Bayesian Interval Estimation

Practical Question

A researcher collects standardized exam score differences from a random sample of students. The data are obtained from the R built-in dataset `sleep`, using the variable `extra`. The dataset is as given below

Data

Assume the Bayesian model:

- Likelihood:

$$X_i \mid \theta \sim \mathcal{N}(\theta, \sigma^2)$$

- Prior:

$$\theta \sim \mathcal{N}(0, 100).$$

(a) Derive the posterior distribution of $\theta \mid x$.

(b) Construct a 95% equal-tailed credible interval for θ.

(c) Interpret the credible interval.

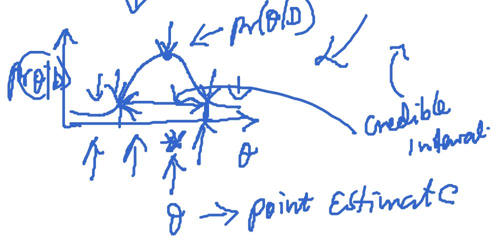
$Pr(\theta)$

1st infor Prior about θ

② Current Data Likelihood function

$Pr(D|\theta)$

future. θ Estimate incidence of malaria using bayesian framework



Prior → Normal $(\underset{\mu_0}{0}, \underset{\tau^2}{100})$

Likelihood → Normal (θ, σ^2)

Posterior → Normal.

θ/x

$$\theta/x \sim N(\mu_n, \delta_n^2)$$

$$\delta_n^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1} \qquad \mu_n = \delta_n^2\left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\tau^2}\right)$$

ID	Group	Extra	$(x_i-\bar{x})^2$
1	1	0.7	0.7056
2	1	-1.6	
3	1	-0.2	
4	1	-1.2	
5	1	-0.1	
6	1	3.4	
7	1	3.7	
8	1	0.8	
9	1	0.0	
10	1	2.0	
1	2	1.9	
2	2	0.8	
3	2	1.1	
4	2	0.1	
5	2	-0.1	
6	2	4.4	
7	2	5.5	
8	2	1.6	
9	2	4.6	
10	2	3.4	3.4596

$$\mu_0 = 0, \quad \tau^2 = 100, \qquad \bar{x} = \frac{0.7 + -1.6 + \cdots + 3.4}{20} = 1.54$$

$$n = 20$$

$$\sigma^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 4.54$$

$$\delta_n^2 = \left(\frac{20}{4.54} + \frac{1}{100}\right)^{-1} = 0.2265$$

$$\mu_n = 0.2265\left(\frac{20 \times 1.54}{4.54} + \frac{0}{100}\right) = 1.54$$

sum $\Rightarrow \sum(x_i - \bar{x})^2 \Rightarrow$

$$\theta/x \sim N(1.54, 0.2265).$$

(b) 95% equal tail interval

Interval $[\mu_n - 1.96\,\sigma_n, \mu + 1.96\,\sigma_n]$.

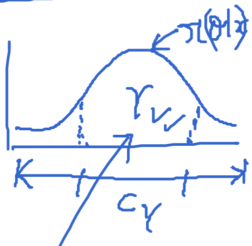
$$\sigma_n = \sqrt{\sigma_n^2} = \sqrt{0.2265} = 0.476.$$

$$CI = [1.54 - 1.96(0.476),\ 1.54 + 1.96(0.476)].$$

$$= [0.61,\ 2.47]$$

(c) Interpretation: There is 95% posterior probability that the true mean lies between 0.61 and 2.47 given the observed data & prior infor.

Highest posterior Density Interval (HPD) / Highest Density interval (HDI)



Let $\pi(\theta/x)$ — be the posterior density of θ

A set C_γ is called γ-level HPD credible set if

$$Pr(\theta \in C_\gamma \mid x) = \int_{\theta \in C_\gamma} \pi(\theta/x)\, d\theta = \gamma.$$

the highest density condition:

