta}(\Theta)$ $\quad \int \pi(\theta)\,d\theta = 1$ (maybe) proper

posterior $\boxed{f(\theta|x)} = \dfrac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)\,d\theta} = \dfrac{f(x|\theta)\pi(\theta)}{f(x)}$ or $m(x)$

sample $X_1 \dots X_n$ $\quad \boxed{f(\theta|\underset{\sim}{x})} = \dfrac{f(\underset{\sim}{x}|\theta)\pi(\theta)}{f(\underset{\sim}{x})}$

$\propto L(\theta; \underset{\sim}{x})\pi(\theta)$

is "everything"

AoS box in §11.2

$f(\underset{\sim}{x}) > 0$

$X_1, \ldots, X_n$ i.i.d. Bernoulli (p)

$$f(\underset{\sim}{x} \mid p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} = p^s (1-p)^{n-s} \qquad s = \sum x_i$$

$$f(s \mid p) = \binom{n}{s} p^s (1-p)^{n-s}, \quad 0 \le p \le 1; \quad 0 \le s \le n, \; s \in \{0, \ldots, n\}$$

$$L(p \mid s) = p^s (1-p)^{n-s}$$

$$f(p \mid s) = \cancel{\binom{n}{s}} \underbrace{p^s (1-p)^{n-s}} \cdot \underbrace{\pi(p)} \Big/ \int \cancel{\binom{n}{s}} p^s (1-p)^{n-s} \, dp$$

$$\pi(p) = \frac{p^{a-1} (1-p)^{b-1}}{B(a,b)} \qquad 0 \le p \le 1$$

Beta (a, b) dist$^{\underline{n}}$

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$\pi(p|s) \propto \underline{p^s(1-p)^{n-s} \cdot p^{a-1}(1-p)^{b-1}}$$

$$\hookrightarrow \pi(p|s) = \frac{p^{s+a-1}(1-p)^{n-s+b-1}}{\int \cdots\cdots \, dp}$$

$$\pi(p|s) = \frac{p^{s+a-1}(1-p)^{n-s+b-1}}{B(a+s,\ b+n-s)}$$

special case $a = b = 1$ $\boxed{\pi(p) = 1}$ $0 \le p \le 1$

$$\pi_{sp}(p|s) = \frac{p^s(1-p)^{n-s}}{B(s+1,\ n-s+1)} \qquad 0 \le p \le 1$$

$$= \frac{\Gamma(s+1)\Gamma(n-s+1)}{\Gamma(n+2)} \cdot p^s(1-p)^{n-s} \qquad B(a,b) = \underline{\Gamma(a)\Gamma(b)}$$

$$E\{B(a,b)\} = \frac{a}{a+b} \qquad E(p|s) = \frac{s+1}{n+2}$$

$$\hat{p} = \frac{s}{n}$$

E.g. if we decide to estimate p by $E(p|s)$, then $\tilde{p}_B = \hat{p}\,\cancel{W} + (1-\cancel{W}) \cdot \frac{1}{2}$

$$\tilde{p}_B = \underset{\substack{\uparrow \\ \text{sample mean}}}{\hat{p}} \cdot w + \underset{\substack{\uparrow \\ \text{prior mean}}}{\tfrac{1}{2}}(1-w) \qquad w = \frac{n}{n+2} \simeq 1$$

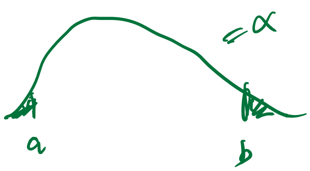
$E\{U(0,1)\}$

~~get~~ $\pi(\theta \mid \underset{\sim}{x})$ $(\theta \in \mathbb{R})$

→ est. θ by $E(\theta \mid \underset{\sim}{x})$

est. θ by $\arg\sup_{\theta} \pi(\theta \mid x)$

→ posterior interval



$$\int_a^b \pi(\theta \mid \underset{\sim}{x})\, d\theta = 1-\alpha$$

credible interval

Table 3.1 *Scores from two tests taken by 22 students,* mechanics *and* vectors.

