Chapter 15

SOME TECHNIQUES FOR FINDING POINT ESTIMATORS OF PARAMETERS

A statistical population consists of all the measurements of interest in a statistical investigation. Usually a population is described by a random variable X. If we can gain some knowledge about the probability density function $f(x;\theta)$ of X, then we also gain some knowledge about the population under investigation.

A sample is a portion of the population usually chosen by method of random sampling and as such it is a set of random variables $X_1, X_2, ..., X_n$ with the same probability density function $f(x;\theta)$ as the population. Once the sampling is done, we get

$$X_1 = x_1,\ X_2 = x_2,\ \cdots,\ X_n = x_n$$

where $x_1, x_2, ..., x_n$ are the *sample data*.

Every statistical method employs a random sample to gain information about the population. Since the population is characterized by the probability density function $f(x;\theta)$, in statistics one makes statistical inferences about the population distribution $f(x;\theta)$ based on sample information. A statistical inference is a statement based on sample information about the population. There are three types of statistical inferences (1) estimation (2) hypothesis testing and (3) prediction. The goal of this chapter is to examine some well known point estimation methods.

In point estimation, we try to find the parameter θ of the population distribution $f(x;\theta)$ from the sample information. Thus, in the parametric point estimation one assumes the functional form of the pdf $f(x;\theta)$ to be known and only estimate the unknown parameter θ of the population using information available from the sample.

Definition 15.1. Let X be a population with the density function $f(x;\theta)$, where θ is an unknown parameter. The set of all admissible values of θ is called a parameter space and it is denoted by Ω, that is

$$\Omega = \{\theta \in \mathbb{R}^n \mid f(x;\theta) \text{ is a pdf } \}$$

for some natural number m.

Example 15.1. If $X \sim EXP(\theta)$, then what is the parameter space of θ ?

Answer: Since $X \sim EXP(\theta)$, the density function of X is given by

$$f(x;\theta) = \frac{1}{\theta}\, e^{-\frac{x}{\theta}}.$$

If θ is zero or negative then $f(x;\theta)$ is not a density function. Thus, the admissible values of θ are all the positive real numbers. Hence

$$\begin{aligned}\Omega &= \{\theta \in \mathbb{R} \mid 0 < \theta < \infty\} \\ &= \mathbb{R}_+.\end{aligned}$$

Example 15.2. If $X \sim N\left(\mu, \sigma^2\right)$, what is the parameter space?

Answer: The parameter space Ω is given by

$$\begin{aligned}\Omega &= \left\{\theta \in \mathbb{R}^2 \mid f(x;\theta) \sim N\left(\mu, \sigma^2\right)\right\} \\ &= \left\{(\mu, \sigma) \in \mathbb{R}^2 \mid -\infty < \mu < \infty,\ 0 < \sigma < \infty\right\} \\ &= \mathbb{R} \times \mathbb{R}_+ \\ &= \text{upper half plane.}\end{aligned}$$

In general, a parameter space is a subset of $\mathbb{R}^m$. Statistics concerns with the estimation of the unknown parameter θ from a random sample $X_1, X_2, ..., X_n$. Recall that a statistic is a function of $X_1, X_2, ..., X_n$ and free of the population parameter θ.

Definition 15.2. Let $X \sim f(x;\theta)$ and $X_1, X_2, ..., X_n$ be a random sample from the population X. Any statistic that can be used to guess the parameter θ is called an estimator of θ. The numerical value of this statistic is called an estimate of θ. The estimator of the parameter θ is denoted by $\widehat{\theta}$.

One of the basic problems is how to find an estimator of population parameter θ. There are several methods for finding an estimator of θ. Some of these methods are:

(1) Moment Method
(2) Maximum Likelihood Method
(3) Bayes Method
(4) Least Squares Method
(5) Minimum Chi-Squares Method
(6) Minimum Distance Method

In this chapter, we only discuss the first three methods of estimating a population parameter.

15.1. Moment Method

Let $X_1, X_2, ..., X_n$ be a random sample from a population X with probability density function $f(x;\theta_1,\theta_2,...,\theta_m)$, where $\theta_1,\theta_2,...,\theta_m$ are m unknown parameters. Let

$$E\left(X^k\right) = \int_{-\infty}^{\infty} x^k\, f(x;\theta_1,\theta_2,...,\theta_m)\, dx$$

be the k^{th} population moment about 0. Further, let

$$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

be the k^{th} sample moment about 0.

In moment method, we find the estimator for the parameters $\theta_1,\theta_2,...,\theta_m$ by equating the first m population moments (if they exist) to the first m sample moments, that is

$$\begin{aligned} E\left(X\right) &= M_1 \\ E\left(X^2\right) &= M_2 \\ E\left(X^3\right) &= M_3 \\ &\vdots \\ E\left(X^m\right) &= M_m \end{aligned}$$

The moment method is one of the classical methods for estimating parameters and motivation comes from the fact that the sample moments are in some sense estimates for the population moments. The moment method was first discovered by British statistician Karl Pearson in 1902. Now we provide some examples to illustrate this method.

Example 15.3. Let $X \sim N\left(\mu, \sigma^2\right)$ and $X_1, X_2, ..., X_n$ be a random sample of size n from the population X. What are the estimators of the population parameters μ and σ^2 if we use the moment method?

Answer: Since the population is normal, that is

$$X \sim N\left(\mu, \sigma^2\right)$$

we know that

$$\begin{aligned} E\left(X\right) &= \mu \\ E\left(X^2\right) &= \sigma^2 + \mu^2. \end{aligned}$$

Hence

$$\begin{aligned} \mu &= E\left(X\right) \\ &= M_1 \\ &= \frac{1}{n} \sum_{i=1}^{n} X_i \\ &= \overline{X}. \end{aligned}$$

Therefore, the estimator of the parameter μ is $\overline{X}$, that is

$$\widehat{\mu} = \overline{X}.$$

Next, we find the estimator of σ^2 equating $E(X^2)$ to M_2. Note that

$$\begin{aligned} \sigma^2 &= \sigma^2 + \mu^2 - \mu^2 \\ &= E\left(X^2\right) - \mu^2 \\ &= M_2 - \mu^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \overline{X}^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2. \end{aligned}$$

The last line follows from the fact that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2-2\,X_i\,\overline{X}+\overline{X}^2\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}X_i^2-\frac{1}{n}\sum_{i=1}^{n}2\,X_i\,\overline{X}+\frac{1}{n}\sum_{i=1}^{n}\overline{X}^2\\ &= \frac{1}{n}\sum_{i=1}^{n}X_i^2-2\,\overline{X}\,\frac{1}{n}\sum_{i=1}^{n}X_i+\overline{X}^2\\ &= \frac{1}{n}\sum_{i=1}^{n}X_i^2-2\,\overline{X}\,\overline{X}+\overline{X}^2\\ &= \frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2.\end{aligned}$$

Thus, the estimator of σ^2 is $\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$, that is

$$\widehat{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2.$$

Example 15.4. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population X with probability density function

$$f(x;\theta)=\begin{cases}\theta\,x^{\theta-1} & \text{if } 0<x<1\\ \\ 0 & \text{otherwise,}\end{cases}$$

where $0<\theta<\infty$ is an unknown parameter. Using the method of moment find an estimator of θ ? If $x_1=0.2, x_2=0.6, x_3=0.5, x_4=0.3$ is a random sample of size 4, then what is the estimate of θ ?

