Computational Statistics Homework 8

吴嘉骜 21307130203

2023 年 11 月 27 日

Problem 1

Prove the following results about conjugate priors in Bayesian analysis.
(a) Beta distribution is the conjugate prior for the success probability parameter p of a geometric distribution. That is, let the prior of p be Beta(α, β). Given n independent and identically distributed random samples $X_1, \ldots, X_n$ from the geometric distribution with parameter p, then the posterior distribution of p is still Beta. Recall that the probability density function of Beta(α, β) is

$$f(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}, \quad 0 \leq y \leq 1 \text{ and } \alpha, \beta > 0.$$

(b) Inverse Gamma (IG) distribution is the conjugate prior for variance parameter σ^2 of a normal distribution with known mean parameter μ_0. That is, let the prior of σ^2 be IG(α, β). Given n independent and identically distributed random samples $X_1, \ldots, X_n$ from $N(\mu_0, \sigma^2)$, then the posterior distribution of σ^2 is still IG. Recall that the probability density function of Inverse Gamma(α, β) is

$$f(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} (1/y)^{\alpha+1} e^{-\beta/y}, \quad y > 0 \text{ and } \alpha, \beta > 0.$$

Proof:

(a) Since the prior of p is Beta(α, β), the prior distribution of p is $\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$.
The likelihood function is $f(x|p) = \prod_{i=1}^n p(1-p)^{x_i} = p^n (1-p)^{\sum\limits_{i=1}^n x_i}$, as X_i are i.i.d.
The posterior distribution of p is

$$f(p|x) = \frac{f(x|p)\pi(p)}{\int_0^1 f(x|p)\pi(p)dp} \propto p^n (1-p)^{\sum\limits_{i=1}^n x_i} p^{\alpha-1}(1-p)^{\beta-1} = p^{\alpha+n-1}(1-p)^{\beta+\sum\limits_{i=1}^n x_i - 1}$$

which is the form of the density function of Beta$(\alpha+n, \beta+\sum\limits_{i=1}^n x_i)$.
Since $f(p|x)$ is a density function, the normalizing constant must be aligned with that of the density function Beta$(\alpha+n, \beta+\sum\limits_{i=1}^n x_i)$, which is $\frac{\Gamma(\alpha+n+\beta+\sum_{i=1}^n x_i)}{\Gamma(\alpha+n)\Gamma(\beta+\sum_{i=1}^n x_i)}$.
Therefore, the posterior distribution of p is Beta$(\alpha+n, \beta+\sum\limits_{i=1}^n x_i)$.
(b) Since the prior of σ^2 is IG(α, β), the prior distribution of σ^2 is $\pi(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)}(1/\sigma^2)^{\alpha+1} e^{-\beta/\sigma^2}$.
The likelihood function is

$$f(x|\sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu_0)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{\sum_{i=1}^n (x_i-\mu_0)^2}{2\sigma^2}\right\},$$

as X_i are i.i.d.
The posterior distribution of σ^2 is

$$\begin{aligned}
f(\sigma^2|x) &= \frac{f(x|\sigma^2)\pi(\sigma^2)}{\int_0^\infty f(x|\sigma^2)\pi(\sigma^2)d\sigma^2} \\
&\propto (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{\sum_{i=1}^n (x_i-\mu_0)^2}{2\sigma^2}\right\} (1/\sigma^2)^{\alpha+1} e^{-\beta/\sigma^2} \\
&= (1/\sigma^2)^{\alpha+\frac{n}{2}+1} \exp\left\{-\frac{\frac{1}{2}\sum_{i=1}^n (x_i-\mu_0)^2+\beta}{\sigma^2}\right\}
\end{aligned}$$

which is the form of the density function of $\mathrm{IG}(\alpha+\frac{n}{2}, \frac{1}{2}\sum_{i=1}^n (x_i-\mu_0)^2+\beta)$.
Similar to (a), it can be indicated that the normalizing constant of $f(\sigma^2|x)$ is $\frac{\left(\frac{1}{2}\sum_{i=1}^n (x_i-\mu_0)^2+\beta\right)^{\alpha+\frac{n}{2}}}{\Gamma(\alpha+\frac{n}{2})}$.
Therefore, the posterior distribution of σ^2 is $\mathrm{IG}(\alpha+\frac{n}{2}, \frac{1}{2}\sum_{i=1}^n (x_i-\mu_0)^2+\beta)$.

Problem 2

Consider the Bayesian estimation of the success probability parameter for a rare event. Suppose n i.i.d. Bernoulli experiments with success probability $\theta \in [0,1]$ are conducted. Then the number of successes y follows a binomial distribution $\mathrm{Bin}(n,\theta)$. Our interest is in estimating θ. Take $\mathrm{Beta}(a,b)$ as a prior for θ.
(a) Derive the posterior distribution $\theta \mid y$.
(b) Express the posterior mean of $\theta \mid y$ as a linear combination of the sample average $\bar{y} = y/n$ and the prior expectation of θ.
(c) Comment on the effect of $\bar{y}$ on the shift of the posterior from the prior.
Solution:
(a) The prior distribution of θ is $\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$.
The likelihood function is $f(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$.
Then the posterior distribution of θ is

$$\begin{aligned}
f(\theta|y) &= \frac{f(y|\theta)\pi(\theta)}{\int_0^1 f(y|\theta)\pi(\theta)d\theta} \\
&\propto \theta^y(1-\theta)^{n-y}\theta^{a-1}(1-\theta)^{b-1} \\
&= \theta^{a+y-1}(1-\theta)^{b+n-y-1}
\end{aligned}$$

which indicates that the posterior distribution of θ is $\mathrm{Beta}(a+y, b+n-y)$.
(b) The prior expectation of θ is $\mathrm{E}(\theta) = \frac{a}{a+b}$.
The posterior mean of θ is $\mathrm{E}(\theta|y) = \frac{a+y}{a+b+n}$.

We can derive the above expression as:

$$\begin{aligned}
\mathrm{E}(\theta|y) &= \frac{a+y}{a+b+n} \\
&= \frac{a}{a+b+n} + \frac{y}{a+b+n} \\
&= \frac{1}{\frac{n}{a+b}+1} \cdot \frac{a}{a+b} + \frac{y/n}{\frac{a+b}{n}+1} \\
&= \frac{1}{\frac{n}{a+b}+1} \cdot \mathrm{E}(\theta) + \frac{1}{\frac{a+b}{n}+1} \cdot \bar{y}
\end{aligned}$$

which is a linear combination of $\mathrm{E}(\theta)$ and $\bar{y}$.
(c) $\bar{y} = y/n$ represents the proportion of successful experiments in the sample. As $\bar{y}$ increases, indicating more observed successes, the expected probability of success in the posterior distribution also increases. Specifically, if all other factors remain the same, an increase in $\bar{y}$ elevates the posterior mean $\mathrm{E}(\theta|y)$, causing the posterior to shift up from the prior and align more closely with the higher rate of success observed in the data.