Lecture 11/3/2019

Slides 020 – single ...

1 – 20

θ parameter of a model ($\theta \in \Theta$)

$p(y|\theta)$ likelihood

$\pi(\theta)$ prior on θ

$$\left.\begin{array}{l} p(y|\theta) \\ \pi(\theta) \end{array}\right\} \quad p(y,\theta) = \pi(\theta)\, p(y|\theta)$$

$$p(y) = \int_{\Theta} p(y,\theta)\, d\theta = \underbrace{\int_{\Theta} p(y|\theta)\, \pi(\theta)\, d\theta}_{\text{average over } \theta}$$

model for y

$y_1, \dots, y_n$ $\quad y_i \underset{\text{IID}}{\sim} p(x \mid \theta)$ usual assumption

$y_1, \dots, y_n$ indep. i.i.d. conditional on θ

y_{n+1} ?

predict based on $y_1, \dots, y_n$

this would be meaningless if $y_1, \dots, y_n, y_{n+1}$ were indep.

$$p(y) = p(y_1, \ldots, y_n) = \int_{\Theta} p(y_1, \ldots, y_n \mid \theta)\, \pi(\theta)\, d\theta$$

$$= \int_{\Theta} \prod_{i=1}^{n} p(y_i \mid \theta)\, \pi(\theta)\, d\theta$$

$\boxed{\text{EXCHANGEABILITY}}$

The order of observations does not count.

$$= \int_{\Theta} p(y_{i_1}, \ldots, y_{i_n} \mid \theta)\, \pi(\theta)\, d\theta$$

$$= p(y_{i_1}, \ldots, y_{i_n})$$

$i_1, \ldots, i_n$ permutation of $1, \ldots, n$

If $y_i = \begin{cases} 0 \\ 1 \end{cases}$ then if

$$P(Y_1 = y_1 \wedge \dots \wedge Y_n = y_n) = P(Y_{i_1} = y_{i_1} \wedge \dots \wedge Y_{i_n} = y_{i_n})$$

then there exists a r.v. $\theta \in [0,1]$ such that

$Y_1, \dots, Y_n \mid \theta$ are IID with $P(Y_i = 1 \mid \theta) = \theta$

De Finetti theorem (see also slide 9)

$$\pi(\theta | y) = \frac{\pi(\theta)\, p(y|\theta)}{p(y)}$$

$p(y)$ can be derived analytically only for simple problems

$$\boxed{\pi(\theta | y) \propto \pi(\theta)\, p(y|\theta)}$$

$\tilde{y}$ potentially observable

y observed on a new stat. unit

Behaves analogously to $y_1, \dots, y_n$

If $y_1, \dots, y_n \sim p(y \mid \vartheta)$

$y_1, \dots, y_n \mid \vartheta$ independent

then $\tilde{y} \sim p(\tilde{y} \mid \vartheta)$

$\tilde{y}$ indep. of $y_1, \dots, y_n$

PRIOR PREDICTIVE DISTRIBUTION

$$p(\tilde{y}) = \int_{\Theta} p(\tilde{y} \mid \theta)\, \pi(\theta)\, d\theta$$

$$= \int_{\Theta} p(\tilde{y}, \theta)\, d\theta$$

POSTERIOR PREDICTIVE DISTRIBUTION

$$p(\tilde{y} \mid y) = \int_{\Theta} p(\tilde{y} \mid \theta)\, \pi(\theta \mid y)\, d\theta$$

$(y_1, \dots, y_n)$

$$= \int_{\Theta} p(\tilde{y}, y, \theta)\, d\theta$$

$$= \int_{\Theta} \underbrace{p(\tilde{y} \mid y, \theta)}_{= p(\tilde{y} \mid \theta)}\, \pi(\theta \mid y)\, d\theta$$

because of independence

Husband not affected

Data: $y_i = \begin{cases} 0 & \text{otherwise} \\ 1 & \text{if } i\text{-th son is affected} \end{cases}$

y_1, y_2 indep. cond on θ

$P(y_i = 1 \mid \theta = 1) = 0.5$

$D = (y_1 = 0 \wedge y_2 = 0)$

$p(y \mid \theta) \to P(D \mid \theta)$

$P(D \mid \theta = 0) = 1$

$P(D \mid \theta = 1) = \frac{1}{4}$

$\pi(\theta \mid y)|$

$$P(\theta = 1 \mid D) = \frac{P(D \mid \theta = 1)\, P(\theta = 1)}{P(D \mid \theta = 1)\, P(\theta = 1) + P(D \mid \theta = 0)\, P(\theta = 0)}$$

$$= \frac{0.25 \times 0.5}{0.25 \times 0.5 + 1 \times 0.5} = 0.20$$

$$\frac{P(\theta = 1 \mid D)}{P(\theta = 0 \mid D)} = \frac{P(D \mid \theta = 1)}{P(D \mid \theta = 0)} \; \frac{P(\theta = 1)}{P(\theta = 0)}$$