Answer: To find an estimator, we shall equate the population moment to the sample moment. The population moment $E(X)$ is given by

$$\begin{aligned}E(X) &= \int_0^1 x\,f(x;\theta)\,dx\\ &= \int_0^1 x\,\theta\,x^{\theta-1}\,dx\\ &= \theta\int_0^1 x^{\theta}\,dx\\ &= \frac{\theta}{\theta+1}\,\left[x^{\theta+1}\right]_0^1\\ &= \frac{\theta}{\theta+1}.\end{aligned}$$

We know that $M_1 = \overline{X}$. Now setting M_1 equal to $E(X)$ and solving for θ, we get

$$\overline{X} = \frac{\theta}{\theta + 1}$$

that is

$$\theta = \frac{\overline{X}}{1 - \overline{X}},$$

where $\overline{X}$ is the sample mean. Thus, the statistic $\frac{\overline{X}}{1-\overline{X}}$ is an estimator of the parameter θ. Hence

$$\widehat{\theta} = \frac{\overline{X}}{1 - \overline{X}}.$$

Since $x_1 = 0.2, x_2 = 0.6, x_3 = 0.5, x_4 = 0.3$, we have $\overline{X} = 0.4$ and

$$\widehat{\theta} = \frac{0.4}{1 - 0.4} = \frac{2}{3}$$

is an estimate of the θ.

Example 15.5. What is the basic principle of the moment method?

Answer: To choose a value for the unknown population parameter for which the observed data have the same moments as the population.

Example 15.6. Suppose $X_1, X_2, ..., X_7$ is a random sample from a population X with density function

$$f(x;\beta) = \begin{cases} \frac{x^6 \, e^{-\frac{x}{\beta}}}{\Gamma(7)\,\beta^7} & \text{if } 0 < x < \infty \\ \\ 0 & \text{otherwise.} \end{cases}$$

Find an estimator of β by the moment method.

Answer: Since, we have only one parameter, we need to compute only the first population moment $E(X)$ about 0. Thus,

$$\begin{aligned} E(X) &= \int_0^\infty x \, f(x;\beta)\, dx \\ &= \int_0^\infty x \, \frac{x^6 \, e^{-\frac{x}{\beta}}}{\Gamma(7)\,\beta^7}\, dx \\ &= \frac{1}{\Gamma(7)} \int_0^\infty \left(\frac{x}{\beta}\right)^7 e^{-\frac{x}{\beta}}\, dx \\ &= \beta \, \frac{1}{\Gamma(7)} \int_0^\infty y^7 \, e^{-y}\, dy \\ &= \beta \, \frac{1}{\Gamma(7)} \, \Gamma(8) \\ &= 7\,\beta. \end{aligned}$$

Since $M_1 = \overline{X}$, equating $E(X)$ to M_1, we get

$$7\,\beta = \overline{X}$$

that is

$$\beta = \frac{1}{7}\,\overline{X}.$$

Therefore, the estimator of β by the moment method is given by

$$\widehat{\beta} = \frac{1}{7}\,\overline{X}.$$

Example 15.7. Suppose $X_1, X_2, ..., X_n$ is a random sample from a population X with density function

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ \\ 0 & \text{otherwise.} \end{cases}$$

Find an estimator of θ by the moment method.

Answer: Examining the density function of the population X, we see that $X \sim UNIF(0, \theta)$. Therefore

$$E(X) = \frac{\theta}{2}.$$

Now, equating this population moment to the sample moment, we obtain

$$\frac{\theta}{2} = E(X) = M_1 = \overline{X}.$$

Therefore, the estimator of θ is

$$\widehat{\theta} = 2\,\overline{X}.$$

Example 15.8. Suppose $X_1, X_2, ..., X_n$ is a random sample from a population X with density function

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } \alpha < x < \beta \\ \\ 0 & \text{otherwise.} \end{cases}$$

Find the estimators of α and β by the moment method.

Answer: Examining the density function of the population X, we see that $X \sim UNIF(\alpha, \beta)$. Since, the distribution has two unknown parameters, we need the first two population moments. Therefore

$$E(X) = \frac{\alpha+\beta}{2} \qquad \text{and} \qquad E(X^2) = \frac{(\beta-\alpha)^2}{12} + E(X)^2.$$

Equating these moments to the corresponding sample moments, we obtain

$$\frac{\alpha+\beta}{2} = E(X) = M_1 = \overline{X}$$

that is

$$\alpha + \beta = 2\overline{X} \tag{1}$$

and

$$\frac{(\beta-\alpha)^2}{12} + E(X)^2 = E(X^2) = M_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

which is

$$\begin{aligned}(\beta-\alpha)^2 &= 12\,\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - E(X)^2\right]\\ &= 12\,\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right]\\ &= 12\,\left[\frac{1}{n}\sum_{i=1}^{n} \left(X_i^2 - \overline{X}\right)^2\right].\end{aligned}$$

Hence, we get

$$\beta - \alpha = \sqrt{\frac{12}{n}\sum_{i=1}^{n} \left(X_i^2 - \overline{X}\right)^2}. \tag{2}$$

Adding equation (1) to equation (2), we obtain

$$2\beta = 2\overline{X} \pm 2\sqrt{\frac{3}{n}\sum_{i=1}^{n} \left(X_i^2 - \overline{X}\right)^2}$$

that is

$$\beta = \overline{X} \pm \sqrt{\frac{3}{n}\sum_{i=1}^{n} \left(X_i^2 - \overline{X}\right)^2}.$$

Similarly, subtracting (2) from (1), we get

$$\alpha = \overline{X} \mp \sqrt{\frac{3}{n}\sum_{i=1}^{n} \left(X_i^2 - \overline{X}\right)^2}.$$

Since, $\alpha < \beta$, we see that the estimators of α and β are

$$\widehat{\alpha} = \overline{X} - \sqrt{\frac{3}{n} \sum_{i=1}^{n} \left(X_i^2 - \overline{X}\right)^2} \qquad \text{and} \qquad \widehat{\beta} = \overline{X} + \sqrt{\frac{3}{n} \sum_{i=1}^{n} \left(X_i^2 - \overline{X}\right)^2}.$$

15.2. Maximum Likelihood Method

The maximum likelihood method was first used by Sir Ronald Fisher in 1912 for finding estimator of a unknown parameter. However, the method originated in the works of Gauss and Bernoulli. Next, we describe the method in detail.

