Chapter 7 Estimation

Yurun (Ellen) Ying

Sep 27, 2022

7.2 Prior and Posterior Distributions

Problem 1

Consider again the situation described in Example 7.2.8. This time, suppose that the experimenter believes that the prior distribution of θ is the gamma distribution with parameters 1 and 5000. What would this experimenter compute as the value of $\Pr(X_6 > 3000 \mid \boldsymbol{x})$?

Answer: Based on the given conditions, the prior p.d.f. for $\theta > 0$ is given by

$$\xi(\theta) = 5000 \cdot e^{-5000\theta}$$

Also, $\xi(\theta) = 0$ for $\theta \leq 0$. The p.d.f. for each X_i conditioned on θ is given by

$$f(x \mid \theta) = \theta e^{-\theta x}$$

for $x > 0$. Therefore, the conditional joint p.d.f. for the first five florescent lights is given by

$$f_n(\boldsymbol{x} \mid \theta) = \prod_i^n f(x_i \mid \theta)$$
$$= \theta^n e^{-\theta y}$$

where $y = \sum_i^n x_i$.
From $\xi(\theta \mid \boldsymbol{x}) \propto f(\boldsymbol{x} \mid \theta)\xi(\theta)$, we have

$$\xi(\theta \mid \boldsymbol{x}) \propto \theta^n e^{-\theta(5000+y)}$$

We can now compute $g_n(\boldsymbol{x})$, which is given by

$$g_n(\boldsymbol{x}) = \int_0^\infty \theta^n e^{-\theta(5000+y)} d\theta = \frac{\Gamma(n+1)}{(y+5000)^{n+1}}$$

From the previous two equations, we have

$$\xi(\theta \mid \boldsymbol{x}) = \frac{(y+5000)^{n+1}}{\Gamma(n+1)} \theta^n e^{-\theta(5000+y)}$$

When $n = 5$ and $y = 16,178$, this equation becomes

$$\xi(\theta \mid x_1, \ldots, x_5) = \frac{(21,178)^6}{\Gamma(6)} \theta^5 e^{-21,178\theta}$$

To compute $\Pr(X_6 > 3000 \mid \boldsymbol{x})$, we first need the p.d.f. of X_6 conditioned on the $X_1, \ldots, X_5$,

$$f(x_6 \mid x_1, \ldots, x_5) = \int_0^\infty f(x_6 \mid \theta)\xi(\theta \mid x_1, \ldots, x_5)$$
$$= \frac{(21,178)^6}{\Gamma(6)} \int_0^\infty \theta^6 e^{-(x_6+21,178)\theta} d\theta$$
$$= \frac{6(21,178)^6}{(x_6 + 21,178)^7}.$$

Therefore, we have

$$\Pr(X_6 > 3000 \mid \boldsymbol{x}) = 6(21,178)^6 \int_{3000}^\infty \frac{1}{(x_6 + 21,178)^7} dx_6 \approx 0.4516.$$

Problem 3

Suppose that the number of defects on a roll of magnetic recording tape has a Poisson distribution for which the mean λ is either 1.0 or 1.5, and the prior p.f. of λ is as follows:

$$\xi(1.0) = 0.4 \text{ and } \xi(1.5) = 0.6.$$

If a roll of tape selected at random is found to have three defects, what is the posterior p.f. of λ?

Answer: Based on the information given, the conditional p.d.f. of the X, the number of defects, can be written as

$$p(x \mid \lambda) = e^{-\lambda} \frac{\lambda^x}{x!},$$

for $x \leq 0$ and where $\lambda > 0$. To compute the posterior p.f. of λ, we first need the marginal distribution of X,

$$g(x) = p(x \mid 1.0)\,\xi(1.0) + p(x \mid 1.5)\,\xi(1.5) = \frac{0.4e^{-1} + 0.5 \cdot 1.5^x e^{-1.5}}{x!}.$$

Therefore, the posterior distribution of X is given by

$$\xi(1.0 \mid X = 3) = \frac{p(X = 3 \mid 1.0)\,\xi(1.0)}{g(3)} \approx 0.2457$$

and

$$\xi(1.5 \mid X = 3) = \frac{p(X = 3 \mid 1.5)\,\xi(1.5)}{g(3)} \approx 0.7543$$

Problem 5

Suppose that the prior distribution of some parameter θ is a beta distribution for which the mean is $1/3$ and the variance is $1/45$. Determine the prior p.d.f. of θ.