	1	2	3	4	5	6	7	8	9	10	11
mechanics	7	44	49	59	34	46	0	32	49	52	44
vectors	51	69	41	70	42	40	40	45	57	64	61

	12	13	14	15	16	17	18	19	20	21	22
mechanics	36	42	5	22	18	41	48	31	42	46	63
vectors	59	60	30	58	51	63	38	42	69	49	63

Table 3.1 shows the scores on two tests, mechanics and vectors, achieved by $n = 22$ students. The sample correlation coefficient between the two scores is $\hat{\theta} = 0.498$,

$$\hat{\theta} = \sum_{i=1}^{22}(m_i - \bar{m})(v_i - \bar{v}) \Bigg/ \left[\sum_{i=1}^{22}(m_i - \bar{m})^2 \sum_{i=1}^{22}(v_i - \bar{v})^2\right]^{1/2}, \quad (3.10)$$

with m and v short for mechanics and vectors, $\bar{m}$ and $\bar{v}$ their averages. We wish to assign a Bayesian measure of posterior accuracy to the true correlation coefficient θ, "true" meaning the correlation for the hypothetical population of all students, of which we observed only 22.

If we assume that the joint (m, v) distribution is bivariate normal (as

$$(X_i, Y_i) \sim N_2\left(\begin{pmatrix}\mu_X \\ \mu_Y\end{pmatrix}, \Sigma\right)$$

$$\Sigma = \begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}$$

$$\pi(\rho)$$

$$\pi(\rho \mid \underset{\sim}{x}, y)$$

$$\hat{\rho} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

(mle)

bb

$$\pi(\rho \mid \underline{x}, y) = f(\hat{\rho} \mid \rho) \cdot \pi(\rho) \Big/ \left(\int f(\hat{\rho} \mid \rho) \pi(\rho) d\rho \right)$$

$$-1 \le \rho \le 1$$

$$\propto \frac{1}{\pi} (n-2)(1-\rho^2)^{(n-1)/2} (1-\hat{\rho}^2)^{(n-4)/2} \cdot \int_0^\infty \frac{dw}{(\cosh w - \rho\hat{\rho})^{n-1}} \pi(\rho)$$

$$\pi_1(\rho) = \frac{1}{2} \quad -1 \le \rho \le 1 \qquad -1 \le \rho \le 1$$

$$\pi_2(\rho) = \frac{1}{1-\rho^2} \quad -1 \le \rho \le 1$$

$$\pi_3(\rho) = 1 - |\rho|, \quad -1 \le \rho \le 1$$

Example: Bivariate normal

$n = 22$

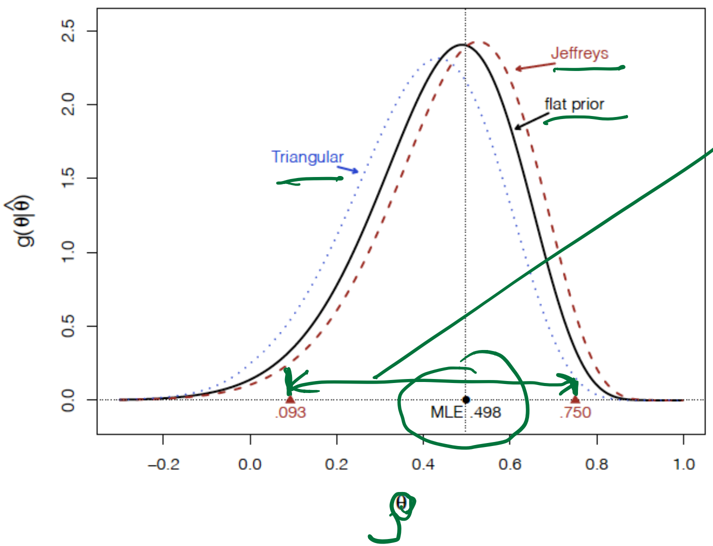


Figure 3.2 Student scores data; posterior density of correlation θ for three possible priors.