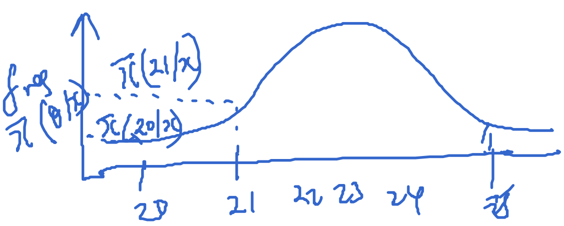
for an $\theta_1 \in C_\gamma$ and $\theta_2 \notin C_\gamma$ $\quad \int_{\theta_1} \quad \int \pi(\theta/x) + \int \pi(\theta/x)$, $\theta_1 \quad \theta_2$ $\quad \Downarrow \int_\theta \pi(\theta/x)$

$$\pi(\theta_1/x) \geq \pi(\theta_2/x) =$$

$$\theta = \{20, 21, 22, 23, 24, 25\} \qquad C_\gamma = \{21, 22, 23, 24\}$$

θ_1

θ_2



$$\pi(21/x) \geq \pi(21/x)$$

$$\pi(21/x) + \pi(22/x) + \pi(23/x) + \pi(24/x) = \gamma \leftarrow \text{HPD}$$

90

$$C_\gamma = [21 \quad 24] \leftarrow \text{HPD}$$

$$C_{90} = [21, 24]$$

~~The~~ Equivalent definition:

HPD $\quad C_\gamma = \{\theta : \pi(\theta|x) \geq k_\gamma\},$

where k is the constant chosen s.t

$$\int_{\pi(\theta|x) \geq k_\gamma} \pi(\theta|x)\, d\theta = \gamma .$$

HPD consist of all parameter θ values whose posterior density $\pi(\theta|x)$ is above certain cut off k and the cut off is chosen so that exactly γ of ~~posterior~~ mass is included.

$\theta = [20, 21, 22, 23, 24, 25]$ $\quad K = 0.05$

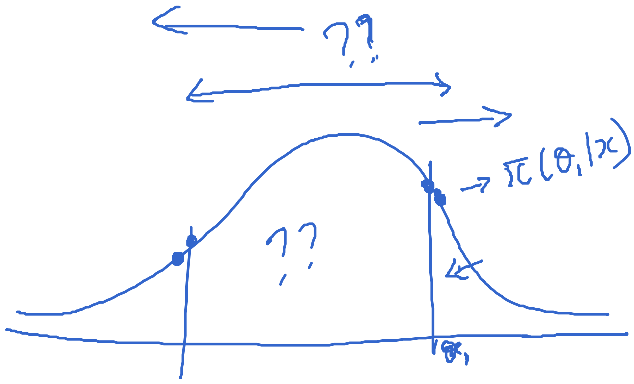
$\pi(\theta/x) = [\pi(20/x), \pi(21/x) \; \cdots \; \pi(24/x), \pi(25/x)]$

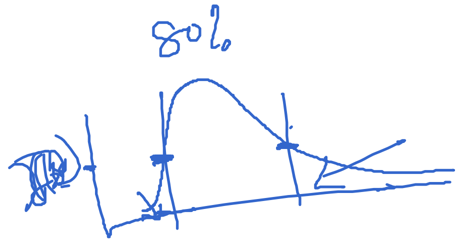
0.01 $\quad$ 0.3 $\quad$ 0.2 $\quad$ 0.02

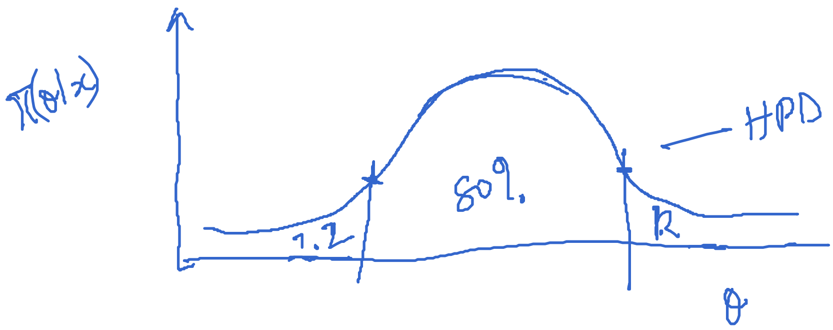
$\pi(20/x) < K$ $\quad$ γ proportion $\quad$ $\pi(25/x) < K$

$\gamma = \frac{4^2}{6_3} \times \frac{33}{100} = 66\% = [$

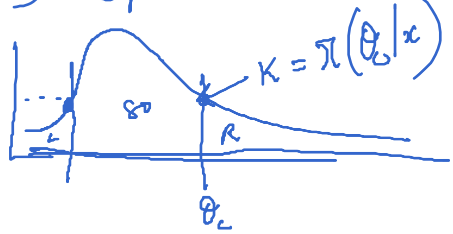
66% Credible interval $= [21, 24]$.



