Bayesian Statistics

Instructor: Malay Ghosh Taken by Yu Zheng

Conjugate Priors

For a given likelihood, $L(\theta)$, $\theta \in \Theta$ if f_{θ} constitutes a family of priors such that the resulting posterior also belongs to the family \mathcal{F}_{θ} , then the family of priors f_{θ} is to be a conjugate family.

Examples:

- Binomial-Beta
- Poisson-Gamma
- Normal-Normal
- Uniform-Pareto:

$$egin{aligned} f(x \mid heta) &= rac{1}{ heta} I_{[0 < x \leq heta]} \ \pi(heta) \propto heta^{-lpha} \end{aligned}$$

Bayesian Inference

Point estimation: Posterior mean/variance/mode/...

Improper Prior

$$X_1,\ldots,X_n \mid heta \stackrel{iid}{\sim} \mathcal{N}(heta,1), \pi(heta) = 1, -\infty < heta < \infty.$$

$$egin{aligned} \pi(heta \mid x_1,\ldots,x_n) \propto& \exp[-rac{1}{2}\sum_{i=1}^n (x_i- heta)^2] \ &=& \exp[-rac{n}{2}(ar{x}- heta)^2 - rac{1}{2}\sum_{i=1}^n (x_i-ar{x})^2] \ &\propto& \exp[-rac{n}{2}(heta-ar{x})^2] \ &\Rightarrow& heta \mid x_1,\ldots,x_n \sim \mathcal{N}(ar{x},1/n). \end{aligned}$$

Non-conjugate Priors

 $X_1,\ldots,X_n \mid heta \stackrel{iid}{\sim} Bin(1, heta), 0 < heta < 1.$

$$egin{aligned} \phi = ext{logit}(heta) = ext{logit}(rac{ heta}{1- heta}) \ \phi &\sim \mathcal{N}(\mu,\sigma^2) \end{aligned}$$

Conjugacy is lost.

Credible Intervals and Sets

$$egin{aligned} X_1,\ldots,X_n \mid heta \sim f \ ext{Prior} \ \pi(heta), \ heta \in \Theta \end{aligned}$$

A set $C \subset \Theta$ is said to be a credible set of size $1 - \alpha$ if $P(\theta \in C \mid X) = 1 - \alpha$.

Remarks:

- In a frequentist framework, it is not possible to make a probabilistic statement related to a confidence interval. In a Bayesian framework, this is possible, since we are looking at the conditional distribution of a parameter given the data.
- Often the confidence set is an interval. E.g., $[\theta_1(x), \theta_2(x)]$ such that $P(\theta_1(x) \le \theta \le \theta_2(x) \mid x) = 1 \alpha$.

Example:

$$egin{aligned} X_1,\ldots,X_n \mid heta \stackrel{iid}{\sim} \mathcal{N}(heta, \underbrace{\sigma^2}_{ ext{known}}) \ heta &\sim \mathcal{N}(\mu, au^2) \ \ heta & = \frac{\sigma^2/n}{\sigma^2/n + au^2} \end{aligned}$$

An equal-tailed CI for heta of size $1 - \alpha$ is given by

$$(1-B)ar{x}+B\mu\pm z_{lpha/2}rac{\sigma}{\sqrt{n}}\sqrt{1-B}$$

• Equal-tailed intervals does not always provide the shortest interval.

E.g. Binomial with beta prior, Poisson with gamma prior, etc.

Bayesians have recommended what they call **Highest Posterior Density (HPD)** interval or more generally HPD set.

Suppose the posterior density of θ is unimodal. Then the HPD interval for θ is given by $C = \{\theta : \pi(\theta \mid x) \ge k\}$ where k is chosen so that $P(\theta \in C \mid x) = 1 - \alpha$.

For symmetric unnimodal distributions, HPD intervals are equal-tailed.

Credible sets

$$egin{aligned} X & \mid heta \sim \mathcal{N}(heta, \Sigma) \ & heta \sim \mathcal{N}(\mu, \Gamma) \end{aligned}$$

 $X= heta+e, \quad e\sim \mathcal{N}(0,\Sigma), \; e ext{ independent of } heta$

$$egin{split} \begin{pmatrix} X \\ heta \end{pmatrix} \sim \mathcal{N}_{2p} \left[\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \Sigma + \Gamma & \Gamma \\ \Gamma & \Gamma \end{pmatrix}
ight] \ \Theta \mid X \sim \mathcal{N}[\mu + \Gamma(\Sigma + \Gamma)^{-1}(X - \mu), \Gamma - \Gamma(\Sigma + \Gamma)^{-1}\Gamma \ B \stackrel{\Delta}{=} \Sigma(\Sigma + \Gamma)^{-1}, \ I - B = I - (\Sigma + \Gamma - \Gamma)(\Sigma + \Gamma)^{-1} = \Gamma(\Sigma + \Gamma)^{-1} \end{split}$$

$$egin{aligned} &\Gamma - \Gamma(\Sigma+\Gamma)^{-1}\Gamma = &\Gamma - (\Gamma+\Sigma-\Sigma)(\Sigma+\Gamma)^{-1}\Gamma\ &=&\Gamma - \Gamma + \Sigma(\Sigma+\Gamma)^{-1}\Gamma\ &=&B\Gamma = (I-B)\Sigma\ &(I-B)\Sigma = &\Sigma - \Sigma(\Sigma+\Gamma)^{-1}\Sigma\ &=&\Sigma - (\Sigma+\Gamma-\Gamma)(\Sigma+\Gamma)^{-1}\Sigma\ &=&\Gamma B = (I-B)\Sigma \end{aligned}$$

