苏、зh 源、 2019010448 贝叶斯 HW1

5.

a) $P(y=k)=\int_0^1 C_n^k \theta^k (1-\theta)^{n-k} d\theta = \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} C_n^k = \frac{1}{n+1}$

b) $\theta \sim Bata(\alpha, \beta) \Rightarrow p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$

$p(y|\theta) \propto \theta^y (1-\theta)^{n-y}$, $p(\theta|y) \propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$

$\therefore E[\theta|y] = \frac{\alpha+y}{\alpha+\beta+n} = \frac{y}{n} + \frac{\alpha+\beta}{\alpha+\beta+n}\left(\frac{\alpha}{\alpha+\beta} - \frac{y}{n}\right)$

$\frac{\alpha+\beta}{\alpha+\beta+n} \in (0,1)$ $\therefore E[\theta|y] \in \left(\min\left(\frac{\alpha}{\alpha+\beta}, \frac{y}{n}\right), \max\left(\frac{\alpha}{\alpha+\beta}, \frac{y}{n}\right)\right)$

c) $\theta \sim U(0,1)$ $Var(\theta) = \frac{1}{12}$

$Var[\theta|y] = \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}$. 注意到 $(\alpha+y)+(\beta+n-y) = \alpha+\beta+n$

$\therefore Var[\theta|y] \leq \frac{1}{4} \cdot \frac{1}{\alpha+\beta+n+1} < \frac{1}{4} \cdot \frac{1}{1+1+0+1} = \frac{1}{12} = Var[\theta]$

d) $Var[\theta] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$, $Var[\theta|y] = \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}$

取 $\alpha=1$, $\beta=9$, $n=8$, $y=6$

$Var[\theta] = 8\times10^{-3}$, $Var[\theta|y] = 10^{-2} > Var[\theta]$

8. $\theta \sim N(\mu_0, \sigma_0^2) = N(180, 40^2)$, $\sigma^2 = 20^2$,

a) $\theta | Y \sim N(\mu_n, \sigma_n^2)$, $\mu_n = \dfrac{\frac{1}{\sigma_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}$, $\frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}$

b) $\tilde{y} | Y \sim N(\mu_n, \sigma_n^2 + \sigma^2)$, μ_n与σ_n^2同上

c) $n=10$时，$\mu_n = 150.73$. $\sigma_n^2 = 39.02$, $\sigma_n^2 + \sigma^2 = 439.02$

∴ θ的95%置信区间为 150.73 ± 12.24

$\tilde{y}$的95%置信区间为 150.73 ± 40.07

d) $n=100$时，(方差,均值)向$(150, 20^2)$移动. σ_n^2变小

故95%置信区间变窄.

e) $\frac{n}{\sigma^2} = \frac{1}{\tau_0^2} \Rightarrow n = \frac{\sigma^2}{\tau_0^2} = \left(\frac{20}{40}\right)^2 = \frac{1}{4}$

19.

a) $p(y|\theta) = \prod_{i=1}^{n} \theta e^{-\theta y_i}$ $\qquad$ $p(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}$

$p(\theta|y) \propto \theta^{\alpha-1+n} e^{-(\beta+n\bar{y})\theta} \sim \text{Gamma}(\alpha+n, \beta+n\bar{y})$

b) $p(\frac{1}{\theta})|d(\frac{1}{\theta})| = p(\theta)|d\theta|$

$\Rightarrow p(\frac{1}{\theta}) = p(\theta)\cdot\theta^2 = p(\frac{1}{\phi})\cdot\frac{1}{\phi^2} \propto \phi^{-\alpha-1} e^{-\frac{\theta}{\phi}}$

$\therefore \phi \sim \text{InvGamma}(\alpha, \theta)$

c) $E[\theta|y] = \dfrac{\alpha+n}{\beta+n\bar{y}}$, $Var[\theta|y] = \dfrac{\alpha+n}{(\beta+n\bar{y})^2}$

$Var[\theta|y]^{1/2} / E[\theta|y] = \dfrac{1}{\sqrt{\alpha+n}}$, $n=0$时, $\dfrac{1}{\sqrt{\alpha+n}} = \dfrac{1}{2} \Rightarrow \alpha = 4$

$\therefore \dfrac{1}{\sqrt{\alpha+n}} = 0.1 \Rightarrow n = 96$

Add. 前提：每天访问量服从正态分布、

根据当前策略(无chatGPT)下14天内每天访问量，计算均值μ_0，方差σ_0^2

先验：加上chatGPT后每天访问量服从 $N(\mu_0, \sigma_0^2)$

根据加上chatGPT后每天访问量计算后验参数μ'

$\mu' = \left(\dfrac{\mu_0}{\sigma_0^2} + \dfrac{n\bar{y}}{\sigma^2}\right) \Big/ \left(\dfrac{1}{\sigma_0^2} + \dfrac{n}{\sigma^2}\right)$, 其中$\bar{y}$为加上chatGPT后均值

σ^2 — — — 外 方差

$\dfrac{1}{\sigma'^2} = \dfrac{1}{\sigma_0^2} + \dfrac{1}{\sigma^2}$ $\qquad$ 若μ_0落在μ'的95%置信区间则用新方案