Chapter 5: Bayesian Estimation

In the last two Chapters 3 and 4, we assumed the random sample came from some known probability distribution $f(x, \theta)$ and we used the classic method to estimate the unknown parameter θ which was some fixed. In this Chapter, we will estimate θ using the Bayesian method which is define the unknown parameter θ as a random variable and has a distribution depending on previous informations called prior distribution.

Prior and Posterior Distributions

Consider a random variable X that has a distribution of probability that depends upon the symbol θ, where θ is an element of a well-defined set Ω. Let us now introduce a random variable Θ that has a distribution of probability over the set Ω. The probability distribution $h(\theta)$ is called the **prior distribution** of Θ. Moreover, we now denote the probability distribution of X by $f(x|\theta)$ since we think of it as a conditional distribution of X, given $\Theta = \theta$. For clarity in this chapter, we will use the following summary of this model:

$$X|\theta \sim f(x|\theta)$$

$$\Theta \sim h(\theta)$$

Thus, we can write the joint conditional distribution of X, given $\Theta = \theta$, as

$$L(x|\theta) = f(x_1|\theta)f(x_2|\theta) \dots f(x_n|\theta)$$

Thus, the joint distribution of X and Θ is

$$g(x, \theta) = L(x|\theta)h(\theta) \tag{5.1}$$

The marginal distribution of X is given by

$$g_1(x) = \begin{cases} \int_\theta \; g(x,\theta)d\theta, & \text{if } \Theta \text{ is a continuous} \\ \sum_\theta \; g(x,\theta), & \text{if } \Theta \text{ is a discrete} \end{cases}$$

In either case the conditional distribution of Θ, given the sample X, is

$$k(\theta|x) = \frac{g(x,\theta)}{g_1(x)} = \frac{L(x|\theta)h(\theta)}{g_1(x)} \tag{5.2}$$

The distribution defined by this conditional distribution is called the **posterior distribution**. The prior distribution reflects the subjective belief of Θ before the

sample is drawn while the posterior distribution is the conditional distribution of Θ after the sample is drawn. Further discussion on these distributions follows an illustrative example.

Example 5.1:

Consider the model

$$X_i|\theta \sim \text{iid } Poisson\ (\theta)$$

$$\Theta \sim \Gamma(\alpha, \beta), \alpha \text{ and } \beta \text{ are known}$$

Hence, the random sample is drawn from a Poisson distribution with mean θ and the prior distribution is $\Gamma(\alpha, \beta)$ distribution. Thus, in this case, the joint conditional pdf of X, given $\Theta = \theta$, is

$$L(x|\theta) = \frac{\theta^{x_1} e^{-\theta}}{x_1!} \ldots . \frac{\theta^{x_n} e^{-\theta}}{x_n!}, x_i = 0, 1, 2, \ldots, i = 1, 2, \ldots ., n,$$

and the prior pdf is

$$h(\theta) = \frac{\theta^{\alpha-1} e^{-\frac{\theta}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}, 0 < \theta < \infty$$

Hence, the joint mixed continuous discrete pdf is given by

$$g(x, \theta) = L(x|\theta)h(\theta) = \left[\frac{\theta^{x_1} e^{-\theta}}{x_1!} \ldots \frac{\theta^{x_n} e^{-\theta}}{x_n!}\right] \left[\frac{\theta^{\alpha-1} e^{-\frac{\theta}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}\right]$$

$$= \frac{\theta^{\sum_{i=1}^{n} x_i + \alpha - 1} e^{-\left(\frac{n\beta+1}{\beta}\right)\theta}}{\prod_{i=1}^{n} x_i! \, \Gamma(\alpha)\beta^{\alpha}}$$