POSTERIOR ODDS $\theta = 1$	LIKELIHOOD RATIO	PRIOR ODDS FOR $\theta = 1$
$\frac{1}{4}$	$\frac{1}{4}$	1

$$P(Y_1 = 1) = P(Y_1 = 1 \mid \theta = 0)\, P(\theta = 0) + P(Y_1 = 1 \mid \theta = 1)\, P(\theta = 1)$$

$$= 0 \times 0.5 + 0.5 \times 0.5 = 0.25$$

Consider a third son Y_3

$$P(Y_3 = 1) = \underline{0.25}$$ PRIOR PRED. DISTRIB.

$$P(Y_3 = 1 \mid D) = P(Y_3 = 1 \mid \cancel{D}, \underline{\theta = 0})\, P(\theta = 0 \mid D) + P(Y_3 = 1 \mid \cancel{D}, \underline{\theta = 1})\, P(\theta = 1 \mid D)$$

POSTERIOR PREDICTIVE

$$= 0 \times 0.8 + 0.5 \times 0.2 = \underline{0.1}$$

$$y_1, y_2 \xrightarrow{\text{tell us about}} \theta \longrightarrow y_3$$

→ leads to a dependence of y_3 on y_1, y_2

$$p(y_3 \mid y_1, y_2) \neq p(y_3)$$

new observation $Y_3 = 0$

$$D' = D \wedge Y_3 = 0 \quad = \quad Y_1 = 0 \wedge Y_2 = 0 \wedge Y_3 = 0$$

$$P(\theta = 1 \mid D')$$

→ start from prior $P(\theta = 1) = 0.5$ and update it based on $P(Y_1 = Y_2 = Y_3 = 0 \mid \theta)$

→ start from $P(\theta = 1 \mid Y_1 = Y_2 = 0)$ and update it based on $P(Y_3 = 0 \mid \theta)$

