Bayesian Statistics

Thursday, November 13, 2014 2:23 PM

{ Bayesian / Frequentist

population mean μ. or proportion p

Frequentist: μ. p are fixed quantity.

Bayesian: μ and p are not fixed anymore

μ and p reflects researcher belief about some event

μ and p are viewed as random variables.

A be the event that there is a tiger in the street

$p = P(A)$

Bayesian. p is the kid's ~~belief~~ that ..

No data. p very small, close to 0 "prior" information

data: three people — update his belief regarding the parameter p.

p is very large. $p = 1$.

Bayesian Inference is always based on the current sample only

① Setting up the probability model, "p" nuisance parameter

② Given the prior for the parameters (model X)

③ Update your belief: compute the distribution of the parameter given the data $\theta | Y$. → posterior distribution

Do the inference. estimate. intervals. testing

④ Evaluating the model fit

§2. Posterior distribution.

Use y to denote the observation. Θ to denote the parameter.

① $Y|\theta \sim f(y|\theta)$

② $\theta \sim f(\theta)$

Joint density of (y, θ) is $f(y,\theta) = f(y|\theta) f(\theta)$

Bayes Theorem

$$f(\theta|y) = \frac{f(y|\theta) f(\theta)}{\int f(y|\theta) f(\theta)\, d\theta} = \frac{f(y|\theta) f(\theta)}{f(y)} \propto f(y|\theta) f(\theta)$$

$$\int f(\theta|y)\, d\theta = \int \frac{f(y|\theta) f(\theta)}{\int f(y|\theta) f(\theta)\, d\theta}\, d\theta = 1$$

Example: Binomial Model

Assume that $\begin{cases} X \sim Bin(n, p) \\ p \sim Beta(\alpha, \beta) \end{cases}$ $\quad \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}$

$f(p|x)$?

$$f(x|p) \propto p^{x}(1-p)^{n-x}$$

$$f(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$$

$$f(p|x) \propto p^{x+\alpha-1}(1-p)^{n-x+\beta-1}$$

$$p|x \sim Beta(\alpha+x, n-x+\beta)$$

Conjugate Prior distributions

$\mathcal{F}$ is a class of sampling distributions $p(y|\theta)$.
and $\mathcal{P}$ is a class of prior for θ. then $\mathcal{P}$ is conjugate for $\mathcal{F}$ if

$$f(\theta|y) \in \mathcal{P}, \quad \forall f(y|\theta) \in \mathcal{F}, \quad f(\theta) \in \mathcal{P}.$$

$\mathcal{P}$ consists of all possible distribution of θ.

Example 8.2.2 Normal - Normal Model

① $Y|\theta \sim N(\theta, \sigma^2)$

② $\theta \sim N(\mu, \tau^2)$

$\theta|Y$?

$$f(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\theta)^2}{2\sigma^2}\right\} \propto \exp\left\{-\frac{(y-\theta)^2}{2\sigma^2}\right\}$$

$$f(\theta) \propto \exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}$$

$$f(\theta|y) \propto \exp\left\{-\frac{1}{2}\left(\frac{(y-\theta)^2}{\sigma^2} + \frac{(\theta-\mu)^2}{\tau^2}\right)\right\}$$

$$\frac{(y-\theta)^2}{\sigma^2} + \frac{(\theta-\mu)^2}{\tau^2} = \frac{\theta^2 - 2\theta y + y^2}{\sigma^2} + \frac{\theta^2 - 2\mu\theta + \mu^2}{\tau^2}$$

$$= \theta^2\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right) - 2\cdot\theta\left(\frac{y}{\sigma^2} + \frac{\mu}{\tau^2}\right) + \text{constant}$$

$$= \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)\left(\theta^2 - 2\theta\cdot\frac{\frac{y}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right) + \text{constant}$$

$$= \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)\left(\theta - \frac{\frac{y}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right)^2 + \text{constant}$$

$$f(\theta|y) \propto \exp\left\{-\frac{1}{2}\frac{\left(\theta - \frac{\frac{y}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right)^2}{\frac{1}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}}\right\}$$

$$\theta|y \sim N\left(\frac{\frac{y}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}, \quad \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right)$$

$$M = \frac{\tau^2}{\sigma^2 + \tau^2} = \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}$$

$$\theta|y \sim N\left(My + (1-M)\mu, \ M\sigma^2\right)$$

8.3 Bayesian Inference

How to estimate θ ?