credible interval

Jeffreys' prior

each

$$\int \pi_k(\rho \mid \hat{\rho})\,d\rho = 1$$

$k = 1, 2, 3$

are "normalized"

$$\pi_k(\rho \mid \hat{\rho}) \propto L(\hat{\rho}; \rho)\,\pi(\rho)$$

Example: Binomial

SM Ex.11.11

Table 11.2 Mortality rates r/m from cardiac surgery in 12 hospitals (Spiegelhalter *et al.*, 1996b, p. 15). Shown are the numbers of deaths r out of m operations.

A	0/47	*B*	18/148	*C*	8/119	*D*	46/810	*E*	8/211	*F*	13/196
G	9/148	*H*	31/215	*I*	14/207	*J*	8/97	*K*	29/256	*L*	24/360

provided the mode lies inside the parameter space. Here $\tilde{J}(\theta)$ is the second derivative matrix of $-\tilde{\ell}(\theta)$. This expansion corresponds to a posterior multivariate normal

12 hospitals
mortality rate

1. hospital A only $m = 47$ $r = 0$ $\hat{p} = 0$

2. all hospitals together

$r_H \sim \text{Bin}(m_H, p_{\cancel{H}})$ p same

$\pi(p) \propto U(0,1)$: $\tilde{p}_{Bayes,A} = \dfrac{0+1}{47+2}$

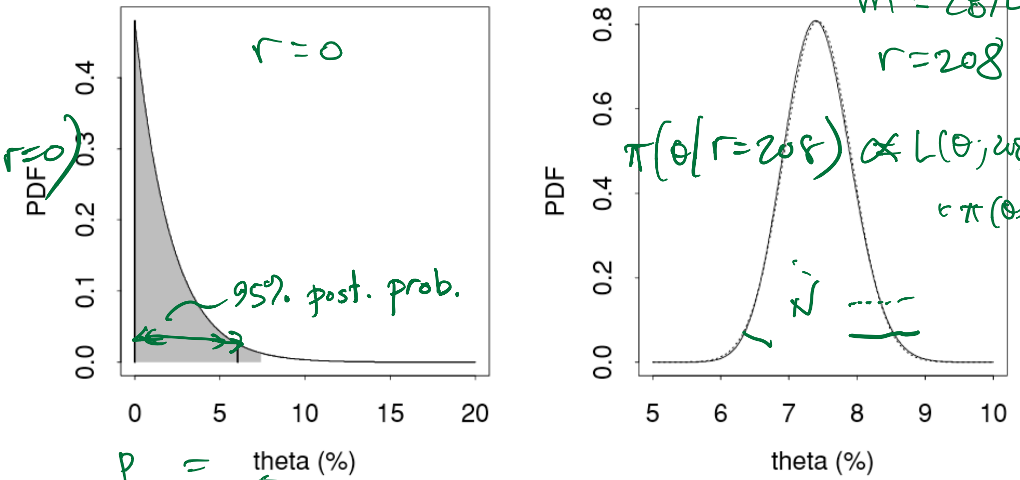
Example: Binomial

SM Ex.11.11

580 — 11 · Bayesian Models

$m = 47$ — S.S.

$r = 0$



$\pi(\theta \mid r=0)$

95% post. prob.

p = theta (%)

A

$m = 2814$ — S.S.

$r = 208$

$\pi(\theta \mid r=208) \propto L(\theta; 208) \times \pi(\theta)$

N

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).

all tog.