Definition 15.3. Let $X_1, X_2, ..., X_n$ be a random sample from a population X with probability density function $f(x; \theta)$, where θ is an unknown parameter. The likelihood function, $L(\theta)$, is the distribution of the sample. That is

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

This definition says that the likelihood function of a random sample $X_1, X_2, ..., X_n$ is the joint density of the random variables $X_1, X_2, ..., X_n$.

The θ that maximizes the likelihood function $L(\theta)$ is called the maximum likelihood estimator of θ, and it is denoted by $\widehat{\theta}$. Hence

$$\widehat{\theta} = Arg \sup_{\theta \in \Omega} L(\theta),$$

where Ω is the parameter space of θ so that $L(\theta)$ is the joint density of the sample.

The method of maximum likelihood in a sense picks out of all the possible values of θ the one most likely to have produced the given observations $x_1, x_2, ..., x_n$. The method is summarized below: (1) Obtain a random sample $x_1, x_2, ..., x_n$ from the distribution of a population X with probability density function $f(x; \theta)$; (2) define the likelihood function for the sample $x_1, x_2, ..., x_n$ by $L(\theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)$; (3) find the expression for θ that maximizes $L(\theta)$. This can be done directly or by maximizing $\ln L(\theta)$; (4) replace θ by $\widehat{\theta}$ to obtain an expression for the maximum likelihood estimator for θ; (5) find the observed value of this estimator for a given sample.

Example 15.9. If $X_1, X_2, ..., X_n$ is a random sample from a distribution with density function

$$f(x;\theta) = \begin{cases} (1-\theta)\, x^{-\theta} & \text{if } 0 < x < 1 \\ \\ 0 & \text{elsewhere,} \end{cases}$$

what is the maximum likelihood estimator of θ ?

Answer: The likelihood function of the sample is given by

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

Therefore

$$\begin{aligned} \ln L(\theta) &= \ln \left(\prod_{i=1}^{n} f(x_i;\theta) \right) \\ &= \sum_{i=1}^{n} \ln f(x_i;\theta) \\ &= \sum_{i=1}^{n} \ln \left[(1-\theta)\, {x_i}^{-\theta} \right] \\ &= n\, \ln(1-\theta) - \theta \sum_{i=1}^{n} \ln x_i. \end{aligned}$$

Now we maximize $\ln L(\theta)$ with respect to θ.

$$\begin{aligned} \frac{d \ln L(\theta)}{d\theta} &= \frac{d}{d\theta} \left(n\, \ln(1-\theta) - \theta \sum_{i=1}^{n} \ln x_i \right) \\ &= -\frac{n}{1-\theta} - \sum_{i=1}^{n} \ln x_i. \end{aligned}$$

Setting this derivative $\frac{d \ln L(\theta)}{d\theta}$ to 0, we get

$$\frac{d \ln L(\theta)}{d\theta} = -\frac{n}{1-\theta} - \sum_{i=1}^{n} \ln x_i = 0$$

that is

$$\frac{1}{1-\theta} = -\frac{1}{n} \sum_{i=1}^{n} \ln x_i$$

or
$$\frac{1}{1-\theta} = -\frac{1}{n} \sum_{i=1}^{n} \ln x_i = -\overline{\ln x}.$$
or
$$\theta = 1 + \frac{1}{\overline{\ln x}}.$$
This θ can be shown to be maximum by the second derivative test and we leave this verification to the reader. Therefore, the estimator of θ is
$$\widehat{\theta} = 1 + \frac{1}{\overline{\ln X}}.$$

Example 15.10. If $X_1, X_2, ..., X_n$ is a random sample from a distribution with density function
$$f(x;\beta) = \begin{cases} \frac{x^6 \, e^{-\frac{x}{\beta}}}{\Gamma(7)\,\beta^7} & \text{if } 0 < x < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$
then what is the maximum likelihood estimator of β ?

Answer: The likelihood function of the sample is given by
$$L(\beta) = \prod_{i=1}^{n} f(x_i;\beta).$$

Thus,
$$\begin{aligned} \ln L(\beta) &= \sum_{i=1}^{n} \ln f(x_i, \beta) \\ &= 6 \sum_{i=1}^{n} \ln x_i - \frac{1}{\beta} \sum_{i=1}^{n} x_i - n \ln(6!) - 7n \ln(\beta). \end{aligned}$$
Therefore
$$\frac{d}{d\beta} \ln L(\beta) = \frac{1}{\beta^2} \sum_{i=1}^{n} x_i - \frac{7n}{\beta}.$$
Setting this derivative $\frac{d}{d\beta} \ln L(\beta)$ to zero, we get
$$\frac{1}{\beta^2} \sum_{i=1}^{n} x_i - \frac{7n}{\beta} = 0$$
which yields
$$\beta = \frac{1}{7n} \sum_{i=1}^{n} x_i.$$

This β can be shown to be maximum by the second derivative test and again we leave this verification to the reader. Hence, the estimator of β is given by

$$\widehat{\beta} = \frac{1}{7}\,\overline{X}.$$

Remark 15.1. Note that this maximum likelihood estimator of β is same as the one found for β using the moment method in Example 15.6. However, in general the estimators by different methods are different as the following example illustrates.

Example 15.11. If $X_1, X_2, ..., X_n$ is a random sample from a distribution with density function

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ \\ 0 & \text{otherwise,} \end{cases}$$

then what is the maximum likelihood estimator of θ ?

Answer: The likelihood function of the sample is given by

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\ &= \prod_{i=1}^{n} \left(\frac{1}{\theta}\right) \qquad\qquad \theta > x_i \quad (i = 1, 2, 3, ..., n) \\ &= \left(\frac{1}{\theta}\right)^n \qquad\qquad \theta > \max\{x_1, x_2, ..., x_n\}. \end{aligned}$$

Hence the parameter space of θ with respect to $L(\theta)$ is given by

$$\Omega = \{\theta \in \mathbb{R} \mid x_{\max} < \theta < \infty\} = (x_{\max},\, \infty).$$

Now we maximize $L(\theta)$ on Ω. First, we compute $\ln L(\theta)$ and then differentiate it to get

$$\ln L(\theta) = -n \ln(\theta)$$

and

$$\frac{d}{d\theta} \ln L(\theta) = -\frac{n}{\theta} < 0.$$

Therefore $\ln L(\theta)$ is a decreasing function of θ and as such the maximum of $\ln L(\theta)$ occurs at the left end point of the interval $(x_{\max},\, \infty)$. Therefore, at

$\theta = x_{\max}$ the likelihood function achieve maximum. Hence the likelihood estimator of θ is given by

$$\widehat{\theta} = X_{(n)}$$

where $X_{(n)}$ denotes the n^{th} order statistic of the given sample.