HPD Confidence set is an ellipsoid

$$\left\{ \theta : (\theta - ((I - B)X + B\mu))^T ((I - B)\Sigma)^{-1} (\theta - ((I - B)X + B\mu)) \le K \right\}$$

K is the chisquare percentile point.

Bayesian Hypothesis Testing

$$egin{array}{ll} H_0: heta\in\Theta_0 & H_a: heta\in\Theta_1 \ & ext{prior} & \pi(heta) \end{array}$$

Posteriror odds in favor of H_0 : $\frac{P(H_0|x)}{P(H_1|x)}$.

Prior odds in favor of H_0 : $\frac{\Pi(\Theta_0)}{\Pi(\Theta_1)} = \frac{\int_{\Theta_0} \pi(\theta) d\theta}{\int_{\Theta_1} \pi(\theta) d\theta}$.

 $\Pi_0 = P(heta \in \Theta_0)$

Neukel Prior: $\Pi_0 = 1/2$

 $g_0(heta)$: density of heta when $heta \in \Theta_0$; $g_1(heta)$: density of heta when $heta \in \Theta_1$. We have $\int_{\Theta_0} g_0(heta) d heta = \int_{\Theta_0} g_1(heta) d heta = 1$.

$$\pi(heta)=\pi_0g_0(heta)+(1-\pi_0)g_1(heta)
onumber \ m_\pi(x)=\int f(x\mid heta)\pi(heta)d heta=\pi_0\int_{\Theta_0}f(x\mid heta)g_0(heta)d heta+(1-\pi_0)\int_{\Theta_1}f(x\mid heta)g_1(heta)d heta$$

$$\pi(heta \mid x) = rac{f(x \mid heta)\pi(heta)}{m_{\pi}(x)} = egin{cases} rac{f(x \mid heta)g_0(heta)}{m_{\pi}(x)}, & heta \in \Theta_0 \ rac{f(x \mid heta)g_1(heta)}{m_{\pi}(x)}, & heta \in \Theta_1 \ \end{cases} \ p^{\pi}(H_0 \mid x) = rac{\pi_0 \int_{\Theta_0} f(x \mid heta)g_0(heta)d heta}{\pi_0 \int_{\Theta_0} f(x \mid heta)g_0(heta)d heta + (1 - \pi_0) \int_{\Theta_1} f(x \mid heta)g_1(heta)d heta} \end{cases}$$

Posterior odds in favor of H_0 :

$$rac{P(H_0 \mid x)}{P(H_1 \mid x)} = rac{\pi_0 \int_{\Theta_0} f(x \mid heta) g_0(heta) d heta}{(1-\pi_0) \int_{\Theta_1} f(x \mid heta) g_1(heta) d heta}$$

Prior odds in favor of H_0 :

$$\frac{\pi_0}{1-\pi_0}$$

$$egin{aligned} &\mathrm{BF}_{01} = \mathrm{Bayes} \ \mathrm{Factor} \ \mathrm{in} \ \mathrm{favor} \ \mathrm{of} \ H_0 \ &= rac{\mathrm{Posterior} \ \mathrm{odds} \ \mathrm{in} \ \mathrm{favor} \ \mathrm{of} \ H_0 \ &}{\mathrm{Prior} \ \mathrm{odds} \ \mathrm{in} \ \mathrm{favor} \ \mathrm{of} \ H_0 \ &} \ &= rac{\int_{\Theta_0} f(x \mid heta) g_0(heta) d heta}{\int_{\Theta_1} f(x \mid heta) g_1(heta) d heta} \end{aligned}$$

Remarks:

- While BF_{01} does not depend on the choice of π_0 , it does depend on the choice of g_0 and g_1 .
- $\operatorname{BF}_{10} = \frac{1}{\operatorname{BF}_{01}}.$
- $\log(BF_{01}) = \log(Posterior odds in favor of H_0) \log(Prior odds in favor of H_0)$, change from prior odds to posterior odds in the log scale.

