MA322: Advanced Statistical Methods – Work Sheet 2

1. Suppose that $X_1, X_2, \ldots,$ are Bernoulli trials with success probability θ. Let t denote the observed number of successes in n such trials. Assume the beta prior distribution with parameters α and β. Show that

$$Pr(X_{n+1} = 1|t) = \frac{t+\alpha}{n+\alpha+\beta}.$$

Solution:

We know that the posterior of θ is the beta distribution:

$$\pi(\theta|t) = \frac{1}{B(t+\alpha, n-t+\beta)}\theta^{t+\alpha-1}(1-\theta)^{n-t+\beta}, 0 < \theta < 1.$$

Now

$$\begin{aligned} Pr(X_{n+1} = x|t) &= \textstyle\int_0^1 \theta^x (1-\theta)^(1-x) \frac{1}{B(t+\alpha,n-t+\beta)}\theta^{t+\alpha-1}(1-\theta)^{n-t+\beta} \\ &= \frac{B(x+t+\alpha,1-x+n-t+\beta)}{B(t+\alpha,n-t+\beta)}. \end{aligned}$$

When $x = 1$, we have

$$\begin{aligned} Pr(X_{n+1} = 1|t) &= \frac{B(1+t+\alpha,n-t+\beta)}{B(t+\alpha,n-t+\beta)} \\ &= \frac{t+\alpha}{n+\alpha+\beta} \end{aligned}$$

by using the relation:

$$B(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}.$$

□

2. Suppose that the number of defects on a roll of magnetic recording tape has a Poisson distribution for which the mean θ is unknown and that the prior distribution of θ is a gamma distribution with parameters $\alpha = 3$ and $\beta = 1$. When five rolls of this tape are selected at random and inspected, the number of defects found on the rolls are 2, 2, 6, 0, and 3. Let t denote the sum of the observations. Show that

$$Pr(X_{n+1} = x|t) = \frac{(n+\beta)^{t+\alpha}}{(n+1+\beta)^{t+\alpha+x}} \frac{\Gamma(t+\alpha+x)}{\Gamma(t+\alpha)} \frac{1}{x!}.$$

Write a S-Plus function to calculate the probability. Obtain the answers for $x = 0, 1$ and 2.

Solution: The posterior here is the gamma distribution with parameters $t+\alpha$ and $n+\beta$. Now

$$\begin{aligned} Pr(X_{n+1} = x|t) &= \textstyle\int_0^\infty e^{-\theta} \frac{\theta^x}{x!} \frac{(n+\beta)^{t+\alpha}}{\Gamma(t+\alpha)}\theta^{t+\alpha-1}e^{-\theta(n+\beta)}d\theta \\ &= \frac{(n+\beta)^{t+\alpha}}{\Gamma(t+\alpha)} \frac{1}{x!} \textstyle\int_0^\infty \theta^{x+t+\alpha-1}e^{-\theta(1+n+\beta)}d\theta \\ &= \frac{(n+\beta)^{t+\alpha}}{\Gamma(t+\alpha)} \frac{1}{x!} \frac{\Gamma(x+t+\alpha)}{(1+n+\beta)^{(x+t+\alpha)}}. \end{aligned}$$

□

3. Suppose $X_1, \ldots, X_n$ is a random sample from the distribution with pdf $f(x|\theta) = \theta e^{-\theta x}$. Suppose the prior for θ is given by $\pi(\theta) = \beta e^{-\beta\theta}$ for some known $\beta > 0$. Let t denote the sum of the observations. Show that

$$f(x_{n+1}|t) = (n+1)\frac{(t+\beta)^{n+1}}{(x_{n+1}+t+\beta+)^{n+2}}.$$

Suppose that $n = 5, t = 1.36, \beta = 1$. Calculate the probability $Pr(X_{n+1} < a|t)$.

Solution: The posterior here is the gamma distribution with parameters $n+1$ and $t+\beta$. Now

$$\begin{aligned} f(x_{n+1}|t) &= \int_0^\infty \theta e^{-\theta x_{n+1}} \frac{(t+\beta)^{n+1}}{\Gamma(n+1)} \theta^{n+1-1} e^{-\theta(t+\beta)} d\theta \\ &= \frac{(t+\beta)^{n+1}}{\Gamma(n+1)} \int_0^\infty \theta^{n+2-1} e^{-\theta(x_{n+1}+t+\beta)} d\theta \\ &= \frac{(t+\beta)^{n+1}}{\Gamma(n+1)} \frac{\Gamma(n+2)}{(x_{n+1}+t+\beta)^{n+2}} \\ &= (n+1)\frac{(t+\beta)^{n+1}}{(x_{n+1}+t+\beta)^{n+2}}, \end{aligned}$$

using the relation

$$\Gamma(a+1) = a\,\Gamma(a).$$

Now

$$\begin{aligned} Pr(X_{n+1} < a|t) &= \int_0^a (n+1)\frac{(t+\beta)^{n+1}}{(x_{n+1}+t+\beta)^{n+2}} dx_{n+1} \\ &= (n+1)(t+\beta)^{n+1} \int_0^a (x_{n+1}+t+\beta)^{-(n+2)} dx_{n+1} \\ &= (n+1)(t+\beta)^{n+1} \left[\frac{(x_{n+1}+t+\beta)^{-(n+2)+1}}{-(n+2)+1}\right]_0^a \\ &= (t+\beta)^{n+1} \left\{\frac{1}{(t+\beta)^{n+1}} - \frac{1}{(a+t+\beta)^{n+1}}\right\}. \end{aligned}$$

□

4. Suppose that $X_1, \ldots, X_n$ is a random sample from $N(\theta, \sigma^2)$ population where $\sigma^2 = 100$. Let the observed value of $\bar{X}$ be 118.7 where $n = 10$. Assume the prior distribution $N(100, 225)$ for θ. Obtain the posterior predictive distribution of X_{n+1} and hence obtain the best 95% predictive interval for X_{n+1}.

Solution: Here the posterior distribution of $\theta|x_1, \ldots, x_n$ is the normal distribution with

$$\text{mean} = \frac{n\bar{x}/\sigma^2 + \mu/\tau^2}{n/\sigma^2 + 1/\tau^2} \text{ and Variance} = \frac{1}{n/\sigma^2 + 1/\tau^2}.$$

The predictive distribution is also normal with

$$\text{mean} = \frac{n\bar{x}/\sigma^2 + \mu/\tau^2}{n/\sigma^2 + 1/\tau^2} \text{ and Variance} = \sigma^2 + \frac{1}{n/\sigma^2 + 1/\tau^2}.$$

Substiuting all the numbers we have that the predictive distribution is normal with

$$\text{mean} = 117.9 \text{ and Variance} = 109.6.$$

Therfore, the 95% predictive interval is $(117.9 \pm 1.96\sqrt{109.6})$ which is (97.38, 138.42).

□