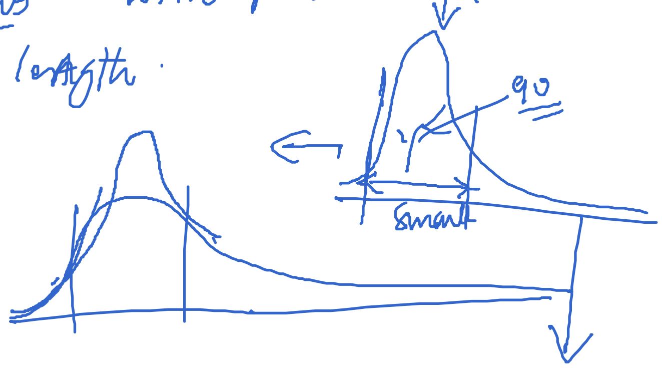




$\rightarrow$ Equal tail



Among all credible intervals with probability γ the HPD interval has the smallest length.

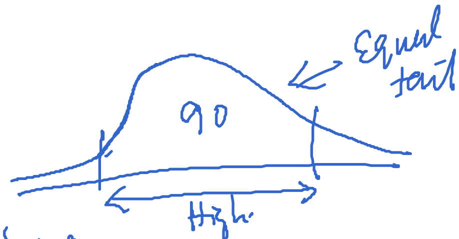


(a) $\theta/x \sim N(\mu, \sigma^2)$ $\gamma = 0.95$

HPD = Equal tail

$C_{0.95} = [\mu - 1.96\sigma, \mu + 1.96\sigma]$

95% CI

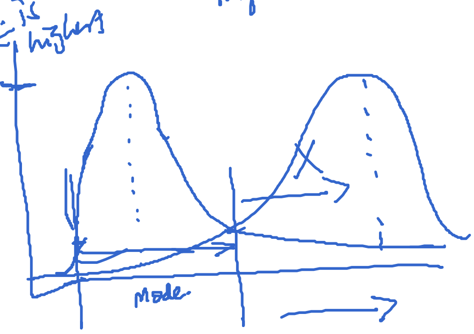


(b) when posterior distribution is not normal (Typical)

NOT normal: Gamma distn | Beta | Lognormal

(1) Equal tail is not shortest

(2) HPD moves towards the mode



HPD — contains the mode, it always contain (MAP Estimator)

$\hat{\theta}_{MAP} = \arg\max_{\theta} (\pi(\theta/x))$

(3) HPD will always here an, equal tails

Example :-

Suppose the posterior distn of θ is

$$\theta/x \sim \text{gamma}(\alpha = 4, \beta = 1)$$

cdf of gamma

$$\pi(\theta/x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \quad \theta > 0$$

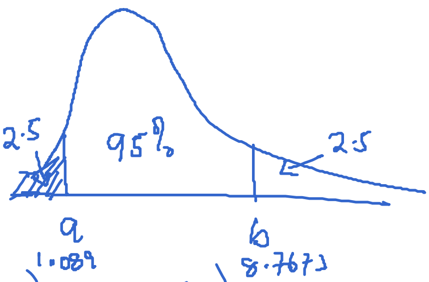
$$= \frac{1}{\Gamma(4)} \theta^{3} e^{-\theta} \quad \theta > 0$$

$$\Gamma(4) = 3\Gamma(3) = 3\times 2\,\Gamma(2) = 3\times 2\times 1\,\Gamma(1) = 6 \quad (\Gamma(1)=1)$$

$$= \frac{1}{6}\theta^3 e^{-\theta}$$

$$F(\theta) = pr(\theta \le \theta/x) = \int_0^\theta \pi(t/x)\,dt.$$

Equal tail 95%



95% CI → [a, b]