Example:

$$egin{aligned} X_1,\ldots,X_n \mid \mu,\sigma^2 \stackrel{iid}{\sim} \mathcal{N}(\mu,\sigma^2) \ H_0:\mu=0 ext{ vs } H_1:\mu
eq 0 \ & \left\{ egin{aligned} g_0(\sigma^2)=(\sigma^2)^{-1} \ g_1(\mu,\sigma^2)=\mathcal{N}(\mu\mid 0,\sigma^2)(\sigma^2)^{-1} \end{aligned}
ight. \end{aligned}$$

Simple Null vs Simple Alternative

$$H_0: heta= heta_0 \quad H_1: heta= heta_1$$

Posterior odds in favor of $H_0 = \frac{\pi_0 f(x|\theta_0)}{(1-\pi_0)f(x|\theta_1)}$

 $BF_{01} = \frac{f(x|\theta_0)}{f(x|\theta_1)}$ Neyman-Pearson Likelihood Ratio

Example:

$$egin{aligned} X_1,\ldots,X_n \mid heta, \underbrace{\sigma^2}_{ ext{known}} \stackrel{iid}{\sim} \mathcal{N}(heta, \sigma^2) \ & H_0: heta \leq heta_0, \quad H_1: heta > heta_0 \ & heta \sim \mathcal{N}(\mu, au^2) \ & heta = \mathcal{N}(\mu, au^2) \ & heta \mid x \sim \mathcal{N}((1-B)ar{x} + B\mu, rac{\sigma^2}{n}(1-B)) \end{aligned}$$

Posterior odds in favor of H_0 :

$$rac{P(heta \leq heta_0 \mid x)}{P(heta > heta_0 \mid x)} = rac{\Phi(rac{ heta_0 - (1-B)ar{x} + B\mu}{rac{\sigma}{n}\sqrt{1-B}})}{1 - \Phi(rac{ heta_0 - (1-B)ar{x} + B\mu}{rac{\sigma}{n}\sqrt{1-B}})}$$

Examples:

• Sugar level for a person two hours after his breakfast.

Let $X \mid \theta \sim \mathcal{N}(\theta, 100), \theta$ being he true level, and $\theta \sim \mathcal{N}(100, 900).$

$$egin{aligned} & heta \mid X = x \sim \mathcal{N}((1-B)x + 100B, 100(1-B)) \xrightarrow{B=100/(100+900)=0.1}{\mathcal{N}(0.9x+10, 90)}. \end{aligned}$$
 Observe $x = 130.$ Then $P(heta \leq 130 \mid x) = 0.624.$

• IQ Test:

$$egin{aligned} X & \mid heta \sim \mathcal{N}(heta, 100) \ heta & \sim \mathcal{N}(100, 15^2) \end{aligned}$$

Posterior:

$$egin{aligned} & heta \mid X = x \sim \mathcal{N}((1-B)x + 100B, 100(1-B)) = & = 100/(100+225) = 4/13 \ \mathcal{N}(rac{9}{13}x + rac{400}{13}, rac{900}{13}) \ & = P(heta \leq 130 \mid x = 130) = ?. \end{aligned}$$

Point Null Hypothesis

$$egin{aligned} H_0: heta &= heta_0 \quad H_1: heta
eq heta_0 \ \pi_0 &= P(heta &= heta_0) \ \pi(heta) &= \pi_0 I_{[heta = heta_0]} + (1 - \pi_0) g_1(heta) I_{[heta
eq heta_0]} \ m_\pi(x) &= \pi_0 f(x \mid heta_0) + (1 - \pi_0) \int_{ heta
eq heta_0} f(x \mid heta) g_1(heta) d heta \end{aligned}$$

Spike and Slab Priors

$$\pi(heta)=\pi_0 I_{[heta=0]}+(1-\pi_0)g_1(heta)
onumber \ m_\pi(x)=\pi_0 f(x\mid 0)+(1-\pi_0)\int_{ heta
eq 0}f(x\mid heta)g_1(heta)d heta$$

 $g_1(heta)$ is typically $\mathcal{N}(heta \mid \mu, au^2).$

$$egin{aligned} P(heta = heta_0 \mid x) = & rac{\pi_0 f(x \mid heta_0)}{m_\pi(x)} = rac{\pi_0 f(x \mid heta_0)}{\pi_0 f(x \mid heta_0) + (1 - \pi_0) \int_{ heta
eq heta_0} f(x \mid heta) g_1(heta) d heta} \ &= & \left[1 + rac{1 - \pi_0}{\pi_0} \cdot rac{\int_{ heta
eq heta_0} f(x \mid heta) g_1(heta) d heta}{f(x \mid heta_0)}
ight]^{-1} \ & ext{BF}_{01} = rac{f(x \mid heta_0)}{\int_{ heta
eq heta_0} f(x \mid heta) g_1(heta) d heta} \end{aligned}$$

Interpretation of Bayes Factors (A. Raftry, Bren 1996 251-266)

BF_{10}	$2\log_e(\mathrm{BF}_{10})$	Evidence for H_0
<1	<0	Negative (Supports H_0)
1 - 3	0 - 2.2	Not worth more than a mention
3 - 20	2.2 - 6	Positive
20 - 150	6 - 10	Strong
>150	>10	Very strong

(Paper) Bayes Factors: What they are what they are not

Parameter space $\Omega; \Omega_H \subset \Omega, \Omega_A = \Omega - \Omega_H; f(x \mid \theta)$: pdf of x given $\theta \in \Omega$.