$\underline{\theta | y}$ is the posterior.

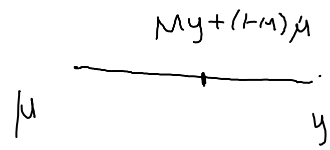
$\underline{E\,\theta | y}$ Median $(\theta | y)$

In the normal - normal model.

$$\hat{\theta} = E\,\theta | y = \underline{M y} + \underline{(1-M)\mu}.$$

weighted average of y and μ.

$$M = \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}.$$ $\sigma^2 \ll \tau^2$. M is close to 1



$y_1, y_2 \cdots y_n \sim N(\theta, \sigma^2)$ $\quad \bar{y} \sim N(\theta, \frac{\sigma^2}{n})$

$$E(\theta | \bar{y}) = M\bar{y} + (1-M)\mu. \quad M = \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$$

~~B~~Binomial $X \sim \text{Bin}(n, p)$ $\quad \hat{p} = \frac{X}{n}$

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

$$p | x \sim \text{Beta}(\alpha + x, n - x + \beta)$$

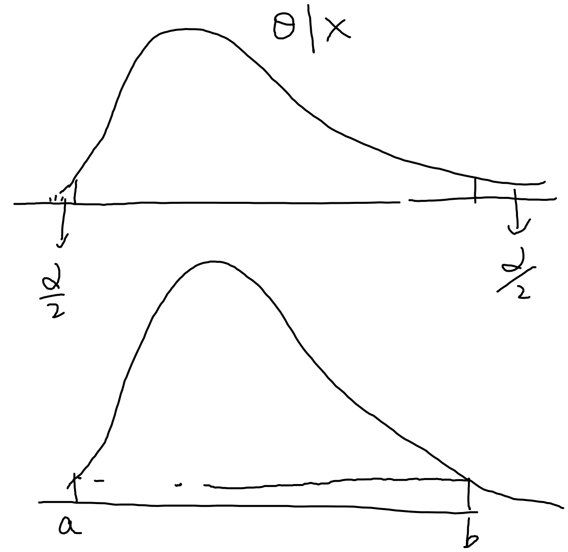
$$\hat{p} = E\,p | x = \frac{\alpha + x}{\underline{\alpha + x} + \underline{n - x + \beta}} = \frac{\alpha + x}{n + \alpha + \beta} = \frac{x}{n+\alpha+\beta} + \frac{\alpha}{n+\alpha+\beta}$$

$$= \frac{n}{n+\alpha+\beta} \frac{x}{n} + \frac{\alpha+\beta}{n+\alpha+\beta} \frac{\alpha}{\alpha+\beta}$$

$$= w \cdot \frac{x}{n} + (1 - w) \cdot \frac{\alpha}{\alpha+\beta}.$$

$$\text{Risk} = E_{(x,\theta)}(\underline{\hat{\theta}} - \theta)^2 \qquad \hat{\theta} = E\,\theta | X$$

Risk 2 $= E_{(X,\theta)} |\hat{\theta} - \theta| \Rightarrow \hat{\theta} = \text{median}(\theta | X)$.



8.3.2: Credible Interval

equal tail Interval

HPD Interval

Highest Posterior density

(i) $f(a) = f(b)$

(ii) $F(b) - F(a) = 1 - \alpha$

$$P(\theta \in CI | X) \geq 1 - \alpha$$

Normal - Normal Model

$$\theta | X \sim N(MY + (1-M)\mu,\ M\sigma^2)$$

$(1-\alpha)$-CI for θ is

$$MY + (1-M)\mu \pm z_{\frac{\alpha}{2}} \sqrt{M\sigma^2}$$

8.3.3 Hypothesis Testing

$H_0: \theta = 0$ $\quad$ $H_0: \theta \in \Theta_0$ $\quad$ $(1, 2)$

$H_a: \theta \neq 0$ $\quad$ $H_a: \theta \in \Theta_0^c$ $\quad$ $(-\infty, 0)$

$\theta | y \to$ posterior.

$P(\theta \in \Theta_0 | Y) < P(\theta \in \Theta_0^c | Y)$, reject H_0

$P(\theta \in \Theta_0 | Y) > P(\theta \in \Theta_0^c | Y)$, accept H_0.