Provided that $x_i = 0, 1, 2, 3, \ldots ., i = 1, 2, \ldots, n$ and $0 < \theta < \infty$. Then, the marginal distribution of the sample, is

$$g_1(x) = \int_0^{\infty} \frac{\theta^{\sum_{i=1}^{n} x_i + \alpha - 1} e^{-\left(\frac{n\beta+1}{\beta}\right)\theta}}{\prod_{i=1}^{n} x_i! \, \Gamma(\alpha)\beta^{\alpha}} d\theta = \frac{\Gamma(\sum_{i=1}^{n} x_i + \alpha)}{\prod_{i=1}^{n} x_i! \, \Gamma(\alpha)\beta^{\alpha} \left(\frac{n\beta + 1}{\beta}\right)^{\sum_{i=1}^{n} x_i + \alpha}}$$

Finally, the posterior pdf of Θ, given $X = x$, is

$$k(\theta|x) = \frac{g(x,\theta)}{g_1(x)} = \frac{\theta^{\sum_{i=1}^{n} x_i+\alpha-1} e^{-\frac{\theta}{\left(\frac{\beta}{n\beta+1}\right)}}}{\Gamma\left(\sum_{i=1}^{n} x_i + \alpha\right)\left(\frac{\beta}{n\beta+1}\right)^{\sum_{i=1}^{n} x_i+\alpha}}$$

Provided that $0 < \theta < \infty$. This conditional pdf is one of the gamma type with parameters $\alpha^* = \sum_{i=1}^{n} x_i + \alpha$ and $\beta^* = \frac{\beta}{n\beta+1}$. Notice that the posterior pdf reflects both prior information (α, β) and sample information $(\sum_{i=1}^{n} x_i)$.

$$\theta|X_i \sim \Gamma\left(\sum_{i=1}^{n} x_i + \alpha, \frac{\beta}{n\beta+1}\right)$$

Remarks:

1. In Example 5.1 notice that it is not really necessary to determine the marginal pdf $g_1(x)$ to find the posterior pdf $k(\theta|x)$. If we divide $L(x|\theta)h(\theta)$ by $g_1(x)$, we must get the product of a factor, which depend upon x but does not depend upon θ, say $c(x)$, That is,

$$k(\theta|x) = c(x)\, \theta^{\sum_{i=1}^{n} x_i+\alpha-1}\, e^{-\frac{\theta}{\left(\frac{\beta}{n\beta+1}\right)}}$$

Provided that $0 < \theta < \infty$, and $x_i = 0,1,2,3, \dots, i = 1,2,\dots,n$. However, $c(x)$ must be that "constant" needed to make $k(\theta|x)$ a pdf, namely

$$c(x) = \frac{1}{\Gamma\left(\sum_{i=1}^{n} x_i + \alpha\right)\left(\frac{\beta}{n\beta+1}\right)^{\sum_{i=1}^{n} x_i+\alpha}}$$

Accordingly, we frequently write that $k(\theta|x)$ is proportional to $L(x|\theta)h(\theta)$; that is, the posterior pdf can be written as

$$k(\theta|x) \propto L(x|\theta)h(\theta) \tag{5.3}$$

Note that in the right-hand member of this expression all factors involving constants and x alone (not θ) can be dropped. For illustration, in solving the problem presented in Example 5.1, we simply write

$$k(\theta|x) \propto \theta^{\sum_{i=1}^{n} x_i+\alpha-1} e^{-\frac{\theta}{\left(\frac{\beta}{n\beta+1}\right)}}$$

$0 < \theta < \infty$. Clearly, $k(\theta|x)$ must be gamma pdf with parameter $\alpha^* = \sum_{i=1}^{n} x_i + \alpha$ and $\beta^* = \frac{\beta}{n\beta+1}$.

2. There is another observation that can be made at this point. Suppose that there exists a sufficient statistic $T = t(X)$ for the parameter so that

$$L(x|\theta) = f_T(t \mid \theta).\, k(X),$$

where now $f_T(t \mid \theta)$ is the pdf of T, given $\Theta = \theta$. Then we note that