Answer: Based on the information given, we have

$$\begin{cases} \frac{\alpha}{\alpha+\beta} = \frac{1}{3} \\ \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{1}{45} \end{cases},$$

which gives $\alpha = 3$ and $\beta = 6$. Therefore, the prior distribution of θ is

$$\xi(\theta) = \frac{\theta^3\,(1-\theta)^6}{\mathrm{B}(3,6)},$$

where

$$\mathrm{B}(3,6) = \frac{\Gamma(3)\,\Gamma(6)}{\Gamma(9)},$$

which is the p.d.f. of a beta distribution with parameters 3 and 6.

Problem 7

Suppose that the proportion θ of defective items in a large manufactured lot is unknown, and the prior distribution of θ is given by

$$\xi(\theta) = \begin{cases} 2\,(1-\theta) & \text{for } 0 < \theta < 1 \\ 0 & \text{otherwise} \end{cases}.$$

When eight items are selected at random from the lot, it is found that exactly three of them are defective. Determine the posterior distribution of θ.

Answer: To determine the posterior distribution of θ, we first need to compute the conditional joint distribution of the result of first n draws of items, with each denoted by X_i. Note that each X_i has a bernoulli distribution with a conditional p.f. $p(x_i \mid \theta) = \theta^{x_i}\,(1-\theta)^{1-x_i}$. Since each draw

is independent (assuming the total number of items is very large), we can write the conditional joint distribution as

$$p_n(\boldsymbol{x} \mid \theta) = \sum_i^n p(x_i \mid \theta) = \theta^d \, \theta^{1-d},$$

where $d = \sum_i^n x_i$, the number of defect items in the first n draws.
We now have

$$\xi(\theta \mid x_1, \ldots, x_8) \propto p_8(\boldsymbol{x} \mid \theta) \, \xi(\theta)$$
$$\propto \theta^3 \, (1 - \theta)^6.$$

The proportionality constant is given by

$$\int_0^1 \theta^3 \, (1 - \theta)^6 d\theta = \mathrm{B}(4, \, 7).$$

Therefore,

$$\xi(\theta \mid x_1, \ldots, x_8) = \frac{1}{\mathrm{B}(4, \, 7)} \, \theta^3 \, (1 - \theta)^6,$$

which is the p.d.f. of a beta distribution with parameters 4 and 7.

7.3 Conjugate Prior Distribution

Problem 1

Consider again the situation described in Example 7.3.10. Once again, suppose that the prior distribution of θ is a normal distribution with mean 0, but this time let the prior variance be $v^2 > 0$. If the posterior mean of θ is 0.12, what value of v^2 was used?

Answer: From the information given, we have the following equation:

$$\frac{\sigma^2 \mu_0 + n v^2 \bar{x}_n}{\sigma^2 + n v^2} = 0.12,$$

where $\sigma^2 = 100$, $\mu_0 = 0$, $\bar{x}_n = 0.125$, and $n = 20$. Solving for v^2 yields $v^2 = 120$.

Problem 3

Suppose that the proportion θ of defective items in a large shipment is unknown and that the prior distribution of θ is the beta distribution with parameters 2 and 200. If 100 items are selected at random from the shipment and if three of these items are found to be defective, what is the posterior distribution of θ?

Answer: The posterior distribution of θ is a beta distribution with parameter $\alpha = 5$ and $\beta = 297$.