either way we obtain the same result $P(\theta = 1 \mid Y_1 = Y_2 = Y_3 = 0)$

Estimating a probability

$y_1, \ldots, y_n \qquad y_i \in \{0, 1\}$

$p(y_i = 1 \mid \theta) = \theta$

$y_1, \ldots, y_n \mid \theta$ independent

$\Rightarrow$ $y_1, \ldots, y_n$ are EXCHANGEABLE

$$p(y_1, \ldots, y_n) = p(y_{i_1}, \ldots, y_{i_n})$$

sufficies to know $\underline{y = \sum_{i=1}^{n} y_i}$

$$p(y|\theta) = \binom{n}{y} \theta^{y} (1-\theta)^{n-y}$$ LIKELIHOOD

$$y|\theta \sim \text{Binomial}(n, \theta)$$

$$\left[\, p(y|\theta) = \text{Binomial}(n, \theta) \,\right]$$

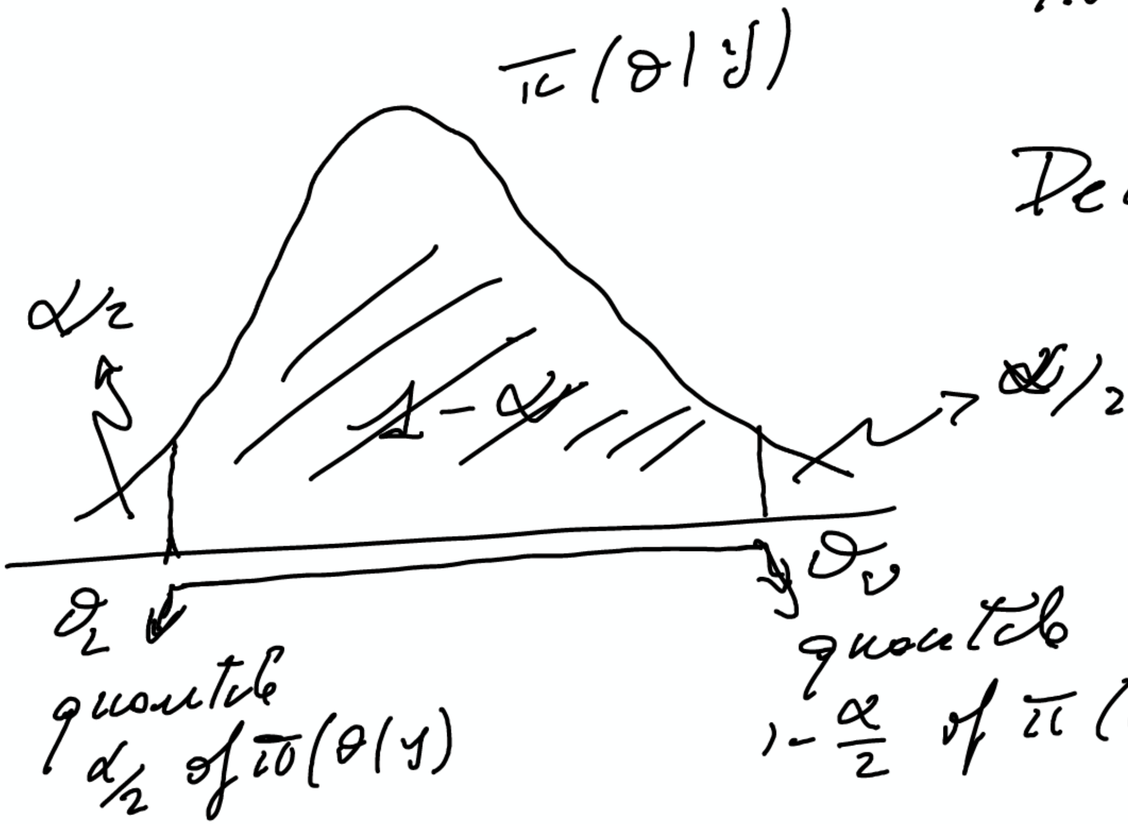
$\pi(\theta) = 1$ on $[0,1]$ uniform on 0, 1

$$p(\theta, y) = \binom{n}{y} \theta^{y} (1-\theta)^{n-y}$$

$$\pi(\theta|y) \propto \theta^{y} (1-\theta)^{n-y} \sim \text{Beta}(y+1, n-y+1)$$