$$H_0: heta\in\Omega_H,\quad H_1: heta\in\Omega_A=\Omega-\Omega_H$$

$$f_H(x) = rac{\int_{\Omega_H} f(x \mid heta) d\mu(heta)}{\underbrace{\mu(\Omega_H)}_{:=p}}, \quad f_A(x) = rac{\int_{\Omega_A} f(x \mid heta) d\mu(heta)}{\mu(\Omega_A)}$$

$$egin{aligned} ext{Posterior odds} &= rac{p f_H(x)}{(1-p) f_A(x)} \ ext{Prior odss} &= rac{p}{1-p} \ ext{BF}_{01} &= rac{f_H(x)}{f_A(x)} \end{aligned}$$

 $\Omega=\{0,1/2,1\}$

6 hypotheses:

- $H_1: \theta = 1$
- $H_2: heta=1/2$
- $H_3: \theta = 0$
- $H_4: \theta \neq 1$

- $H_5: heta
 eq 1/2$
- $H_6: heta
 eq 0$

A coint is tossed 4 times, all ending in heads.

$$egin{aligned} f_{H_2}(x) &= rac{(1/2)^4 \mu(\{1/2\})}{\mu(\{1/2\})} = rac{1}{16} \ f_{H_5}(x) &= rac{(1)^4 \mu(\{1\}) + 0^4 \mu(\{0\})}{\mu(\{1\}) + \mu(\{0\})} = rac{\mu(\{1\})}{\mu(\{1\}) + \mu(\{0\})} \end{aligned}$$

 $\begin{array}{l} \text{Posterior expected cost of rejection } H = clP(H \text{ is true } \mid x) \\ \text{Posterior expected cost of accepting } H = lP(H \text{ is not true } \mid x) \\ \text{Bayesian reject } H \text{ if } \frac{clP(H \text{ is true } \mid x)}{lP(H \text{ is not true } \mid x)} < 1 \\ \text{i.e. } \frac{P(H \text{ is true } \mid x)}{1 - P(H \text{ is true } \mid x)} < \frac{1}{c} \\ \Leftrightarrow P(H \text{ is true } \mid x) < \frac{1}{c+1}, \\ \text{reject } H \text{ if Bayes Factor in favor of } H \text{ is less than } k \end{array}$

$$egin{aligned} f_{H_1}(x) &= rac{(1)^4 \mu(\{1\})}{\mu(\{1\})} = 1 \ f_{H_2}(x) &= rac{1}{16} \ f_{H_3}(x) &= rac{0^4 \mu(\{0\})}{\mu(\{0\})} = 0 \end{aligned}$$

Assume $\mu(\{1\})=0.01, \mu(\{1/2\})=0.98, \mu(\{0\})=0.01$

$$egin{aligned} f_{H_4}(x) &= rac{\int_{H_2 \cup H_3} f(x \mid heta) d\mu(heta)}{\mu(H_2 \cup H_3)} \ &= rac{(1/2)^4 \mu(\{1/2\}) + 0^4 \mu(\{0\}))}{\mu(\{1/2\}) + \mu(\{0\})} \ &= rac{0.98/16}{0.98 + 0.01} = 0.0619 \ f_{H_5}(x) &= rac{0^4 \mu(\{0\}) + 1^4 \mu(\{1\})}{\mu(\{0\}) + \mu(\{1\})} = rac{0.01}{0.01 + 0.01} = 0.5 \ f_{H_6}(x) &= rac{1^4 \mu(\{1\}) + (1/2)^4 \mu(\{1/2\})}{\mu(\{1\}) + \mu(\{1/2\})} = rac{0.01 + 0.98/16}{0.99} = 0.72 \ &= rac{f_{H_4}(x)}{f_{H_1}(x)} = 0.0619, \quad rac{f_{H_2}(x)}{f_{H_5}(x)} = rac{0.0625}{0.5} = 0.125 \end{aligned}$$

If we now choose $k \in (0.0619, 0.125)$, we have contradictory results.

 $\text{Assume } f_{H_3}(x) < \min(f_{H_1}(x), f_{H_2}(x))$

$$egin{aligned} f_{H_4}(x) =& rac{\int_{H_2\cup H_3} f(x\mid heta) d\mu(heta)}{\mu(H_2\cup H_3)} \ &= rac{\int_{H_2} f(x\mid heta) d\mu(heta) + \int_{H_3} f(x\mid heta) d\mu(heta)}{\mu(H_2) + \mu(H_3)} \ &= rac{\mu(H_2) \int_{H_2} rac{f(x\mid heta)}{\mu(H_2)} d\mu(heta) + \mu(H_3) \int_{H_3} rac{f(x\mid heta)}{\mu(H_3)} d\mu(heta)}{\mu(H_2) + \mu(H_3)} \ &= rac{\mu(H_2)}{\mu(H_2) + \mu(H_3)} f_{H_2}(x) + rac{\mu(H_3)}{\mu(H_2) + \mu(H_3)} f_{H_3}(x) \ &< f_{H_2}(x) \end{aligned}$$