$$k(\theta|x) \propto f_T(t \mid \theta)\, h(\theta) \tag{5.4}$$

5.1 Bayesian Point Estimation

Suppose we want a point estimator of θ. From the Bayesian viewpoint, this really amounts to selecting a decision function δ, so that $\delta(x)$ is a predicted value of θ (an experimental value of the random variable Θ) when both the computed value x and the conditional pdf $k(\theta|x)$ are known. Now, in general, how would we predict an experimental value of any random variable, say W, if we want our prediction to be "reasonably close" to the value to be observed?. Many statisticians would predict the mean, $E(W)$, of the distribution of W; others would predict a median (perhaps unique) of the distribution of W, and some would have other predictions. However, it seems desirable that the choice of the decision function should depend upon a loss function $\mathcal{L}[\theta, \delta(x)]$. One way in which this dependence upon the loss function can be reflected is to select the decision function δ in such a way that the conditional expectation of the loss is minimum. A **Bayes' estimate** is a decision function δ that minimizes the expectation of the loss function $E\{\mathcal{L}[\Theta, \delta(x)]|X = x\}$ and then

$$\delta(x) = E\{\mathcal{L}[\Theta, \delta(x)]|X = x\}$$

$$= \begin{cases} \int_\theta \mathcal{L}[\theta, \delta(x)]k(\theta|x)\, d\theta, & \text{if } \Theta \text{ is a continuous} \\ \sum_\theta \mathcal{L}[\theta, \delta(x)]k(\theta|x), & \text{if } \Theta \text{ is a continuous} \end{cases} \tag{5.5}$$

is called **Bayes' estimator** of θ.

Some Possible Loss Functions:

1. **Squared Error Loss Function:**
   The squared error loss function is given by
   $$\mathcal{L}[\theta, \delta(x)] = [\theta - \delta(x)]^2$$
   Then, the Bayes' estimate is the mean of the conditional distribution of Θ, given $X = x$
   $$\delta(x) = E(\Theta|x)$$

2. **Absolute Error Loss Function:**
   The absolute error loss function is given by
   $$\mathcal{L}[\theta, \delta(x)] = |\theta - \delta(x)|$$
   Then, a median of the conditional distribution of Θ, given $X = x$, is the Bayes' solution
   $$\delta(x) = \text{Median of } \Theta$$
   where the median, m, is the solution of
   $$\int_{-\infty}^{m} k(\theta|x) d\theta = \frac{1}{2}$$

It is easy to generalize this to estimate a function of θ, for a specified function $\tau(\theta)$. For the loss function $\mathcal{L}[\theta, \delta(x)]$, a **Bayes estimate** of $\tau(\theta)$ is a decision function δ that minimizes

$$E\{\mathcal{L}[\tau(\Theta), \delta(x)]|X = x\} = \int_{-\infty}^{\infty} \mathcal{L}[\tau(\theta), \delta(x)]\, k(\theta|x) d\theta \qquad (5.6)$$

The random variable $\delta(X)$ is called **Bayes' estimator** of $\tau(\theta)$.

Example 5.2:

Consider the model

$$X_i|\theta \sim \text{iid } Binomial\ (1, \theta)$$

$$\Theta \sim Beta(\alpha, \beta)\ , \alpha \text{ and } \beta \text{ are known}$$

That is, the prior pdf is

$$h(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}\ , 0 < \theta < 1$$

when α and β are assigned positive constants. We seek a decision function δ that is a Bayes' solution. The sufficient statistic is $Y = \sum_1^n X_i$, which has a $Binomial\ (n, \theta)$ distribution. Thus, the conditional pdf of Y given $\Theta = \theta$ is

$$g(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y} \quad y = 0, 1, \ldots, n$$

Thus by Equation (5.4), the conditional posterior pdf of Θ, given $Y = y$ at positive probability density, is

$$k(\theta|y) \propto \theta^y (1-\theta)^{n-y}\, \theta^{\alpha-1}(1-\theta)^{\beta-1}, 0 < \theta < 1$$