Problem 5

Suppose that the number of defects in a 1200-foot roll of magnetic recording tape has a Poisson distribution for which the value of the mean θ is unknown and that the prior distribution of θ is the gamma distribution with parameters $\alpha = 3$ and $\beta = 1$. When five rolls of this tape are selected at random and inspected, the numbers of defects found on the rolls are 2, 2, 6, 0, and 3. Determine the posterior distribution of θ.

Answer: The posterior distribution is a gamma distribution with parameter $\alpha = 16$ and $\beta = 6$.

Problem 7

Suppose that the heights of the individuals in a certain population have a normal distribution for which the value of the mean θ is unknown and the standard deviation is 2 inches. Suppose also that the prior distribution of θ is a normal distribution for which the mean is 68 inches and the standard deviation is 1 inch. If 10 people are selected at random from the population, and their average height is found to be 69.5 inches, what is the posterior distribution of θ?

Answer: From the information given, we have

$$\mu_1 = \frac{4 \times 68 + 10 \times 1 \times 69.5}{4 + 10 \times 1} = \frac{967}{14} \text{ and } v_1^2 = \frac{4 \times 1}{4 + 10 \times 1} = \frac{2}{7}.$$

Therefore, the posterior distribution of θ is a normal distribution with $\mu_1 = \frac{967}{14}$ and $v_1^2 = \frac{2}{7}$.

Problem 13

Suppose that the time in minutes required to serve a customer at a certain facility has an exponential distribution for which the value of the parameter θ is unknown.
For a distribution with mean $\mu \neq 0$ and standard deviation $\sigma > 0$, the *coefficient of variation* of the distribution is defined as $\sigma/|\mu|$. Suppose that the coefficient of variation of the prior gamma distribution of θ is 2. What is the smallest number of customers that must be observed in order to reduce the coefficient of variation of the posterior distribution to 0.1?

Answer: The coefficient of variation of the prior gamma distribution is

$$\frac{v_0}{\mu_0} = \frac{\sqrt{\alpha_0}/\beta_0}{\alpha_0/\beta_0} = \frac{1}{\sqrt{\alpha_0}} = 2,$$

which gives $\alpha_0 = \frac{1}{4}$. By the same token, the coefficient of variation of the posterior distribution is

$$\frac{1}{\sqrt{\alpha_1}} = \frac{1}{\sqrt{\alpha_0 + n}}.$$

To reduce this value to 0.1, the value of $\alpha_0 + n$ should at least be 100. This means that n at least needs to be 100.

7.4 Bayes Estimator

Problem 1

In a clinical trial, let the probability of successful outcome θ have a prior distribution that is the uniform distribution on the interval $[0, 1]$, which is also the beta distribution with parameters 1 and 1. Suppose that the first patient has a successful outcome. Find the Bayes estimates of θ that would be obtained for both the squared error and absolute error loss functions.

Answer: When the squared error loss function is used, the Bayes estimator is the mean of the posterior distribution of θ, which is a beta distribution with parameter $\alpha_1 = 2$ and $\beta_1 = 1$. We have the Bayes estimator as follows,

$$\delta^*(\boldsymbol{X}) = \mathbb{E}[\theta \mid \boldsymbol{x}] = \frac{\alpha_0 + \sum_i^n X_i}{\alpha_0 + \beta_0 + n}.$$

Based on the information given, $\sum X_i = 1$ and $n = 1$. Therefore, $\delta^*(\boldsymbol{x}) = \frac{2}{3}$.
When the absolute error loss function is used, the Bayes estimator is the median of the posterior distribution of θ. Note that the p.d.f. of the posterior distribution of θ is

$$\xi(\theta \mid \boldsymbol{x}) = \frac{\theta}{B(2, 1)},$$

and there for the cumulative distribution function is

$$F_{\theta \mid \boldsymbol{x}}(x) = \frac{1}{B(2, 1)} \int_0^x \theta d\theta.$$

Set this function equal to 0.5 and solve the equation, which gives $x = \frac{\sqrt{2}}{2}$. This means that the Bayes estimate of θ when absolute error loss function is used is $2^{-1/2}$.