s.t. $pr(\theta < a) = 0.025$, $pr(\theta > b) = 0.025$

$F(a) = 0.025$

$\Downarrow$ $F(b) = pr(\theta < b) = 0.975$

$$[a, b] = [\theta_{0.025}, \theta_{0.975}] = [1.089, 8.7673]$$

$C_\gamma = [a, b]$

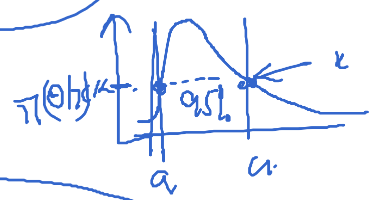
$\theta \in C_\gamma$

$\theta \in C_{0.95}$

HPD 95% Credible Interval

$\gamma = 95\%$

$$\int_a^b \pi(\theta/x)\,d\theta = 0.95$$



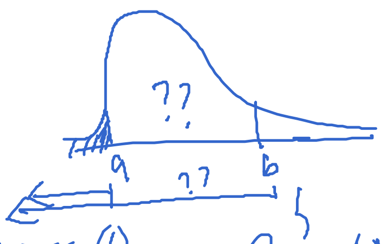
$$k = \pi(a|x) = \pi(b|x)$$

$$\pi(\theta/x) = \frac{1}{6}\theta^3 e^{-\theta}$$

$$\frac{1}{6}a^3 e^{-a} = \frac{1}{6}b^3 e^{-b}$$

$$a^3 e^{-a} = b^3 e^{-b}$$

$$3\ln a - a = 3\ln b - b \;\ldots\ldots (i)$$



$$\int_a^b \pi(\theta|x)\,d\theta.$$

Also

$$pr(a < \theta < b \,|\, x) = 0.95$$

$$F(b) - F(a) = 0.95 \;\ldots (ii)$$

$a = 0.7125$

$b = 7.9458$ ← Short dist

$[a,b] = [0.7125,\ 7.9483]$ ← HPD

$= [1.089,\ 8.7673]$ ← Equal tail.

long distance

Example 2 Monthly snakebite cases recorded in a county over one year are:

$$y = (3, 5, 4, 6, 2, 7, 5, 4, 6, 5, 3, 4).$$

Assume the Bayesian model:

- Likelihood:
$$Y_i \mid \theta \sim \text{Poisson}(\theta).$$
- Prior:
$$\theta \sim \text{Gamma}(\overset{\alpha}{1}, \overset{\beta}{1}),$$
where the second parameter is the rate.

(a) Derive the posterior distribution of $\theta \mid y$.

(b) Compute the 95% equal-tailed credible interval.

(c) Obtain the 95% HPD credible interval.

(d) Compare the two intervals.

Solution:

Likelihood $\quad Y \mid \theta \sim \text{pois}(\theta)$

$$L(y \mid \theta) = \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!}$$

$$= \frac{\theta^{\sum y_i} e^{-n\theta}}{\prod y_i!} \Rightarrow \frac{1}{\prod y_i!}\ \theta^{\sum y_i} e^{-\theta}$$

$$L(y \mid \theta) \propto \boxed{\theta^{\sum y_i} e^{-n\theta}}$$

Prior
$$\pi(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$$

$$\pi(\theta) \propto \boxed{\theta^{\alpha-1} e^{-\beta\theta}}$$

$$\pi(\theta \mid y) \propto \underbrace{\theta^{\overbrace{\alpha+\sum y_i}^{\alpha^*}-1}\, e^{-\overbrace{(\beta+n)}^{\beta^*}\theta}}_{\theta^{\alpha_*-1}\, e^{-\beta_*\theta}}$$

$\sum y_i = 54 \quad n = 12$

$\alpha_* = 1 + 54$

$\beta_* = 1 + 12 = 13$

$\pi(\theta | y) \sim \text{Gamma}(55, 13)$

(b) 95% Equal tails

$(a, b) = [\theta_{0.025}, \theta_{0.975}]$

$= (3.09, 5.88) \longleftarrow$

(c) $pr(a \le \theta \le b)$

$HPD_{0.95} = [3.21, 5.72] \longleftarrow$ shorter length