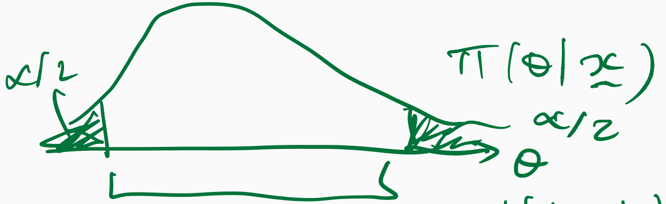
$r = 208 \quad m = 2814$

$\tilde{p} \; (208+1)/(2814+2) \approx \hat{p} = \frac{208}{2814}$

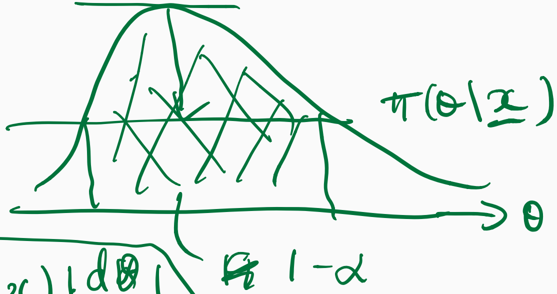
Posterior summaries

point estimates — posterior mean, mode, median

posterior intervals 1 — equi-tailed



posterior intervals 2 — highest post. density intervals



functions of the parameter

$$\pi(\theta|\underset{\sim}{x}) \propto L(\theta;\underset{\sim}{x})\,\pi(\theta)$$

$$\alpha = \alpha(\theta) \qquad \boxed{\pi(\alpha|\underset{\sim}{x}) = \pi(\theta|\underset{\sim}{x})\left|\frac{d\theta}{d\alpha}\right|}$$

Posterior summaries

SM §11.2.1; AoS §11.2

point estimates $\pi(\theta \mid \underline{x})$ — mean, median, mode (posterior)

posterior intervals 1 — $\tilde{\theta} \pm \widetilde{s.e.} \cdot z_{\alpha/2}$ ← approx (normal)

or find $\int_{-\infty}^{\theta_L} \pi(\theta \mid x)\,d\theta = \frac{\alpha}{2}$; $\int_{\theta_u}^{\infty} \pi(\theta \mid x)\,d\theta = \frac{\alpha}{2}$

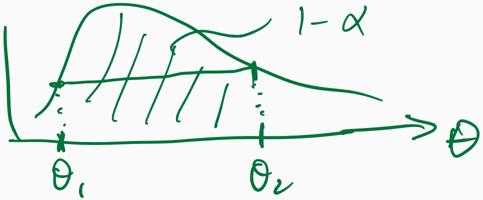
$\Longrightarrow (\theta_L, \theta_u)$ posterior credible int. of prob $1-\alpha$

↑ exact

posterior intervals 2 — highest post. density (HPD)

~~functions of the parameter~~

exact → $\pi(\theta \mid \underline{x})$



Functions of the parameter

$S \sim \text{Bin}(n, p)$ $\psi = \log\{p/(1-p)\}$ parameter of interest

$$\pi(p \mid s) = \frac{p^s (1-p)^{n-s}}{B(s+1, n-s+1)}$$

$0 \leq p \leq 1$ flat

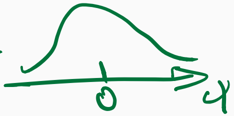
$\pi(p) = 1$

$$p = \frac{e^\psi}{1+e^\psi}; \quad \psi = \log \frac{p}{1-p}$$

$$\pi(\psi \mid s) \propto \{e^\psi/(1+e^\psi)\}^s \{1/(1+e^\psi)\}^{n-s} \cdot \frac{e^\psi}{(1+e^\psi)^2}$$

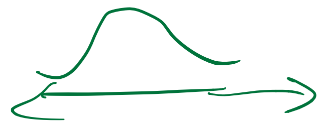
mistake in AoS density

$$\pi(\psi) = \frac{e^\psi}{(1+e^\psi)^2}$$ logistic dens.



