this time: Bernoulli/Beta

next time: Poisson/Gamma

(same readings)

AMS 206

8 Feb 18

AMI case study ①

$$\left\{ \begin{array}{l} (\theta \mid \mathcal{B}) \sim p(\theta \mid \mathcal{B}) \\ \boxed{(Y_i \mid \theta\, \mathcal{B}) \overset{IID}{\sim} \text{Bernoulli}(\theta)} \\ \quad (i = 1, \dots, n) \end{array} \right.$$

sampling dist./likelihood specification came uniquely from de Finetti's theorem

how specify prior information about θ?

Judgments:

wrt prior

- $p(\theta \mid \mathcal{B})$ should be centered (mean) at 0.15
- $p(\theta \mid \mathcal{B})$ should have most (0.95) of its prob. between 0.05 and 0.30

(these judgments come from my personal context)

(omit $\mathcal{B}$ for simplicity)

Assumptions: ?

Bernoulli likelihood:

$\ell(\theta \mid s) =$

$\ell(\theta \mid y) = c\, \theta^{s} (1-\theta)^{n-s}$

$y = (y_1, \dots, y_n)$

$s = \sum_{i=1}^{n} y_i$

if take $p(\theta) = c\, \theta^{\alpha-1} (1-\theta)^{\beta-1}$ $(\alpha, \beta > 0)$

something nice happens:

Bayes's Thm.

$$p(\theta \mid y) = c \; p(\theta) \, \ell(\theta \mid y)$$

$$p(\theta \mid s) = c \underbrace{\boxed{\theta^{\alpha-1} (1-\theta)^{\beta-1}}}_{\text{Beta}(\alpha,\beta)} \theta^{s} (1-\theta)^{n-s}$$

$$= \underbrace{c \; \theta^{(\alpha+s)-1} (1-\theta)^{(\beta+n-s)-1}}_{\text{Beta}[\alpha+s,\, \beta+(n-s)]}$$

(Bayes, 1760)

prior, likelihood & posterior all have same mathematical form

when this happens, prior is said to be *conjugate* to that likelihood

(Raiffa Schlaifer) (1955)

$(\alpha, \beta) > 0$

here, family of Beta(α, β) distributions is conjugate to Bernoulli likelihood

---

$$\theta \sim \text{Beta}(\alpha, \beta) \rightarrow p(\theta) = c \; \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$E(\theta) = \frac{\alpha}{\alpha+\beta} = 0.15$$

$$\int_{0.05}^{0.30} \text{Beta}(\theta \mid \alpha, \beta) \, d\theta = 0.95$$

Judgments: $\to$ Assumption $\qquad \theta \sim \text{Beta}(4.5, 25.5)$

$(\alpha, \beta > 0) \qquad y = (y_1, \dots, y_n)$

$$\left\{ \begin{array}{c} (\theta \mid \mathcal{B}) \sim \text{Beta}(\alpha, \beta) \\ (Y_i \mid \theta\, \mathcal{B}) \stackrel{\text{IID}}{\sim} \text{Bernoulli}(\theta) \\ (i = 1, \dots, n) \end{array} \right\} \to (\theta \mid y\, \mathcal{B}) \sim \text{Beta}(\alpha + s, \beta + n - s) \qquad \left(s = \sum_{i=1}^{n} y_i \right)$$

our first conjugate updating example

$$\begin{array}{cc} \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \hline 0 \\ \vdots \\ 0 \end{pmatrix} & \begin{array}{c} \uparrow \\ \alpha \\ \downarrow \\ \uparrow \\ \beta \\ \downarrow \end{array} \end{array}$$

mean $\dfrac{\alpha}{\alpha + \beta}$

"prior sample size" $= (\alpha + \beta) = n_0$

"prior data set" $\quad 4.5 + 25.5 = 30$

$$\begin{array}{cc} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \hline 0 \\ \vdots \\ 0 \end{bmatrix} & \begin{array}{c} s \\ \\ \\ \\ n - s \\ \\ \end{array} \end{array}$$

mean $\dfrac{s}{n} = \bar{y}$

sample data set

data sample size $= n = 400$

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

posterior mean = prior mean sample ($\bar{y}$) mean

$$\frac{\alpha+s}{\alpha+\beta+n} = \frac{\overbrace{(\alpha+\beta)}^{n_0}\left(\frac{\alpha}{\alpha+\beta}\right) + (n)\left(\frac{s}{n}\right)}{\alpha+\beta+n}$$