Therefore,

$$rac{f_{H_4}(x)}{f_{H_1}(x)} < rac{f_{H_2}(x)}{f_{H_1}(x)}.$$

On the other hand,

$$egin{aligned} f_{H_5}(x) =& rac{\int_{H_1\cup H_3} f(x\mid heta) d\mu(heta)}{\mu(H_1\cup H_3)} \ &= & rac{\mu(H_1)}{\mu(H_1)+\mu(H_3)} f_{H_1}(x) + rac{\mu(H_3)}{\mu(H_1)+\mu(H_3)} f_{H_3}(x) \ &< & f_{H_1}(x) \end{aligned}$$

Therefore,

$$rac{f_{H_2}(x)}{f_{H_5}(x)} > rac{f_{H_2}(x)}{f_{H_1}(x)}$$

Default Priors for Binomial(*p*)

- Laplace: $\pi(p) = 1, 0 \le p \le 1$
- Jeffreys: $\pi(p) = p^{-rac{1}{2}}(1-p)^{-rac{1}{2}}, 0$
- Haldane: $\pi(p) = p^{-1}(1-p)^{-1}, 0$

For the prior $\pi(\sigma^2) = 1$, consider $z = \sigma$. We have $z^2 = \sigma^2 \Rightarrow \frac{d\sigma^2}{dz} = 2z \Rightarrow \pi(z) \propto z$ -- Not invariant under one-to-one transformation.

Jeffreys's Prior

Invariant under 1:1 transformation

$$egin{aligned} &|I(heta)|^{1/2} \ &I(heta) = \mathbb{E}\left[\left(rac{d\log f}{d heta}
ight)^2
ight] = \mathbb{E}\left[-rac{d^2\log f}{d heta^2}
ight] \end{aligned}$$

Suppose $\phi \stackrel{1:1}{\leftrightarrow} \theta$.

$$egin{aligned} &I(\phi) = &\mathbb{E}iggl[rac{d\log f}{d heta}iggr]^2 = &\mathbb{E}iggl[rac{d\log f}{d heta}rac{d heta}{d\phi}iggr]^2 \ &=&\mathbb{E}iggl[rac{d\log f}{d heta}iggr]^2iggl(rac{d heta}{d\phi}iggr)^2 \ &\Rightarrow |I(\phi)|^{1/2} = |I(heta)|^{1/2} \left|rac{d heta}{d\phi}
ight| \end{aligned}$$

Now, for Binomial(p),

$$f(x \mid p) = {n \choose x} p^x (1-p)^{n-x}$$

 $rac{d\log f}{dp} = rac{x}{p} - rac{n-x}{1-p}$
 $rac{d^2\log f}{dp^2} = -rac{x}{p^2} - rac{n-x}{(1-p)^2}$
 $I(p) = rac{np}{p^2} + rac{n(1-p)}{(1-p)^2} = rac{n}{p(1-p)}$
 $\pi_J(p) = p^{-1/2}(1-p)^{-1/2}$

Multiparameter

$$egin{aligned} & heta = (heta_1, \dots, heta_p) \ & \pi(heta) = |I(heta)|^{1/2} \end{aligned}$$

$$\phi = (\phi_1, \dots, \phi_p) \stackrel{1:1}{\Leftrightarrow} (heta_1, \dots, heta_p)$$

$$I(\phi) = \mathbb{E}\left[\frac{\partial \log f}{\partial \phi} \left(\frac{\partial \log f}{\partial \phi}\right)^T\right] = \mathbb{E}\left[\begin{pmatrix}\frac{\partial \log f}{\partial \phi_1}\\ \vdots\\ \frac{\partial \log f}{\partial \phi_p}\end{pmatrix} \left(\frac{\partial \log f}{\partial \phi_1} & \cdots & \frac{\partial \log f}{\partial \phi_p}\end{pmatrix}\right]$$

$$\begin{split} \frac{\partial \log f}{\partial \phi_j} &= \sum_{k=1}^p \frac{\partial \log f}{\partial \theta_k} \frac{\partial \theta_k}{\partial \phi_j} \\ &= \left(\frac{\partial \theta_1}{\partial \phi_j} \quad \cdots \quad \frac{\partial \theta_p}{\partial \phi_j} \right) \begin{pmatrix} \frac{\partial \log f}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log f}{\partial \theta_p} \end{pmatrix} \\ I(\phi) &= \mathbb{E} \left(\begin{pmatrix} \frac{\partial \theta_1}{\partial \phi_1} & \cdots & \frac{\partial \theta_p}{\partial \phi_1} \\ \vdots & \vdots \\ \frac{\partial \theta_1}{\partial \phi_p} & \cdots & \frac{\partial \theta_p}{\partial \phi_p} \end{pmatrix} \begin{pmatrix} \frac{\partial \log f}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log f}{\partial \theta_1} & \cdots & \frac{\partial \log f}{\partial \theta_p} \end{pmatrix} \begin{pmatrix} \frac{\partial \log f}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log f}{\partial \theta_p} \end{pmatrix} \begin{pmatrix} \frac{\partial \log f}{\partial \theta_1} & \cdots & \frac{\partial \log f}{\partial \theta_p} \end{pmatrix} \begin{pmatrix} \frac{\partial \theta_1}{\partial \theta_1} & \cdots & \frac{\partial \theta_1}{\partial \phi_p} \end{pmatrix} \\ &= JI(\theta)J^T \end{split}$$