Problem 3

Suppose that the proportion θ of defective items in a large shipment is unknown, and the prior distribution of θ is the beta distribution for which the parameters are $\alpha = 5$ and $\beta = 10$. Suppose also that 20 items are selected at random from the shipment.
a. For what number of defective items in the sample will the mean squared error of the Bayes estimate be a maximum?
b. For what number will the mean squared error of the Bayes estimate be a minimum?

Answer: From the information given, the posterior distribution of θ will be a beta distribution with parameters $\alpha_1 = 5 + y$ and $\beta_1 = 30 - y$, where y is the number of defective items among the sample of 20 items and $0 \le y \le 20$. We have the mean squared error being

$$\text{Var}(\theta \mid \boldsymbol{x}) = \frac{\alpha_1 \beta_1}{(\alpha_1 + \beta_1)^2 (\alpha_1 + \beta_1 + 1)} = \frac{(5 + y)(30 - y)}{35^2 \times 36} = \frac{-(y - 25/2)^2 + c}{35^2 \times 36}.$$

The estimator is maximized when $y = 12$ or 13 and minimized when $y = 0$.

Problem 5

Suppose that the number of defects in a 1200-foot roll of magnetic recording tape has a Poisson distribution for which the value of the mean θ is unknown, and the prior distribution of θ is the gamma distribution with parameters $\alpha = 3$ and $\beta = 1$. When five rolls of this tape are selected at random and inspected, the numbers of defects found on the rolls are 2, 2, 6, 0, and 3. If the squared error loss function is used, what is the Bayes estimate of θ?

Answer: Based on the information given, the posterior distribution of θ is a gamma distribution with parameters $\alpha_1 = 16$ and $\beta_1 = 6$. Therefore, we have the Bayes estimator being

$$\delta^* X = \frac{\alpha_1}{\beta_1} = \frac{8}{3}.$$

Problem 9

Suppose that a random sample is to be taken from a normal distribution for which the value of the mean θ is unknown and the standard deviation is 2, the prior distribution of θ is a normal distribution for which the standard deviation is 1, and the value of θ must be estimated by using the squared error loss function. What is the smallest random sample that must be taken in order for the mean squared error of the Bayes estimator of θ to be 0.01 or less?

Answer: The mean squared error of the Bayes estimator of θ is given by

$$
\begin{aligned}
\mathbb{E}[L(\theta, \delta^*(\boldsymbol{x})) \mid \boldsymbol{x}] &= \mathbb{E}[(\theta - \delta^*(\boldsymbol{x}))^2 \mid \boldsymbol{x}] \\
&= \mathbb{E}[\theta^2 \mid \boldsymbol{x}] - 2\delta^*(\boldsymbol{x})\mathbb{E}[\theta \mid \boldsymbol{x}] + (\delta^*(\boldsymbol{x}))^2 \\
&= \mathbb{E}[\theta^2 \mid \boldsymbol{x}] - (\mathbb{E}[\theta \mid \boldsymbol{x}])^2 \\
&= \mathrm{Var}[\theta \mid \boldsymbol{x}] \\
&= \frac{\sigma^2 v_0^2}{\sigma^2 + n v_0^2} = \frac{4}{4 + n}.
\end{aligned}
$$

This value being less than or equal to 0.01 gives the minimum value of $n = 396$.

7.5 Maximum Likelihood Estimator

Problem 3

It is not known what proportion p of the purchases of a certain brand of breakfast cereal are made by women and what proportion are made by men. However, we do know that $\frac{1}{2} \leq p \leq \frac{2}{3}$. In a random sample of 70 purchases of this cereal, it was found that 58 were made by women and 12 were made by men. Find the M.L.E. of p.