Thus, Example 15.7 and Example 15.11 say that the if we estimate the parameter θ of a distribution with uniform density on the interval $(0, \theta)$, then the maximum likelihood estimator is given by

$$\widehat{\theta} = X_{(n)}$$

where as

$$\widehat{\theta} = 2\,\overline{X}$$

is the estimator obtained by the method of moment. Hence, in general these two methods do not provide the same estimator of an unknown parameter.

Example 15.12. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density function

$$f(x;\theta) = \begin{cases} \sqrt{\frac{2}{\pi}}\; e^{-\frac{1}{2}\,(x-\theta)^2} & \text{if } x \geq \theta \\ \\ 0 & \text{elsewhere.} \end{cases}$$

What is the maximum likelihood estimator of θ ?

Answer: The likelihood function $L(\theta)$ is given by

$$L(\theta) = \left(\sqrt{\frac{2}{\pi}}\right)^n \prod_{i=1}^{n} e^{-\frac{1}{2}\,(x_i-\theta)^2} \qquad\qquad x_i \geq \theta\ (i = 1, 2, 3, ..., n).$$

Hence the parameter space of θ is given by

$$\Omega = \{\theta \in \mathbb{R} \mid 0 \leq \theta \leq x_{\min}\} = [0, x_{\min}],,$$

where $x_{\min} = \min\{x_1, x_2, ..., x_n\}$. Now we evaluate the logarithm of the likelihood function.

$$\ln L(\theta) = \frac{n}{2}\,\ln\left(\frac{2}{\pi}\right) - \frac{1}{2}\,\sum_{i=1}^{n}(x_i-\theta)^2,$$

where θ is on the interval $[0,\, x_{\min}\,]$. Now we maximize $\ln L(\theta)$ subject to the condition $0 \leq \theta \leq x_{\min}$. Taking the derivative, we get

$$\frac{d}{d\theta}\ln L(\theta) = -\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)\,2(-1) = \sum_{i=1}^{n}(x_i-\theta).$$

In this example, if we equate the derivative to zero, then we get $\theta = \overline{x}$. But this value of θ is not on the parameter space Ω. Thus, $\theta = \overline{x}$ is not the solution. Hence to find the solution of this optimization process, we examine the behavior of the $\ln L(\theta)$ on the interval $[0,\, x_{\min}\,]$. Note that

$$\frac{d}{d\theta} \ln L(\theta) = -\frac{1}{2} \sum_{i=1}^{n} (x_i - \theta)\, 2(-1) = \sum_{i=1}^{n} (x_i - \theta) > 0$$

since each x_i is bigger than θ. Therefore, the function $\ln L(\theta)$ is an increasing function on the interval $[0,\, x_{\min}\,]$ and as such it will achieve maximum at the right end point of the interval $[0,\, x_{\min}\,]$. Therefore, the maximum likelihood estimator of θ is given by

$$\widehat{X} = X_{(1)}$$

where $X_{(1)}$ denotes the smallest observation in the random sample $X_1, X_2, ..., X_n$.

Example 15.13. Let $X_1, X_2, ..., X_n$ be a random sample from a normal population with mean μ and variance σ^2. What are the maximum likelihood estimators of μ and σ^2 ?

Answer: Since $X \sim N(\mu, \sigma^2)$, the probability density function of X is given by

$$f(x; \mu, \sigma) = \frac{1}{\sigma\, \sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

The likelihood function of the sample is given by

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\, \sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}.$$

Hence, the logarithm of this likelihood function is given by

$$\ln L(\mu, \sigma) = -\frac{n}{2}\, \ln(2\pi) - n \ln(\sigma) - \frac{1}{2\sigma^2}\, \sum_{i=1}^{n} (x_i - \mu)^2.$$

Taking the partial derivatives of $\ln L(\mu, \sigma)$ with respect to μ and σ, we get

$$\frac{\partial}{\partial \mu} \ln L(\mu, \sigma) = -\frac{1}{2\sigma^2}\, \sum_{i=1}^{n} (x_i - \mu)\, (-2) = \frac{1}{\sigma^2}\, \sum_{i=1}^{n} (x_i - \mu).$$

and

$$\frac{\partial}{\partial \sigma} \ln L(\mu, \sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\, \sum_{i=1}^{n} (x_i - \mu)^2.$$

Setting $\frac{\partial}{\partial \mu} \ln L(\mu, \sigma) = 0$ and $\frac{\partial}{\partial \sigma} \ln L(\mu, \sigma) = 0$, and solving for the unknown μ and σ, we get

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}.$$

Thus the maximum likelihood estimator of μ is

$$\widehat{\mu} = \overline{X}.$$

Similarly, we get

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

implies

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2.$$

Again μ and σ^2 found by the first derivative test can be shown to be maximum using the second derivative test for the functions of two variables. Hence, using the estimator of μ in the above expression, we get the estimator of σ^2 to be

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Example 15.14. Suppose $X_1, X_2, ..., X_n$ is a random sample from a distribution with density function

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta \\ \\ 0 & \text{otherwise.} \end{cases}$$

Find the estimators of α and β by the method of maximum likelihood.

Answer: The likelihood function of the sample is given by

$$L(\alpha, \beta) = \prod_{i=1}^{n} \frac{1}{\beta - \alpha} = \left(\frac{1}{\beta - \alpha} \right)^n$$

for all $\alpha \leq x_i$ for $(i = 1, 2, ..., n)$ and for all $\beta \geq x_i$ for $(i = 1, 2, ..., n)$. Hence, the domain of the likelihood function is

$$\Omega = \{(\alpha, \beta) \mid 0 < \alpha \leq x_{(1)} \quad \text{and} \quad x_{(n)} \leq \beta < \infty\}.$$

It is easy to see that $L(\alpha, \beta)$ is maximum if $\alpha = x_{(1)}$ and $\beta = x_{(n)}$. Therefore, the maximum likelihood estimator of α and β are

$$\widehat{\alpha} = X_{(1)} \qquad \text{and} \qquad \widehat{\beta} = X_{(n)}.$$

The maximum likelihood estimator $\widehat{\theta}$ of a parameter θ has a remarkable property known as the invariance property. This invariance property says that if $\widehat{\theta}$ is a maximum likelihood estimator of θ, then $g(\widehat{\theta})$ is the maximum likelihood estimator of $g(\theta)$, where g is a function from $\mathbb{R}^k$ to a subset of $\mathbb{R}^m$. This result was proved by Zehna in 1966. We state this result as a theorem without a proof.