So,

$$egin{aligned} |I(\phi)| &= |J| |I(heta)| |J^T| = |J|^2 |I(heta)| \ &\Rightarrow |I(\phi)|^{1/2} = |J| |I(heta)|^{1/2} \end{aligned}$$

Example:

$$\begin{split} X_1, \dots, X_n \mid \mu, \sigma^2 \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2) \\ L(\mu, \sigma) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right] \\ &\left\{ \frac{\partial \log L}{\partial \mu} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^2} = \frac{n(\bar{x} - \mu)}{\sigma^2} \\ \frac{\partial \log L}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 \\ \Rightarrow \begin{cases} \frac{\partial^2 \log L}{\partial \mu \partial \sigma} = -\frac{2n(\bar{x} - \mu)}{\sigma^3}, \\ \frac{\partial^2 \log L}{\partial \sigma^2} &= -\frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \end{cases} \end{split}$$

So,

$$egin{aligned} I(\mu,\sigma) &= egin{pmatrix} rac{n}{\sigma^2} & 0 \ 0 & rac{2n}{\sigma^2} \end{pmatrix} \ &\Rightarrow |I(\mu,\sigma)|^{1/2} \propto \sigma^{-2} \end{aligned}$$

- Jeffreys' general rule prior $|I(\theta)|^{1/2}$; here $|I(\mu,\sigma)|^{1/2} \propto \sigma^{-2}$
- Jeffreys' independence prior $\pi(\mu, \sigma) \propto \sigma^{-1}$ -- Recommended for point estimation and construction of credible sets

If
$$\pi(\mu, \sigma) \propto \sigma^{-1}$$
, consider $z = \sigma^2$. We have $\pi(\mu, z) = z^{-1/2} \frac{1}{2\sqrt{z}} \propto z^{-1}$, i.e., $\pi(\mu, \sigma^2) \propto (\sigma^2)^{-1}$.

$$egin{aligned} L(\mu,\sigma) \propto (\sigma^2)^{-n/2} \exp\left[-rac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2
ight] \ \pi(\mu,\sigma^2) \propto (\sigma^2)^{-1} \end{aligned}$$

Then,

$$\pi(\mu,\sigma^2 \mid X_1,\ldots,X_n) \propto (\sigma^2)^{-n/2-1} \exp\left[-rac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2
ight]$$

Integrating out σ^2 , we get

$$\pi(\mu) \propto \left[\sum_{i=1}^{n} (X_i - \mu)^2\right]^{-n/2}$$
$$= \left[n(\bar{X} - \mu)^2 + \sum_{i=1}^{n} (X_i - \bar{X})^2\right]^{-n/2}$$
$$\propto \left[1 + \frac{n(\bar{X} - \mu)^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}\right]^{-n/2}$$
$$= \left[1 + \frac{(\bar{X} - \mu)^2}{(n-1)(S^2/n)}\right]^{-(n-1)/2 - 1/2}$$

Posterior for μ is Student's t with location μ , scale S/\sqrt{n} , df n-1.

A Student's t-distribution is a Scale-Mixed Normal Dsitribution

$$egin{aligned} X \mid heta, \sigma^2, r \sim \mathcal{N}(heta, \sigma^2/r) \ r \sim ext{Gamma}(
u/2,
u/2) \end{aligned}$$

Then

 $X \mid heta, \sigma^2 \sim ext{Student's t}(
u(ext{df}), heta(ext{location}), \sigma(ext{scale}))$

• If $\nu = 1$, the student's t distribution becomes a Cauchy distribution.

Proof:

$$\begin{split} f(x) &= \int_0^\infty \left(\frac{r}{2\pi\sigma^2}\right)^{1/2} \exp\left[-\frac{r}{2\sigma^2}(x-\theta)^2\right] \exp\left(-\frac{\nu r}{2}\right) \frac{r^{\nu/2-1}\nu^{\nu/2}}{2^{\nu/2}\Gamma(\nu/2)} dr \\ &= (2\pi\sigma^2)^{-1/2} 2^{-\nu/2}\Gamma^{-1}(\nu/2)\nu^{\nu/2} \int_0^\infty \exp\left[-\frac{r}{2}\left[\frac{(x-\theta)^2}{\sigma^2}+\nu\right]\right] r^{(\nu-1)/2} dr \\ &= \cdots \\ &= \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}} \left[1 + \frac{(x-\theta)^2}{\nu\sigma^2}\right]^{-\frac{\nu+1}{2}} \end{split}$$

A Double Exponential Distribution is a Scale-Mixed Normal Distribution

$$X \mid heta, \sigma^2, r \sim \mathcal{N}(heta, \sigma^2/r) \ f(r) = \exp(-rac{1}{2r})rac{1}{2r^2} = \exp(-rac{1}{2r})rac{1}{2}r^{-1-1} \sim \mathrm{InvGamma}(1, 1/2)$$