$P(\theta \in \Theta_0 | Y) < \frac{1}{2}$ $\quad$ reject H_0

$P(\theta \in \Theta_0 | Y) > \frac{1}{2}$ $\quad$ Accept H_0.

p-value

Example 8.3.1. Female birth rate

$n = 980$

$437 \to$ Female

Let Y be the number of female that is born under placenta previa,

θ be the proportion.

① $Y \mid \theta \sim Bin(980, \theta)$

② $\theta \sim Beta(\alpha, \beta)$

$$\theta \mid y \sim Beta(y+\alpha, n-y+\beta) = Beta(437+\alpha, 543+\beta)$$

$$E\theta \mid y = \frac{437+\alpha}{437+\alpha+543+\beta} = \frac{437+\alpha}{980+\alpha+\beta}.$$

posterior mean. $\qquad w = \frac{n}{n+\alpha+\beta} = \frac{980}{980+\alpha+\beta}$

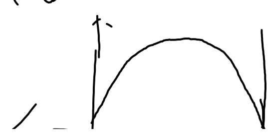
95% credible interval

$\frac{\alpha}{\alpha+\beta}$	$\alpha+\beta$		
0.5	2	0.446	[0.415, 0.477]
0.485	2	0.446	[0.415, 0.477]
0.485	5	0.446	[0.415, 0.477]
0.485	200	0.453	[0.424, 0.481]

① How to choose α and β?

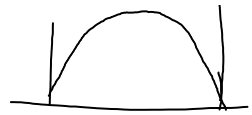
② $\frac{\theta}{1-\theta}$ $\qquad \log\frac{\theta}{1-\theta}$ $\qquad$ Sin($\exp\frac{\theta}{1-\theta}$)

$\alpha = \beta = 2$



choose the prior.

$\alpha = \beta = 2$



choose the prior:

① Informative Prior. model the temprature for tomorrow.

Use the temporature ~~of~~ today as my prior

② Non-informative prior. $Y|\theta \sim Bin(n, \theta)$

$\theta \sim U(0,1)$ $\boxed{\alpha = \beta = 1}$

③ Consider the hierarchical Prior

$$\left.\begin{aligned} Y|\theta &\sim Bin(n, \theta) \\ \theta &\sim Beta(\alpha, \beta) \\ \alpha, \beta &\sim Unif(0, 100) \end{aligned}\right\} \underline{\theta|Y}$$

$$\left\{\begin{aligned} X|\theta &\sim N(\theta, \sigma^2) \\ \theta &\sim N(0, 10000) \end{aligned}\right. \qquad \left\{\begin{aligned} X|\theta &\sim N(\theta, \sigma^2) \\ \theta &\sim N(\mu, \tau^2) \\ \mu &\sim N(0, 10000) \end{aligned}\right.$$

$$\underline{\theta|Y} = \frac{\iint f(y|\theta) f(\theta|\alpha,\beta) f(\alpha,\beta)\, d\alpha\, d\beta}{\int f(y|\theta) f(\theta|\alpha\beta) f(\alpha\beta)\, d\alpha\, d\beta\, d\theta}$$

MCMC

estimate $\frac{\theta}{1-\theta}$, $\log \frac{\theta}{1-\theta}$. Non-informative Prior $\alpha = \beta = 1$

$$\theta|y \sim Beta(437 + \alpha, 543 + \beta) \sim Beta(438, 544)$$

$\frac{\theta}{1-\theta} \Big| y$ $\theta_1, \theta_2 \cdots \theta_{10,000}$

$$\frac{\theta_1}{1-\theta_1} \quad \frac{\theta_2}{1-\theta_2} \quad \cdots \quad \frac{\theta_{10,000}}{1-\theta_{10,000}}$$

$$\widehat{\frac{\theta}{1-\theta}} = \text{mean}\left(\frac{\theta_i}{1-\theta_i}\right)$$

H_0: $\cancel{\theta < 0.485}$ $\theta \geq 0.485$

H_a: $\cancel{\theta \geq 0.485}$ $\theta < 0.485$

$P(\theta < 0.485 | Y) = 0.99221$ ~~Accept H_0~~. Reject H_0.

$P(\theta < 0.485 \mid Y) = \underline{0.99}221$, ~~Accept H_0~~. Reject H_0.

Monte-Carlo Markov Chain (MCMC)

Placenta Previa causes low rate of female birth.