$\pi(p) = 1 \qquad 0 \le p \le 1$

Jeffreys' prior

$\Rightarrow \pi(\psi) =$  $\psi \in \mathbb{R}$

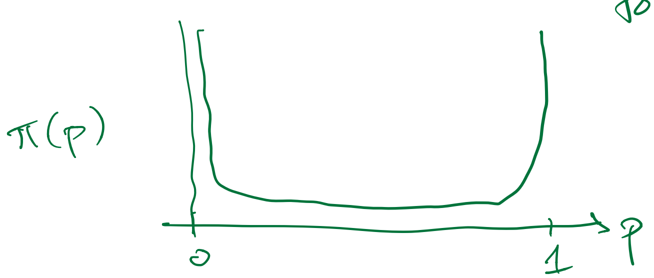
$\pi(\psi) = 1, \qquad -\infty \le \psi \le \infty$ not proper

$$\pi(\psi|s) = \frac{L(\psi; s) \cdot \pi(\psi)}{\int \cdots \pi(\psi)\, d\psi} \leftarrow \text{ok if } < \infty$$

$\int \pi(\psi|s)\, d\psi = 1$

I *think* if $\pi(\psi) = 1$ on $\mathbb{R}$ that $\pi(\psi|s)$ is <u>proper</u> ✓ yes (I checked)

if $\pi(\psi) = 1$ what's the implied prior for p: it's Beta$(\frac{1}{2}, \frac{1}{2})$



see SM Fig 5.4 top left

Priors of <u>convenience</u> can be "misleading"

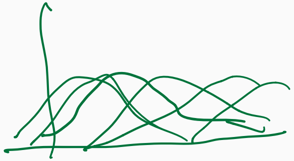
Approximate posterior

$$\pi(\theta \mid \underset{\sim}{x}) = \frac{L(\theta; \underset{\sim}{x})\,\pi(\theta)}{\int L(\theta; \underset{\sim}{x})\,\pi(\theta)\,d\theta}$$

dist$^{\underline{n}}$ of $\pi(\theta \mid \underset{\sim}{X})$

r.v.

$$\underset{\sim}{X} \sim \prod_{i=1}^{n} f(x_i;\theta) = \underline{\underline{f(\underset{\sim}{x};\theta)}}$$



$$= e^{\tilde{\ell}(\theta; \underset{\sim}{x})} \Big/ \int e^{\tilde{\ell}(\theta; \underset{\sim}{x})}\, d\theta$$

$$\tilde{\ell}(\theta; x) = \ell(\theta; x) + \log \pi(\theta)$$

$$= \underline{\tilde{\ell}(\tilde{\theta}; x)} + \cancel{(\theta - \hat{\tilde{\theta}})\,\tilde{\ell}'(\tilde{\theta})} + \underline{\tfrac{1}{2}(\theta - \tilde{\theta})^2\,\tilde{\ell}''(\tilde{\theta})} + \dots$$

assume $\tilde{\ell}'(\hat{\theta}) = 0$

$$\simeq \tilde{\ell}(\tilde{\theta}) + \frac{1}{2}(\theta-\tilde{\theta})^2\,\tilde{\ell}''(\tilde{\theta}) + \frac{1}{3}(\theta-\tilde{\theta})^3\,\tilde{\ell}'''(\theta^*)$$

$$\boxed{\theta \le \theta^* \le \tilde{\theta}}$$

$$\pi(\theta|\underset{\sim}{x}) = \frac{e^{\tilde{\ell}(\theta)}}{\int e^{\tilde{\ell}(\theta)}\,d\theta} = \frac{\cancel{e^{\tilde{\ell}(\tilde{\theta})}}\; e^{-\frac{1}{2}(\theta-\tilde{\theta})^2[-\tilde{\ell}''(\tilde{\theta})]}}{\cancel{e^{\tilde{\ell}(\tilde{\theta})}}\int e^{-\frac{1}{2}(\theta-\tilde{\theta})^2[-\tilde{\ell}''(\tilde{\theta})]}\,d\theta}$$

— N

$\to$ const.

$$\pi(\theta|\underset{\sim}{x}) \simeq N\left(\tilde{\theta}\,;\{-\tilde{\ell}''(\tilde{\theta})\}^{-1}\right)$$