Then

$$f(x) = rac{1}{2\sigma} \mathrm{exp}\left[-rac{|x- heta|}{\sigma^2}
ight]$$

Proof:

$$\begin{split} f(x) &= \int_0^\infty \left(\frac{r}{2\pi\sigma^2}\right)^{1/2} \exp\left[-\frac{r(x-\theta)^2}{2\sigma^2}\right] \exp(-\frac{1}{2r})\frac{dr}{r^2} \\ &= \frac{1}{2\sigma} \int_0^\infty \frac{1}{(2\pi r^3)^{1/2}} \exp\left[-\frac{1}{2r} \left[\frac{r(x-\theta)^2}{\sigma^2} + 1\right]\right] dr \\ &= \cdots \\ &= \frac{1}{2\sigma} \exp\left[-\frac{1}{\sigma}|x-\theta|\right] \end{split}$$

Inverse Gaussian

$$egin{aligned} f(z) &= rac{1}{(2\pi z^3)^{1/2}} ext{exp}\left[-rac{1}{2z}igg(rac{z}{\mu}-1igg)^2
ight] \ &\mathbb{E}Z &= \mu \end{aligned}$$

Logistic pdf

$$egin{aligned} f(x) &= rac{\exp(x)}{(1+\exp(x))^2} = \exp(-x)[1+\exp(-x)]^{-2} \ &= \exp(-x)[1-2\exp(-x)+3\exp(-2x)-4\exp(-3x)+\cdots] \ &= \sum_{r=1}^\infty (-1)^{r-1}r\exp(-rx) \ &= rac{f(x)=f(-x)}{2} 2\sum_{r=1}^\infty (-1)^{r-1}rac{r}{2}\exp(-r|x|) - ext{Mixture of DE} \end{aligned}$$

[More on the lecture notes in paper]

Asymptotic Normality of the Posterior

Heuristics

$$\begin{split} f(x \mid \theta) & \pi(\theta) \\ \pi(\theta \mid x) \\ \log \pi(\theta \mid x_n) = \log \pi(\hat{\theta}_n \mid x_n) + (\theta - \hat{\theta}_n) \frac{d \log \pi(\theta \mid x_n)}{d\theta} \mid_{\theta = \hat{\theta}_n} + \frac{1}{2} (\theta - \hat{\theta}_n)^2 \frac{d^2 \log \pi(\theta \mid x_n)}{d\theta^2} \mid_{\theta = \hat{\theta}_n} \\ = \log \pi(\tilde{\theta}_n \mid x_n) + \frac{1}{2} (\theta - \tilde{\theta}_n)^2 (-\tilde{I}_n(\tilde{\theta}_n)) \\ \Rightarrow \pi(\theta \mid x) = \pi(\tilde{\theta}_n \mid x_n) \exp[-\frac{1}{2} (\theta - \tilde{\theta}_n)^2 \tilde{I}_n(\tilde{\theta}_n)] \\ & \propto \exp[-\frac{1}{2} (\theta - \tilde{\theta}_n)^2 \tilde{I}_n(\tilde{\theta}_n)] \\ & \sim \mathcal{N}(\tilde{\theta}_n, \tilde{I}_n^{-1}(\tilde{\theta}_n)) \end{split}$$

where $\hat{\theta}_n$ is the MLE of θ , $\tilde{\theta}_n$ is very close to $\hat{\theta}_n$.

Berstein von Mises proves asymptotic $\mathcal{N}(\hat{\theta}_n, \hat{I}_n^{-1}(\hat{\theta}_n))$, where $\hat{I}_n^{-1}(\hat{\theta}_n) = -\frac{d^2 \log f(x_n|\theta)}{d\theta^2}|_{\theta=\hat{\theta}_n}$ for normalized posterior.

Total Variation Between Two Densities f_1 and f_2

$$TV(f_1,f_2) = \sup_A \left|\int_A (f_1-f_2) d\mu
ight|$$

Recall

$$TV(f_1,f_2) = rac{1}{2}\int |f_1-f_2| d\mu.$$

Renyi Divergence

$$D_lpha(f_1,f_2)=rac{1-\int f_1^lpha f_2^{1-lpha}d\mu}{lpha(1-lpha)}, \quad 0$$

$$egin{aligned} \lim_{lpha o 0} D_lpha(f_1,f_2) &= \lim_{lpha o 0} rac{1 - \int \left(rac{f_1}{f_2}
ight)^lpha f_2 d\mu}{lpha(1-lpha)} \ &= \lim_{lpha o 0} rac{-\int \left(rac{f_1}{f_2}
ight)^lpha(\lograc{f_1}{f_2})f_2 d\mu}{1-2lpha} \ &= \int (\lograc{f_2}{f_1})f_2 d\mu \ &= KL(f_2,f_1), \end{aligned}$$
 $\lim_{lpha o 1} D(f_1,f_2) &= \lim_{lpha o 0} rac{-\int \left(rac{f_1}{f_2}
ight)^lpha(\lograc{f_1}{f_2})f_2 d\mu}{1-2lpha} \ &= \int \log(rac{f_1}{f_2})f_1 d\mu \ &= KL(f_1,f_2). \end{aligned}$