Theorem 15.1. Let $\widehat{\theta}$ be a maximum likelihood estimator of a parameter θ and let $g(\theta)$ be a function of θ. Then the maximum likelihood estimator of $g(\theta)$ is given by $g\left(\widehat{\theta}\right)$.

Now we give two examples to illustrate the importance of this theorem.

Example 15.15. Let $X_1, X_2, ..., X_n$ be a random sample from a normal population with mean μ and variance σ^2. What are the maximum likelihood estimators of σ and $\mu - \sigma$?

Answer: From Example 15.13, we have the maximum likelihood estimator of μ and σ^2 to be

$$\widehat{\mu} = \overline{X}$$

and

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 =: \Sigma^2 \text{ (say)}.$$

Now using the invariance property of the maximum likelihood estimator we have

$$\widehat{\sigma} = \Sigma$$

and

$$\widehat{\mu - \sigma} = \overline{X} - \Sigma.$$

Example 15.16. Suppose $X_1, X_2, ..., X_n$ is a random sample from a distribution with density function

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta \\ \\ 0 & \text{otherwise.} \end{cases}$$

Find the estimator of $\sqrt{\alpha^2 + \beta^2}$ by the method of maximum likelihood.

Answer: From Example 15.14, we have the maximum likelihood estimator of α and β to be

$$\widehat{\alpha} = X_{(1)} \qquad \text{and} \qquad \widehat{\beta} = X_{(n)},$$

respectively. Now using the invariance property of the maximum likelihood estimator we see that the maximum likelihood estimator of $\sqrt{\alpha^2 + \beta^2}$ is $\sqrt{X_{(1)}^2 + X_{(n)}^2}$.

The concept of information in statistics was introduced by Sir Ronald Fisher, and it is known as Fisher information.

Definition 15.4. Let X be an observation from a population with probability density function $f(x;\theta)$. Suppose $f(x;\theta)$ is continuous, twice differentiable and it's support does not depend on θ. Then the Fisher information, $I(\theta)$, in a single observation X about θ is given by

$$I(\theta) = \int_{-\infty}^{\infty} \left[\frac{d \ln f(x;\theta)}{d\theta} \right]^2 f(x;\theta)\, dx.$$

Thus $I(\theta)$ is the expected value of the square of the random variable $\frac{d \ln f(X;\theta)}{d\theta}$. That is,

$$I(\theta) = E\left(\left[\frac{d \ln f(X;\theta)}{d\theta} \right]^2 \right).$$

In the following lemma, we give an alternative formula for the Fisher information.

Lemma 15.1. The Fisher information contained in a single observation about the unknown parameter θ can be given alternatively as

$$I(\theta) = -\int_{-\infty}^{\infty} \left[\frac{d^2 \ln f(x;\theta)}{d\theta^2} \right] f(x;\theta)\, dx.$$

Proof: Since $f(x;\theta)$ is a probability density function,

$$\int_{-\infty}^{\infty} f(x;\theta)\, dx = 1. \tag{3}$$

Differentiating (3) with respect to θ, we get

$$\frac{d}{d\theta} \int_{-\infty}^{\infty} f(x;\theta)\, dx = 0.$$

Rewriting the last equality, we obtain

$$\int_{-\infty}^{\infty} \frac{df(x;\theta)}{d\theta} \, \frac{1}{f(x;\theta)} \, f(x;\theta)\, dx = 0$$

which is

$$\int_{-\infty}^{\infty} \frac{d \ln f(x;\theta)}{d\theta} \, f(x;\theta)\, dx = 0. \tag{4}$$

Now differentiating (4) with respect to θ, we see that

$$\int_{-\infty}^{\infty} \left[\frac{d^2 \, \ln f(x;\theta)}{d\theta^2} \, f(x;\theta) + \frac{d \ln f(x;\theta)}{d\theta} \, \frac{df(x;\theta)}{d\theta} \right] \, dx = 0.$$

Rewriting the last equality, we have

$$\int_{-\infty}^{\infty} \left[\frac{d^2 \, \ln f(x;\theta)}{d\theta^2} \, f(x;\theta) + \frac{d \ln f(x;\theta)}{d\theta} \, \frac{df(x;\theta)}{d\theta} \frac{1}{f(x;\theta)} \, f(x;\theta) \right] \, dx = 0$$

which is

$$\int_{-\infty}^{\infty} \left(\frac{d^2 \, \ln f(x;\theta)}{d\theta^2} + \left[\frac{d \ln f(x;\theta)}{d\theta} \right]^2 \right) \, f(x;\theta)\, dx = 0.$$

The last equality implies that

$$\int_{-\infty}^{\infty} \left[\frac{d \, \ln f(x;\theta)}{d\theta} \right]^2 f(x;\theta)\, dx = -\int_{-\infty}^{\infty} \left[\frac{d^2 \, \ln f(x;\theta)}{d\theta^2} \right] f(x;\theta)\, dx.$$

Hence using the definition of Fisher information, we have

$$I(\theta) = -\int_{-\infty}^{\infty} \left[\frac{d^2 \, \ln f(x;\theta)}{d\theta^2} \right] f(x;\theta)\, dx$$

and the proof of the lemma is now complete.

The following two examples illustrate how one can determine Fisher information.

Example 15.17. Let X be a single observation taken from a normal population with unknown mean μ and known variance σ^2. Find the Fisher information in a single observation X about μ.

Answer: Since $X \sim N(\mu, \sigma^2)$, the probability density of X is given by

$$f(x;\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

Hence

$$\ln f(x;\mu) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}.$$

Therefore

$$\frac{d\ln f(x;\mu)}{d\mu} = \frac{x-\mu}{\sigma^2}$$

and

$$\frac{d^2\ln f(x;\mu)}{d\mu^2} = -\frac{1}{\sigma^2}.$$

Hence

$$I(\mu) = -\int_{-\infty}^{\infty}\left(-\frac{1}{\sigma^2}\right) f(x;\mu)\,dx = \frac{1}{\sigma^2}.$$

Example 15.18. Let $X_1, X_2, ..., X_n$ be a random sample from a normal population with unknown mean μ and known variance σ^2. Find the Fisher information in this sample of size n about μ.