(recall $X \sim N(\mu,\sigma^2)$: $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

$$\pi(\theta|\underset{\sim}{x}) = \frac{e^{-\frac{1}{2}(\theta-\tilde{\theta})^2\{-\ell''(\tilde{\theta})\}}}{\sqrt{2\pi}\cdot[-\tilde{\ell}''(\tilde{\theta})]^{-1/2}}$$

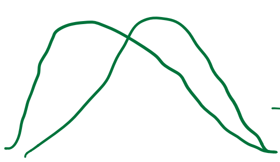
$$= \frac{\{-\tilde{\ell}''(\hat{\theta})\}^{\frac{1}{2}}}{\sqrt{2\pi}}\,e^{-(\theta-\hat{\theta})^2\{-\ell''(\hat{\theta})\}/2}$$

rigorous : $\tilde{\theta} \xrightarrow{\cancel{P}} \theta$ in probability

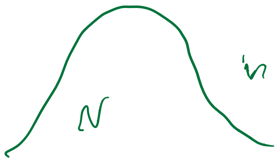
$\sqrt{n}(\tilde{\theta}-\theta) \xrightarrow{\cancel{d}}$ in dist^n (to N)

assume that $\exists\ \tilde{\theta} = \tilde{\theta}(\underset{\sim}{x})$ s.t.

$\pi(\tilde{\theta}|x) \geq \pi(\theta|\underset{\sim}{x})$ all $\underline{\theta} \in \underline{\Theta}$



- that $\tilde{\theta}$ also satisfies $\frac{\partial}{\partial\theta}\log \pi(\tilde{\theta}|x) = 0$
- that $\pi(\theta)$ is proper.



in limit.

$$\frac{\partial}{\partial\theta}\log\pi(\tilde\theta\,|\,\underset{\sim}{x})=0$$

$$=\frac{\partial}{\partial\theta}\ell(\tilde\theta;\underset{\sim}{x})+\frac{\partial}{\partial\theta}\log\pi(\tilde\theta)$$

$$\frac{\partial}{\partial\theta}\ell(\hat\theta;x)+\underbrace{\frac{\partial}{\partial\theta}\log\pi(\tilde\theta)}=0$$

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

leave prior "down"

def: fcale

$$\ell(\theta;\underset{\sim}{x})=\ell(\hat\theta;\underset{\sim}{x})+\cancel{(\theta-\hat\theta)\ell'(\hat\theta;x)}+\frac{1}{2}(\theta-\hat\theta)^2\ell''(\hat\theta;x)$$

$$\pi(\theta|y)\approx\frac{e^{\ell(\hat\theta;\underset{\sim}{x})}\cdot e^{-\frac{1}{2}(\theta-\hat\theta)^2\{-\ell''(\hat\theta)\}}\{\pi(\hat\theta)+\boxed{(\theta-\hat\theta)\pi'(\hat\theta)}+\dots\}\ \boxed{e^{\{\dots\}}}}{\int\ \dots\ \pi(\hat\theta)+\times}$$

$$\approx e^{-\frac{1}{2}(\theta-\hat\theta)^2\{-\ell''(\hat\theta)\}}\cdot\frac{1}{\sqrt{2\pi}}\left|-\ell''(\hat\theta)\right|^{1/2}$$

$$N(\hat\theta,\ j^{-1}(\hat\theta))$$ $j(\theta)=$ F. inf. f^n.

$$\boxed{\hat\theta\pm z_{\alpha/2}\{j(\hat\theta)\}^{1/2}\quad\text{approx } 1-\alpha \text{ credible interval}}$$

$$\theta\,\dot\sim\,N(\underline{\hat\theta},\ \underline{\hat{j}^{-1}})\qquad(\theta-\hat\theta)\hat{j}^{1/2}\,\dot\sim\,N(0,1)$$

$=$ $1-\alpha$ C.I. from N approx to $\hat\theta$

If $\theta \in \mathbb{R}$, and ... regularity ...

post. credible interval $\simeq$ [illegible] conf. interval

[illegible] as $n \to \infty$,

"prior is washed out by the data"