= weighted average of prior mean & sample mean, with weights = (prior sample size) & (data sample size)

posterior

$\boxed{\text{Beta}(\alpha+s, \beta+n-s)}$

$\alpha = 4.5$

$\beta = 25.5$

$n = 400$

$s = 72$

prior

$\boxed{\text{Beta}(\alpha, \beta)}$

$c\,\theta^{\alpha-1}(1-\theta)^{\beta-1}$

likelihood

$\boxed{\text{Beta}(s+1, n-s+1)}$

$c\,\theta^{s}(1-\theta)^{n-s}$

$(\theta \mid s\ B) \sim \text{Beta}(\alpha + s, \beta + n - s)$

$\alpha = 4.5$, $s = 72$, $\beta = 25.5$, $n = 400$, $s = 72$

method	~~[illegible]~~
likelihood	
Bayes	

$E(\theta \mid s\ B) = \dfrac{\alpha + s}{(\alpha + s) + (\beta + n - s)}$

$\dfrac{\alpha + s}{\alpha + \beta + n}$

$= 0.178$ post. mean

Bayesian posterior SD $= \underline{\underline{0.0184}}$

likelihood analysis mean $\bar{y} = \dfrac{s}{n}$

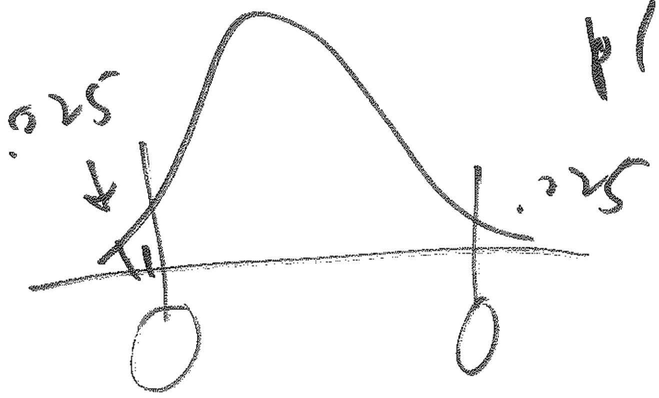
$\hat{\theta}_{MLE} = 0.18$

likelihood approx. 95% CI $\boxed{CLT}$

$\hat{\theta} \pm 1.96\ \hat{SE}(\hat{\theta}) = (0.142, 0.218)$

$0.18 \pm 1.96\ (0.0192)$

$\hat{SE}(\hat{\theta}_{MLE}) = \sqrt{\dfrac{\hat{\theta}(1-\hat{\theta})}{n}} = \underline{\underline{0.0192}}$



$p(\theta \mid y \mathcal{B}) \doteq$ Beta$(\alpha+s, \beta+n-s)$

.143	.215	Bayes
.142	.218	lik

Bayes $\doteq$ likelihood

$\boxed{\text{iff } (*)}$ • small amount of prior info

& • large n (large amount of data info)

---

under these conditions, ok to use likelihood analysis as (fast) (good) approximate Bayes analysis

$P_F(0.14 < \theta < 0.22) \doteq \cancel{0.95}$ undefined

$P_B(0.14 < \theta < 0.22 \mid \text{data, low info content prior context}) \doteq 0.95$ $\boxed{\text{yes}}$

Fisher (likelihood) =

Bayes with flat prior

diffuse

low-information content

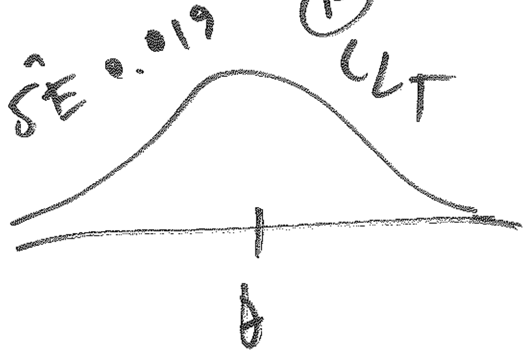
~~noninformative~~

so under iff ⊛ ok

to make frequentist calculation

& interpret it Bayesianly

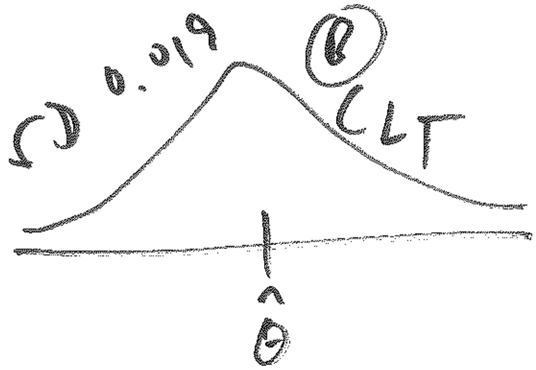
—



repeated-sampling dist of $\hat{\theta}$, n large

Freq.

equal but with different meaning



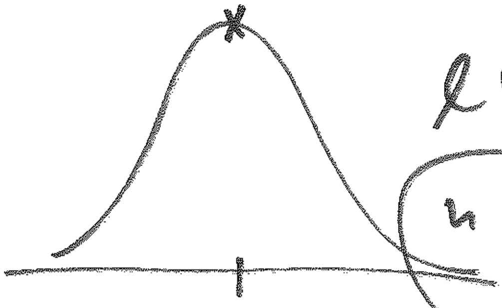
posterior dist. of θ given data, n large

Bayes

Bernstein-Von Mises Theorem

normal density

$$c_1 e^{-c_2(\theta - \hat{\theta})^2}$$



mode

= mean

$$l(\theta \mid y) \doteq p(\theta \mid y)$$

n large

CLT

$$\hat{\theta}_{MLE} \doteq E(\theta \mid y)$$

maximizing

post. mean