Answer: Let $I_n(\mu)$ be the required Fisher information. Then from the definition, we have

$$\begin{aligned}
I_n(\mu) &= -\,E\left(\frac{d^2\ln f(X_1, X_2, ..., X_n;\mu)}{d\mu^2}\right)\\
&= -\,E\left(\frac{d^2}{d\mu^2}\left\{\ln f(X_1;\mu) + \cdots + \ln f(X_n;\mu)\right\}\right)\\
&= -\,E\left(\frac{d^2\ln f(X_1;\mu)}{d\mu^2}\right) - \cdots -\, E\left(\frac{d^2\ln f(X_n;\mu)}{d\mu^2}\right)\\
&= I(\mu) + \cdots + I(\mu)\\
&= n\,I(\mu)\\
&= n\,\frac{1}{\sigma^2} \qquad\qquad \text{(using Example 15.17).}
\end{aligned}$$

This example shows that if $X_1, X_2, ..., X_n$ is a random sample from a population $X \sim f(x;\theta)$, then the Fisher information, $I_n(\theta)$, in a sample of size n about the parameter θ is equal to n times the Fisher information in X about θ. Thus

$$I_n(\theta) = n\,I(\theta).$$

If X is a random variable with probability density function $f(x;\theta)$, where $\theta = (\theta_1, ..., \theta_n)$ is an unknown parameter vector then the Fisher information,

$I(\theta)$, is a $n \times n$ matrix given by

$$\begin{aligned} I(\theta) &= (I_{ij}(\theta)) \\ &= \left(-E \left(\frac{\partial^2 \ln f(X;\theta)}{\partial \theta_i \, \partial \theta_j} \right) \right). \end{aligned}$$

Example 15.19. Let $X_1, X_2, ..., X_n$ be a random sample from a normal population with mean μ and variance σ^2. What is the Fisher information matrix, $I_n(\mu, \sigma^2)$, of the sample of size n about the parameters μ and σ^2?

Answer: Let us write $\theta_1 = \mu$ and $\theta_2 = \sigma^2$. The Fisher information, $I_n(\theta)$, in a sample of size n about the parameter (θ_1, θ_2) is equal to n times the Fisher information in the population about (θ_1, θ_2), that is

$$I_n(\theta_1, \theta_2) = n \, I(\theta_1, \theta_2). \tag{5}$$

Since there are two parameters θ_1 and θ_2, the Fisher information matrix $I(\theta_1, \theta_2)$ is a 2×2 matrix given by

$$I(\theta_1, \theta_2) = \begin{pmatrix} I_{11}(\theta_1, \theta_2) & I_{12}(\theta_1, \theta_2) \\ \\ I_{21}(\theta_1, \theta_2) & I_{22}(\theta_1, \theta_2) \end{pmatrix} \tag{6}$$

where

$$I_{ij}(\theta_1, \theta_2) = -E \left(\frac{\partial^2 \ln f(X; \theta_1, \theta_2)}{\partial \theta_i \, \partial \theta_j} \right)$$

for $i = 1, 2$ and $j = 1, 2$. Now we proceed to compute I_{ij}. Since

$$f(x; \theta_1, \theta_2) = \frac{1}{\sqrt{2 \, \pi \, \theta_2}} \, e^{-\frac{(x-\theta_1)^2}{2\,\theta_2}}$$

we have

$$\ln f(x; \theta_1, \theta_2) = -\frac{1}{2} \, \ln(2 \, \pi \, \theta_2) - \frac{(x - \theta_1)^2}{2 \, \theta_2}.$$

Taking partials of $\ln f(x; \theta_1, \theta_2)$, we have

$$\begin{aligned} \frac{\partial \ln f(x; \theta_1, \theta_2)}{\partial \theta_1} &= \frac{x - \theta_1}{\theta_2}, \\ \frac{\partial \ln f(x; \theta_1, \theta_2)}{\partial \theta_2} &= -\frac{1}{2 \, \theta_2} + \frac{(x - \theta_1)^2}{2 \, \theta_2^2}, \\ \frac{\partial^2 \ln f(x; \theta_1, \theta_2)}{\partial \theta_1^2} &= -\frac{1}{\theta_2}, \\ \frac{\partial^2 \ln f(x; \theta_1, \theta_2)}{\partial \theta_2^2} &= \frac{1}{2 \, \theta_2^2} - \frac{(x - \theta_1)^2}{\theta_2^3}, \\ \frac{\partial^2 \ln f(x; \theta_1, \theta_2)}{\partial \theta_1 \, \partial \theta_2} &= -\frac{x - \theta_1}{\theta_2^2}. \end{aligned}$$

Hence

$$I_{11}(\theta_1,\theta_2) = -E\left(-\frac{1}{\theta_2}\right) = \frac{1}{\theta_2} = \frac{1}{\sigma^2}.$$

Similarly,

$$I_{21}(\theta_1,\theta_2) = I_{12}(\theta_1,\theta_2) = -E\left(-\frac{X-\theta_1}{\theta_2^2}\right) = \frac{E(X)}{\theta_2^2} - \frac{\theta_1}{\theta_2^2} = \frac{\theta_1}{\theta_2^2} - \frac{\theta_1}{\theta_2^2} = 0$$

and

$$\begin{aligned} I_{22}(\theta_1,\theta_2) &= -E\left(-\frac{(X-\theta_1)^2}{\theta_2^3} + \frac{1}{2\theta_2^2}\right) \\ &= \frac{E\left((X-\theta_1)^2\right)}{\theta_2^3} - \frac{1}{2\theta_2^2} = \frac{\theta_2}{\theta_2^3} - \frac{1}{2\theta_2^2} = \frac{1}{2\theta_2^2} = \frac{1}{2\sigma^4}. \end{aligned}$$

Thus from (5), (6) and the above calculations, the Fisher information matrix is given by

$$I_n(\theta_1,\theta_2) = n \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix} = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}.$$

Now we present an important theorem about the maximum likelihood estimator without a proof.

Theorem 15.2. Under certain regularity conditions on the $f(x;\theta)$ the maximum likelihood estimator $\widehat{\theta}$ of θ based on a random sample of size n from a population X with probability density $f(x;\theta)$ is asymptotically normally distributed with mean θ and variance $\frac{1}{n\,I(\theta)}$. That is

$$\widehat{\theta}_{ML} \sim N\left(\theta,\, \frac{1}{n\,I(\theta)}\right) \qquad \text{as } n \to \infty.$$

The following example shows that the maximum likelihood estimator of a parameter is not necessarily unique.

Example 15.20. If $X_1, X_2, ..., X_n$ is a random sample from a distribution with density function

$$f(x;\theta) = \begin{cases} \frac{1}{2} & \text{if } \theta - 1 \leq x \leq \theta + 1 \\ 0 & \text{otherwise,} \end{cases}$$

then what is the maximum likelihood estimator of θ?