That is

$$k(\theta|y) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+y)\Gamma(\beta+n-y)} \theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}, 0 < \theta < 1$$

and $y = 0, 1, \dots, n$. Hence, the posterior pdf is a beta density function with parameters $(\alpha + y, \beta + n - y)$. We take squared error loss, i.e., $\mathcal{L}[\theta, \delta(y)] = [\theta - \delta(y)]^2$, as the loss function. Then, the Bayesian point estimate of θ is the mean of this beta pdf which is

$$\delta(y) = \frac{\alpha + y}{\alpha + \beta + n}$$

5.2 Bayesian Interval Estimation

For fixed α, we can find two functions $u(x)$ and $v(x)$ so that the conditional probability

$$P(u(x) < \Theta < v(x)|X = x) = \int_{u(x)}^{v(x)} k(\theta|x)\, d\theta = 1 - \alpha$$

which is defined to be $100(1-\alpha)\%$ Bayesian interval estimates of θ. This interval is often called **credible interval**, so as not to confuse them with confidence interval.

Example 5.3:

Recall Example 5.1 where $X_1, X_2, \dots, X_n$ is a random sample from a Poisson distribution with mean θ and a $\Gamma(\alpha, \beta)$ prior, with α and β known, is considered. As given, the posterior pdf is a $\Gamma\left(y + \alpha, \frac{\beta}{n\beta+1}\right)$ pdf, where $y = \sum_{i=1}^{n} x_i$, i.e.

$$\theta|X_i \sim \Gamma\left(y + \alpha, \frac{\beta}{n\beta + 1}\right)$$

Find:

a) Bayes' point estimator of Θ using the squared error loss function.
b) Bayes' point estimator of Θ using the absolute error loss function.
c) $(1-\xi)100\%$ credible interval for Θ.
d) a and c when, $\alpha = 2, \beta = 4, n = 12, y = 8, \xi = 0.05$.

Solution:

a) If we use the squared error loss function, the Bayes' point estimate of Θ is the mean of the posterior

$$\delta(y) = \frac{\beta(y+\alpha)}{n\beta+1}$$

b) If we use the absolute error loss function, the Bayes' point estimate of Θ is the median of the posterior or it is the solution, m, of the following equation:

$$\int_0^m \frac{(n\beta+1)^{y+\alpha}\ \theta^{y+\alpha-1} e^{-\frac{(n\beta+1)\theta}{\beta}}}{\Gamma(y+\alpha)\beta^{y+\alpha}} d\theta = \frac{1}{2}$$

c) To obtain a credible interval, from that the posterior distribution of Θ we get that

$$\frac{2(n\beta+1)}{\beta}\Theta \sim \Gamma(y+\alpha,2) \Leftrightarrow \frac{2(n\beta+1)}{\beta}\Theta \sim \chi^2_{(2(y+\alpha))}$$

Based on this, the following interval is a $(1-\xi)100\%$ credible interval for Θ

$$P\left(\chi^2_{(1-\frac{\xi}{2},2(y+\alpha))} < \frac{2(n\beta+1)}{\beta}\Theta < \chi^2_{(\frac{\xi}{2},2(y+\alpha))}\right) = 1-\alpha$$

or $$\Theta \in \left(\frac{\beta}{2(n\beta+1)}\chi^2_{(1-\frac{\xi}{2},2(y+\alpha))}, \frac{\beta}{2(n\beta+1)}\chi^2_{(\frac{\xi}{2},2(y+\alpha))}\right)$$

where $\chi^2_{(1-\frac{\xi}{2},2(y+\alpha))}$ and $\chi^2_{(\frac{\xi}{2},2(y+\alpha))}$ are the lower and upper χ^2 quantiles for a χ^2 distribution with $2(y+\alpha)$ degrees of freedom.

d) If $\alpha = 2, \beta = 4, n = 12, y = 8, \xi = 0.05,$ then the point estimator is

$$\delta(y) = \frac{4(8+2)}{48+1} = 0.8163$$

and the 95% credible interval for Θ is

$$\Theta \in \left(\frac{4}{2(48+1)}\chi^2_{(0.975,20)}, \frac{4}{2(48+1)}\chi^2_{(0.025\ ,20)}\right)$$

From χ^2 distribution Table: $\chi^2_{(0.025\ ,20)} = 34.17$, $\chi^2_{(0.975,20)} = 9.59$, thus

$$\Theta \in (0.3914, 1.3947)$$