Bhattachwyga-Hellinger

$$egin{aligned} H(f_1,f_2) &= \left[\int \left(f_1^{1/2} - f_2^{1/2}
ight)^2 d\mu
ight]^{1/2} \ D_{1/2}(f_1,f_2) =& 4 \left[1 - \int f_1^{1/2} f_2^{1/2} d\mu
ight] \ &= 2 \left[2 - 2 \int f_1^{1/2} f_2^{1/2} d\mu
ight] \ &= 2 \left[\int f_1 d\mu + \int f_2 d\mu - \int f_1^{1/2} f_2^{1/2} d\mu
ight] \ &= 2 \int \left(f_1^{1/2} - f_2^{1/2}
ight)^2 d\mu \ &= 2 H^2(f_1,f_2) \end{aligned}$$

$$egin{aligned} D_{-1}(f_1,f_2) =& rac{1-\int f_1^{-1}f_2^2d\mu}{(-1)(2)} \ &=& rac{1}{2}\left[\int rac{f_2^2}{f_1}d\mu -1
ight] \ &=& rac{1}{2}\int rac{(f_2-f_1)^2}{f_1}d\mu & ext{--Chi-square divergence} \end{aligned}$$

$$\begin{split} H^{2}(f_{1},f_{2}) =& 2 \left[1 - \int f_{1}^{1/2} f_{2}^{1/2} d\mu \right] \\ =& 2 \left[1 - \int (\frac{f_{1}}{f_{2}})^{1/2} f_{2} d\mu \right] \\ \leq& 2 \left[1 - \int [1 + \log(\frac{f_{1}}{f_{2}})^{1/2}] f_{2} d\mu \right] \\ =& 2 \left[1 - \int f_{2} d\mu + \frac{1}{2} \int \log(\frac{f_{2}}{f_{1}}) f_{2} d\mu \right] \\ =& \int \log(\frac{f_{2}}{f_{1}}) f_{2} d\mu \\ =& KL(f_{2},f_{1}) \\ KL(f_{1},f_{2}) =& \int (\log\frac{f_{1}}{f_{2}}) f_{1} d\mu \\ \leq& \int \left(\frac{f_{1}}{f_{2}} - 1 \right) f_{2} d\mu \\ =& \int \frac{f_{1}^{2}}{f_{2}} d\mu - 1 \\ =& \int \frac{(f_{1} - f_{2})^{2}}{f_{2}} d\mu \end{split}$$

Scheffe's Theorem

Let $\{p_n, n \geq 1\}$ be a sequence of pdf's such that $p_n(x) o p(x)$ pointwise. Then $\int |p_n - p| d\mu o 0.$ proof:

$$egin{aligned} &\int |p_n-p| d\mu = \int [p_n+p-2\min(p_n,p)] d\mu \ =& 2[1-\int\min(p_n,p)d\mu] \end{aligned}$$

Because $\min(p_n,p) o p$ and $\min(p_n,p) \le p$ and $\int p d\mu = 1$, apply DCT to get

$$\int |p_n-p|d\mu
ightarrow 2(1-1)=0.$$

<u>**Result:**</u> $TV(p_n,p)
ightarrow 0 \Rightarrow D_{\alpha}(p_n,p)
ightarrow 0$ for all 0 < lpha < 1.

proof:

$$egin{aligned} D_lpha(p_n,p) =& rac{1-\int p_n^lpha p^{1-lpha} d\mu}{lpha(1-lpha)} \ &\leq & rac{1-\int \min(p_n,p) d\mu}{lpha(1-lpha)} \ &\leq & rac{\int |p_n-p| d\mu}{2lpha(1-lpha)} o 0 \end{aligned}$$

<u>**Result:**</u> $D_{1/2}(p_n,p)
ightarrow 0 \Rightarrow TV(p_n,p)
ightarrow 0.$

proof:

$$\begin{split} \int |p_n - p| d\mu &= \int \left| (p_n^{1/2} + p^{1/2}) (p_n^{1/2} - p^{1/2}) \right| d\mu \\ &\leq \left[\int (p_n^{1/2} + p^{1/2})^2 d\mu \right]^{1/2} \left[\int (p_n^{1/2} - p^{1/2})^2 d\mu \right]^{1/2} \\ &= \left[\int (p_n + p + 2p_n^{1/2} p^{1/2}) d\mu \right]^{1/2} H(p_n, p) \\ &= \left[2 \left(1 + \int p_n^{1/2} p^{1/2} d\mu \right) \right]^{1/2} H(p_n, p) \\ &= \left[4 - 2(1 - \int p_n^{1/2} p^{1/2} d\mu) \right]^{1/2} H(p_n, p) \\ &= \left[4 - H^2(p_n, p) \right]^{1/2} H(p_n, p) \to 0 \end{split}$$

To put everything in a nutshell:

 $\underline{ \textbf{Final Result:}} \ D_{1/2}(p_n,p) \to 0 \Rightarrow TV(p_n,p) \to 0 \Rightarrow D_{\alpha}(p_n,p) \to 0 \ \textbf{for all} \ 0 < \alpha < 1.$

Bernstein-von Mises Theorem

See the lecture notes in paper