INTERVAL ESTIMATE

Decide a probab. $1-\alpha$

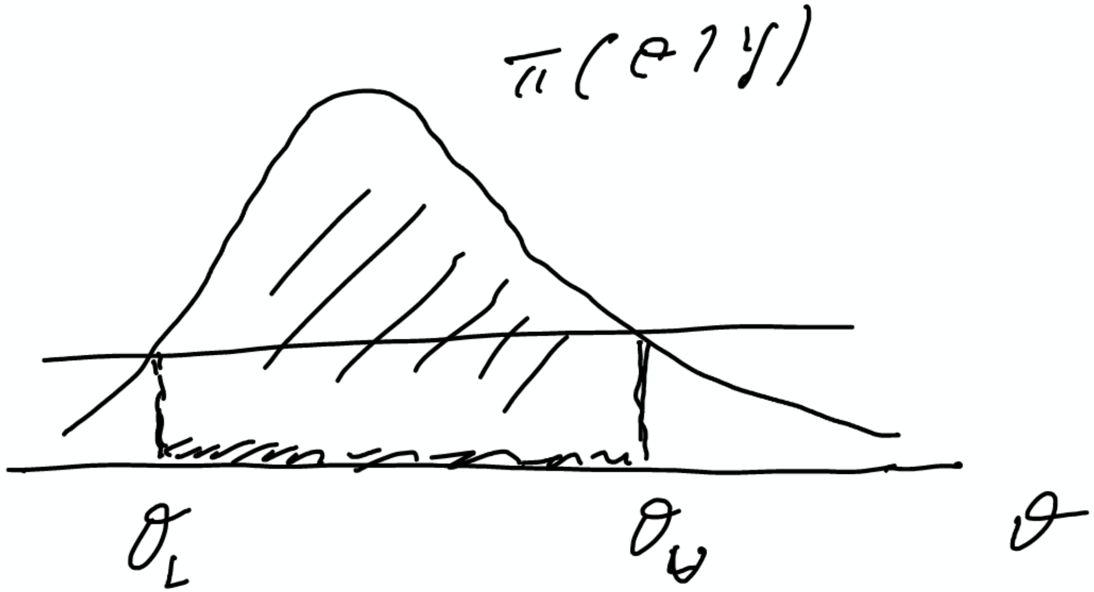


CENTRAL CREDIBILITY INTERVAL OF PROB. $1-\alpha$

$$P(\theta_L \le \theta \le \theta_U \mid y) = 1-\alpha$$

Conf. interv. $[\theta_L, \theta_U]$ of prob. $1-\alpha$

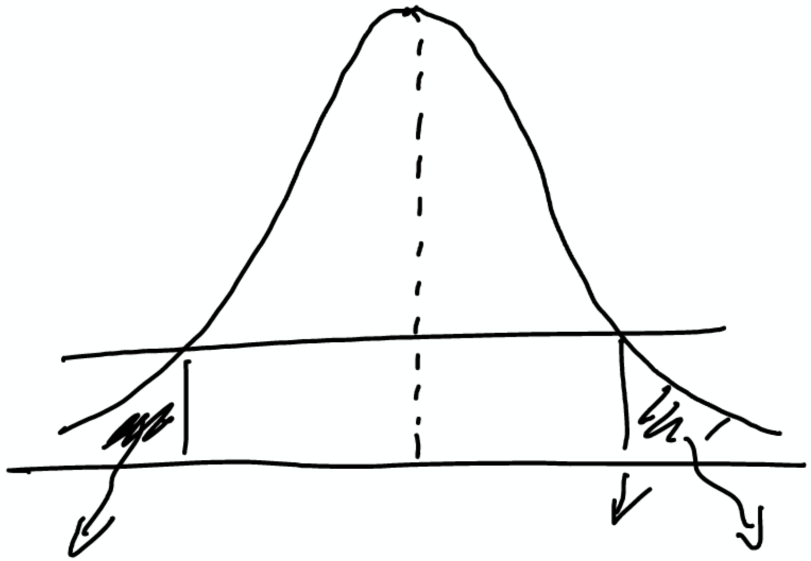
θ has prob. $1-\alpha$ ~~of~~ being in $[\theta_L, \theta_U]$



95%

HIGH POSTERIOR DENSITY INTERVAL OF PROB.

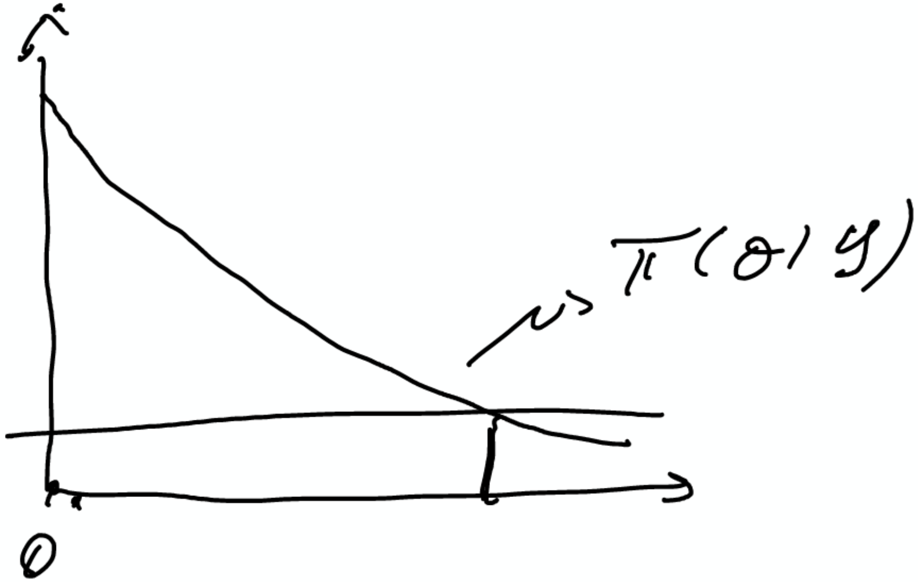
$$\int_{\theta_L}^{\theta_U} \pi(\theta|y)\, d\theta$$