Answer: Based on the information given, the likelihood function of 70 observations is given by

$$f_{70}(\boldsymbol{x} \mid p) = p^{58}(1-p)^{12}.$$

Taking the derivative with respect to p gives

$$\frac{df}{dp} = p^{57}(1-p)^{11}(58(1-p) - 12p).$$

Setting the derivative to 0 and solving for p gives $p = \frac{29}{35}$. The second derivative at this point is negative, implying that the likelihood function reaches a local maximum. However, this value falls outside the given boundary of p. Therefore, the M.L.E. of p is $\frac{2}{3}$.

Problem 5

Suppose that $X_1, \ldots, X_n$ form a random sample from a Poisson distribution for which the mean θ is unknown $(\theta > 0)$.
a. Determine the M.L.E. of θ, assuming that at least one of the observed values is different from 0.
b. Show that the M.L.E. of θ does not exist if every observed value is 0.

Answer.a: Since X_i is a random sample from a Poisson distribution, the p.f. for each observation is

$$f(x_i \mid \theta) = e^{-\theta}\frac{\theta^{x_i}}{x_i!}.$$

Thus the joint p.f. and the likelihood function for all n observations is

$$f_n(x_i \mid \theta) = e^{-n\theta}\frac{\theta^y}{\prod_i^n (x_i!)},$$

where $y = \sum_i^n x_i$. Taking the derivative of the likelihood function with respect to θ gives

$$\frac{\partial f}{\partial \theta} = \frac{1}{\prod_i^n (x_i!)}(-ne^{-n\theta}\theta^y + e^{-n\theta}y\theta^{y-1}).$$

Setting it equal to zero and solving for θ give $\hat{\theta} = \frac{y}{n}$, which is the sample mean $\bar{x}_n$.

Answer.b: When all observations are 0, $\hat{\theta} = 0$. However, $\theta > 0$, therefore the M.L.E. of θ does not exist.

7.6 Properties of Maximum Likelihood Estimators

Problem 3

Suppose that $X_1, \ldots, X_n$ form a random sample from an exponential distribution for which the value of the parameter β is unknown. Determine the M.L.E. of the median of the distribution.

Answer: The p.d.f. of the exponential distribution is

$$f(x \mid \beta) = \beta e^{-\beta x}.$$

Let m be the median of this distribution, which should satisfy the equation

$$\Pr(x \leq m) = \int_0^m \beta e^{-\beta x} dx = \frac{1}{2}.$$

Solving for m yields $m = \frac{\ln 2}{\beta}$. The likelihood function is

$$f_n(\boldsymbol{x} \mid \beta) = \beta^n e^{-\beta y},$$

where $y = \sum_i^n x_i$. Taking the partial derivative with respect to β, setting it equal to 0 and solving for β give the M.L.E. $\hat{\beta} = \frac{1}{\overline{X}_n}$. According to the invariance property, the M.L.E. of the median $\hat{m} = \ln 2 \overline{X}_n$.

Problem 7

Suppose that $X_1, \ldots, X_n$ form a random sample from a normal distribution for which both the mean and the variance are unknown. Find the M.L.E. of $v = \Pr(X > 2)$.

Answer: The M.L.E. of $\theta = (\mu, \sigma^2)$ is

$$\hat{\theta} = (\hat{\mu}, \hat{\sigma^2}) = (\overline{X}_n, \frac{1}{n} \sum_i^n (X_i - \overline{X}_n)^2).$$

According to the invariance property of M.L.E., we have the M.L.E. of v,

$$\hat{v} = \int_2^\infty \frac{1}{\sqrt{2\pi V_n}} \exp[-\frac{(x - \overline{X}_n)^2}{2V_n}] dx,$$

where $V_n = \frac{1}{n} \sum_i^n (X_i - \overline{X}_n)^2$. This also can be expressed as $\hat{v} = 1 - \Phi(\frac{2 - \hat{\mu}}{\hat{\sigma}})$.