Answer: The likelihood function of this sample is given by

$$L(\theta) = \begin{cases} \left(\frac{1}{2}\right)^n & \text{if } \max\{x_1, ..., x_n\} - 1 \leq \theta \leq \min\{x_1, ..., x_n\} + 1 \\ \\ 0 & \text{otherwise.} \end{cases}$$

Since the likelihood function is a constant, any value in the interval $[\max\{x_1, ..., x_n\} - 1,\ \min\{x_1, ..., x_n\} + 1]$ is a maximum likelihood estimate of θ.

Example 15.21. What is the basic principle of maximum likelihood estimation?

Answer: To choose a value of the parameter for which the observed data have as high a probability or density as possible. In other words a maximum likelihood estimate is a parameter value under which the sample data have the highest probability.

15.3. Bayesian Method

In the classical approach, the parameter θ is assumed to be an unknown, but fixed quantity. A random sample $X_1, X_2, ..., X_n$ is drawn from a population with probability density function $f(x; \theta)$ and based on the observed values in the sample, knowledge about the value of θ is obtained.

In Bayesian approach θ is considered to be a quantity whose variation can be described by a probability distribution (known as the prior distribution). This is a subjective distribution, based on the experimenter's belief, and is formulated before the data are seen (and hence the name prior distribution). A sample is then taken from a population where θ is a parameter and the prior distribution is updated with this sample information. This updated prior is called the posterior distribution. The updating is done with the help of Bayes' theorem and hence the name Bayesian method.

In this section, we shall denote the population density $f(x; \theta)$ as $f(x/\theta)$, that is the density of the population X given the parameter θ.

Definition 15.5. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density $f(x/\theta)$, where θ is the unknown parameter to be estimated. The probability density function of the random variable θ is called the prior distribution of θ and usually denoted by $h(\theta)$.

Definition 15.6. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density $f(x/\theta)$, where θ is the unknown parameter to be estimated. The conditional density, $k(\theta/x_1, x_2, ..., x_n)$, of θ given the sample $x_1, x_2, ..., x_n$ is called the posterior distribution of θ.

Example 15.22. Let $X_1 = 1$, $X_2 = 2$ be a random sample of size 2 from a distribution with probability density function

$$f(x/\theta) = \binom{3}{x} \theta^x (1-\theta)^{3-x}, \qquad x = 0, 1, 2, 3.$$

If the prior density of θ is

$$h(\theta) = \begin{cases} k & \text{if } \frac{1}{2} < \theta < 1 \\ \\ 0 & \text{otherwise,} \end{cases}$$

what is the posterior distribution of θ ?

Answer: Since $h(\theta)$ is the probability density of θ, we should get

$$\int_{\frac{1}{2}}^{1} h(\theta)\, d\theta = 1$$

which implies

$$\int_{\frac{1}{2}}^{1} k\, d\theta = 1.$$

Therefore $k = 2$. The joint density of the sample and the parameter is given by

$$\begin{aligned} u(x_1, x_2, \theta) &= f(x_1/\theta) f(x_2/\theta) h(\theta) \\ &= \binom{3}{x_1} \theta^{x_1} (1-\theta)^{3-x_1} \binom{3}{x_2} \theta^{x_2} (1-\theta)^{3-x_2}\, 2 \\ &= 2 \binom{3}{x_1} \binom{3}{x_2} \theta^{x_1+x_2} (1-\theta)^{6-x_1-x_2}. \end{aligned}$$

Hence,

$$\begin{aligned} u(1, 2, \theta) &= 2 \binom{3}{1} \binom{3}{2} \theta^3 (1-\theta)^3 \\ &= 18\, \theta^3 (1-\theta)^3. \end{aligned}$$

The marginal distribution of the sample

$$
\begin{aligned}
g(1,2) &= \int_{\frac{1}{2}}^{1} u(1,2,\theta)\, d\theta \\
&= \int_{\frac{1}{2}}^{1} 18\,\theta^3 (1-\theta)^3\, d\theta \\
&= 18 \int_{\frac{1}{2}}^{1} \theta^3 \left(1 + 3\theta^2 - 3\theta - \theta^3\right)\, d\theta \\
&= 18 \int_{\frac{1}{2}}^{1} \left(\theta^3 + 3\theta^5 - 3\theta^4 - \theta^6\right)\, d\theta \\
&= \frac{9}{140}.
\end{aligned}
$$

The conditional distribution of the parameter θ given the sample $X_1 = 1$ and $X_2 = 2$ is given by

$$
\begin{aligned}
k(\theta/x_1 = 1, x_2 = 2) &= \frac{u(1,2,\theta)}{g(1,2)} \\
&= \frac{18\,\theta^3\,(1-\theta)^3}{\frac{9}{140}} \\
&= 280\,\theta^3\,(1-\theta)^3.
\end{aligned}
$$

Therefore, the posterior distribution of θ is

$$
k(\theta/x_1 = 1, x_2 = 2) = \begin{cases} 280\,\theta^3\,(1-\theta)^3 & \text{if } \frac{1}{2} < \theta < 1 \\ \\ 0 & \text{otherwise.} \end{cases}
$$

Remark 15.2. If $X_1, X_2, ..., X_n$ is a random sample from a population with density $f(x/\theta)$, then the joint density of the sample and the parameter is given by

$$
u(x_1, x_2, ..., x_n, \theta) = h(\theta) \prod_{i=1}^{n} f(x_i/\theta).
$$

Given this joint density, the marginal density of the sample can be computed using the formula

$$
g(x_1, x_2, ..., x_n) = \int_{-\infty}^{\infty} h(\theta) \prod_{i=1}^{n} f(x_i/\theta)\, d\theta.
$$

Now using the Bayes rule, the posterior distribution of θ can be computed as follows:

$$k(\theta/x_1, x_2, ..., x_n) = \frac{h(\theta) \prod_{i=1}^{n} f(x_i/\theta)}{\int_{-\infty}^{\infty} h(\theta) \prod_{i=1}^{n} f(x_i/\theta)\, d\theta}.$$

In Bayesian method, we use two types of loss functions.

Definition 15.7. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density $f(x/\theta)$, where θ is the unknown parameter to be estimated. Let $\widehat{\theta}$ be an estimator of θ. The function

$$\mathcal{L}_2\left(\widehat{\theta}, \theta\right) = \left(\widehat{\theta} - \theta\right)^2$$

is called the squared error loss. The function

$$\mathcal{L}_1\left(\widehat{\theta}, \theta\right) = \left|\widehat{\theta} - \theta\right|$$

is called the absolute error loss.

The loss function $\mathcal{L}$ represents the 'loss' incurred when $\widehat{\theta}$ is used in place of the parameter θ.