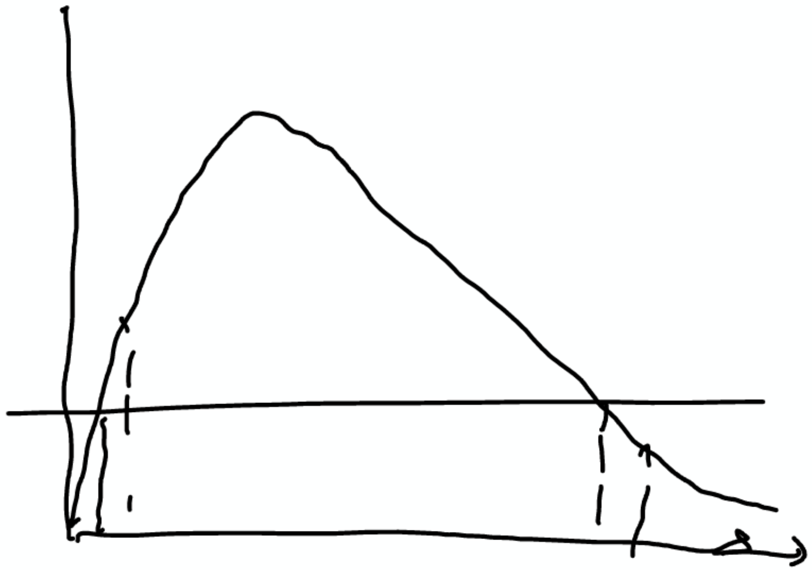


unimodal

symmetric

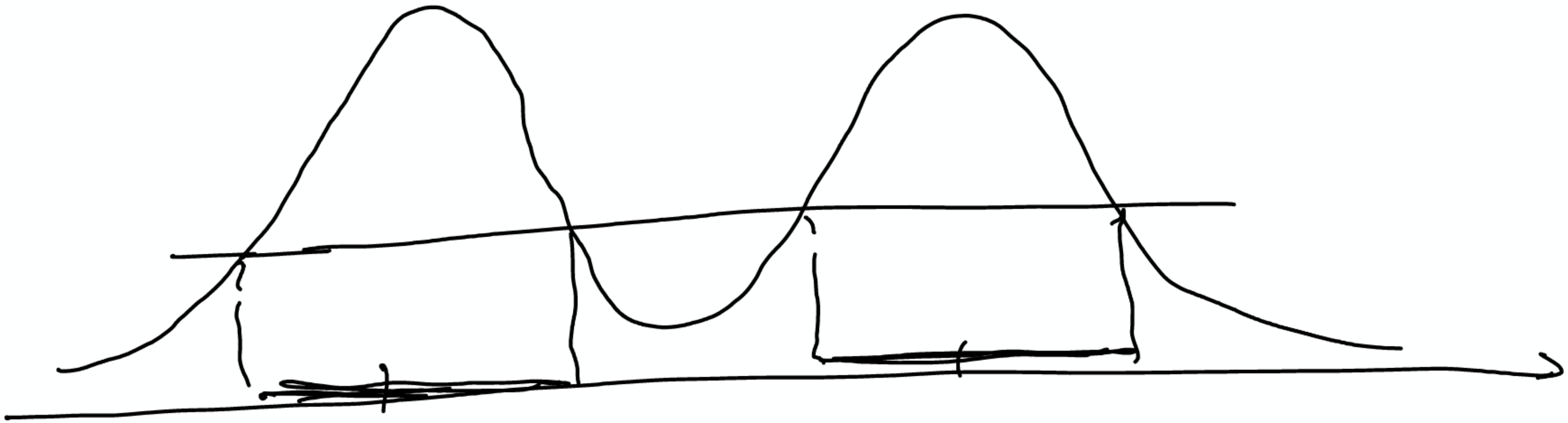
HPD coincides with central interval





HPD: $[0, (1-\alpha)\text{-quantile of } \pi(\theta|y)]$

Central interval



High posterior density region is not an interval (is the union of the two intervals)

$$\pi(\theta|y) \propto \theta^{y}(1-\theta)^{n-y} \; \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$\theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}$$

$$\boxed{\mathrm{Beta}(\alpha+y, \beta+n-y)}$$

CONJUGACY

$$E(\theta|y) = \frac{\alpha+y}{\alpha+y+\beta+n-y}$$

$$=$$

$$\theta \sim \mathrm{Beta}(\alpha, \beta) \qquad E(\theta) = \frac{\alpha}{\alpha+\beta}$$

$$\theta \mid y \sim \mathrm{Beta}(\alpha + y,\ \beta + n - y)$$

$$\underline{\underline{E(\theta \mid y)}} = \frac{\alpha + y}{\alpha + y + \beta + n - y} = \frac{\alpha + y}{\alpha + \beta + n}$$

$$= \frac{\alpha+\beta}{\alpha+\beta+n} \underbrace{\frac{\alpha}{\alpha+\beta}}_{\underline{E(\theta)}} + \frac{n}{\alpha+\beta+n} \underbrace{\frac{y}{n}}_{\text{SAMPLE PROP}}$$