Binomial model

2021年5月14日 15:16

θ 出现正面概率，未知

Y：Coin 出现正面次数，n 已知

Beta - Binomial model

Likelihood (data)：$y|\theta \sim Bin(\overset{const.}{n}, \theta)$ $\quad p(y|\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$

Prior：$\theta \sim Beta(a, b)$, $\theta \in (0,1)$ $\quad p(\theta) = \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}$

$0.5 \quad \theta \quad 1$

$*$ If $a = b = 1$ $\quad Beta(1,1) \overset{d}{=} Uniform(0,1)$

Posterior：$\theta|y \sim Beta(y+a, n-y+b)$

θ credibility ⇐ 看 data 下

$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$ const.

$\propto p(y|\theta)p(\theta)$

$\propto \binom{n}{y}\theta^{y}(1-\theta)^{n-y} \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}$

$\propto \theta^{y+a-1}(1-\theta)^{n-y+b-1} \sim \underline{Beta(y+a, n-y+b)}$

(1/2) prediction

$y^*|y \sim Pr(y^*|y) = \frac{y+a}{n+a+b}$

$\forall y^* \in \{0, 1, \dots, n\}$

$Pr(y^*|y) = \int_0^1 Pr(y^*|\theta, y)\, p(\theta|y)\, d\theta$

$= \int_0^1 \theta\, p(\theta|y)\, d\theta$

$= E(\theta|y) = \frac{y+a}{n+a+b}$ $\quad \forall y^* \in \{1, \dots, n\}$

Property 1：

- $E(\theta|y) = \frac{y+a}{y+a+n-y+b} = \frac{y+a}{n+a+b}$

  $E(\theta) = \frac{a}{a+b}$

  Sample mean：$\bar{y} = \frac{y}{n}$

  $\Rightarrow \underset{\text{prior mean}}{\frac{a}{a+b}} \lesseqgtr \underset{\text{posterior mean}}{\frac{y+a}{n+a+b}} \lesseqgtr \underset{\text{sample mean}}{\frac{y}{n}}$

  $$\underset{\text{posterior}}{\frac{z+a}{N+a+b}} = \underset{\text{data}}{\frac{z}{N}}\underset{\text{weight}}{\frac{N}{N+a+b}} + \underset{\text{prior}}{\frac{a}{a+b}}\underset{\text{weight}}{\frac{a+b}{N+a+b}}.$$

- $Mode(\theta|y) = \frac{y+a-1}{n+a+b-1} = MLE_\theta$

  $\Rightarrow$ unbiased estimator：$E(Mode(\theta|y)) = \theta$

- $Var(\theta|y) = \frac{E(\theta|y)(1-E(\theta|y))}{n+a+b+1}$

Property 2：

n and y become larger with fix a and b

$E(\theta|y) = \frac{y+a}{n+a+b} \approx \frac{y}{n}$

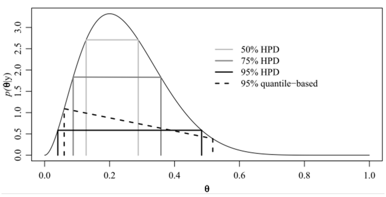
$Var(\theta|y) \approx \frac{1}{n}\frac{y}{n}(1-\frac{y}{n})$ approach to 0 with rate $\frac{1}{n}$

Note：In the limit, the par. of the prior have no influence on the posterior.

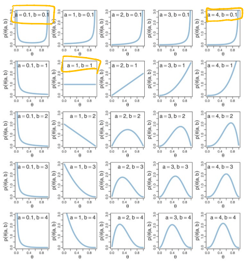
Theorem 1

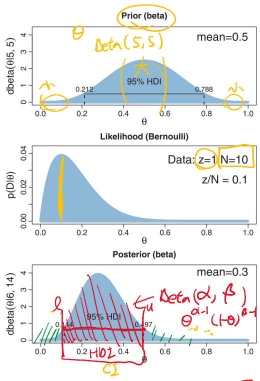
By the central limit theorem

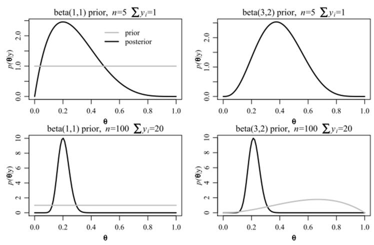
$$\left(\frac{\theta - E(\theta|y)}{\sqrt{Var(\theta|y)}} \,\middle|\, y\right) \longrightarrow N(0,1)$$



One way to summarize the uncertainty is by marking the span of values that are most credible and cover 95% of the distribution. This is called the highest density interval (HDI)





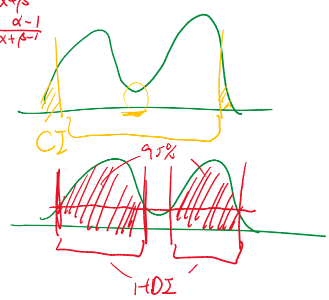


frequency θ = ? $\quad \mu \Leftarrow \bar{x} = 0.387$

bayesian $\theta|y \sim$ distribution

- point：posterior
  - mean $\quad E(\theta|y) = \frac{\alpha}{\alpha+\beta}$
  - mode $\quad Mode(\theta|y) = \frac{\alpha-1}{\alpha+\beta-2}$
  - median $\quad Median(\theta|y)$
- 95% HDI：$Pr(l \le \theta \le u) = 95\%$

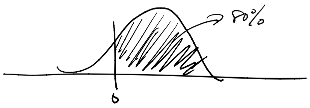
~~95% CI = [0.12, 0.58]~~



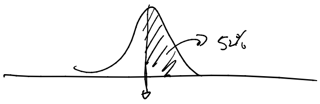
	θ_c conditional		θ_e exposure	
	无	有	没	有
data	$n_c - p_c$	p_c	$n_e - p_e$	p_e
prior	$a = 1 = b$		$a = 1 = b$	
post	○	○	○	○

$\theta_e - \theta_c$

$\theta_e - \theta_c \underline{=} 0 \Leftrightarrow \theta_e = \theta_c$



$Pr(\theta_e - \theta_c \geq 0) = 80\%$



$Pr(\theta_e - \theta_c \geq 0) = 52\%$