Definition 15.8. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density $f(x/\theta)$, where θ is the unknown parameter to be estimated. Let $\widehat{\theta}$ be an estimator of θ and let $\mathcal{L}\left(\widehat{\theta}, \theta\right)$ be a given loss function. The expected value of this loss function with respect to the population distribution $f(x/\theta)$, that is

$$R_{\mathcal{L}}(\theta) = \int \mathcal{L}\left(\widehat{\theta}, \theta\right) \, f(x/\theta)\, dx$$

is called the risk.

The posterior density of the parameter θ given the sample $x_1, x_2, ..., x_n$, that is

$$k(\theta/x_1, x_2, ..., x_n)$$

contains all information about θ. In Bayesian estimation of parameter one chooses an estimate $\widehat{\theta}$ for θ such that

$$k(\widehat{\theta}/x_1, x_2, ..., x_n)$$

is maximum subject to a loss function. Mathematically, this is equivalent to minimizing the integral

$$\int_{\Omega} \mathcal{L}\left(\widehat{\theta}, \theta\right) \, k(\theta/x_1, x_2, ..., x_n)\, d\theta$$

with respect to $\widehat{\theta}$, where Ω denotes the support of the prior density $h(\theta)$ of the parameter θ.

Example 15.23. Suppose one observation was taken of a random variable X which yielded the value 2. The density function for X is

$$f(x/\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ \\ 0 & \text{otherwise,} \end{cases}$$

and prior distribution for parameter θ is

$$h(\theta) = \begin{cases} \frac{3}{\theta^4} & \text{if } 1 < \theta < \infty \\ \\ 0 & \text{otherwise.} \end{cases}$$

If the loss function is $\mathcal{L}(z, \theta) = (z - \theta)^2$, then what is the Bayes' estimate for θ ?

Answer: The prior density of the random variable θ is

$$h(\theta) = \begin{cases} \frac{3}{\theta^4} & \text{if } 1 < \theta < \infty \\ \\ 0 & \text{otherwise.} \end{cases}$$

The probability density function of the population is

$$f(x/\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ \\ 0 & \text{otherwise.} \end{cases}$$

Hence, the joint probability density function of the sample and the parameter is given by

$$\begin{aligned} u(x, \theta) &= h(\theta)\, f(x/\theta) \\ &= \frac{3}{\theta^4}\, \frac{1}{\theta} \\ &= \begin{cases} 3\,\theta^{-5} & \text{if } 0 < x < \theta \quad \text{and} \quad 1 < \theta < \infty \\ \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

The marginal density of the sample is given by

$$\begin{aligned} g(x) &= \int_x^\infty u(x, \theta)\, d\theta \\ &= \int_x^\infty 3\,\theta^{-5}\, d\theta \\ &= \frac{3}{4}\, x^{-4} \\ &= \frac{3}{4\, x^4}. \end{aligned}$$

Thus, if $x = 2$, then $g(2) = \frac{3}{64}$. The posterior density of θ when $x = 2$ is given by

$$\begin{aligned} k(\theta/x=2) &= \frac{u(2,\theta)}{g(2)} \\ &= \frac{64}{3}\, 3\,\theta^{-5} \\ &= \begin{cases} 64\,\theta^{-5} & \text{if } 2 < \theta < \infty \\ \\ 0 & \text{otherwise}. \end{cases} \end{aligned}$$

Now, we find the Bayes estimator by minimizing the expression $E\left[\mathcal{L}(\theta,z)/x=2\right]$. That is

$$\widehat{\theta} = Arg\,\max_{z\in\Omega} \int_\Omega \mathcal{L}(\theta,z)\,k(\theta/x=2)\,d\theta.$$

Let us call this integral $\psi(z)$. Then

$$\begin{aligned} \psi(z) &= \int_\Omega \mathcal{L}(\theta,z)\,k(\theta/x=2)\,d\theta \\ &= \int_2^\infty (z-\theta)^2\,k(\theta/x=2)\,d\theta \\ &= \int_2^\infty (z-\theta)^2\,64\theta^{-5}\,d\theta. \end{aligned}$$

We want to find the value of z which yields a minimum of $\psi(z)$. This can be done by taking the derivative of $\psi(z)$ and evaluating where the derivative is zero.

$$\begin{aligned} \frac{d}{dz}\psi(z) &= \frac{d}{dz}\int_2^\infty (z-\theta)^2\,64\theta^{-5}\,d\theta \\ &= 2\int_2^\infty (z-\theta)\,64\theta^{-5}\,d\theta \\ &= 2\int_2^\infty z\,64\theta^{-5}\,d\theta - 2\int_2^\infty \theta\,64\theta^{-5}\,d\theta \\ &= 2\,z - \frac{16}{3}. \end{aligned}$$

Setting this derivative of $\psi(z)$ to zero and solving for z, we get

$$\begin{aligned} 2z - \frac{16}{3} &= 0 \\ \Rightarrow z &= \frac{8}{3}. \end{aligned}$$

Since $\frac{d^2\psi(z)}{dz^2} = 2$, the function $\psi(z)$ has a minimum at $z = \frac{8}{3}$. Hence, the Bayes' estimate of θ is $\frac{8}{3}$.

In Example 15.23, we have found the Bayes' estimate of θ by directly minimizing the $\int_\Omega \mathcal{L}\left(\widehat{\theta}, \theta\right) k(\theta/x_1, x_2, ..., x_n)\, d\theta$ with respect to $\widehat{\theta}$. The next result is very useful while finding the Bayes' estimate using a quadratic loss function. Notice that if $\mathcal{L}(\widehat{\theta}, \theta) = (\theta - \widehat{\theta})^2$, then $\int_\Omega \mathcal{L}\left(\widehat{\theta}, \theta\right) k(\theta/x_1, x_2, ..., x_n)\, d\theta$ is $E\left((\theta - \widehat{\theta})^2 / x_1, x_2, ..., x_n\right)$. The following theorem is based on the fact that the function ϕ defined by $\phi(c) = E\left[(X - c)^2\right]$ attains minimum if $c = E[X]$.

Theorem 15.3. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density $f(x/\theta)$, where θ is the unknown parameter to be estimated. If the loss function is squared error, then the Bayes' estimator $\widehat{\theta}$ of parameter θ is given by

$$\widehat{\theta} = E(\theta/x_1, x_2, ..., x_n),$$

where the expectation is taken with respect to density $k(\theta/x_1, x_2, ..., x_n)$.

Now we give several examples to illustrate the use of this theorem.

Example 15.24. Suppose the prior distribution of θ is uniform over the interval $(0, 1)$. Given θ, the population X is uniform over the interval $(0, \theta)$. If the squared error loss function is used, find the Bayes' estimator of θ based on a sample of size one.

Answer: The prior density of θ is given by

$$h(\theta) = \begin{cases} 1 & \text{if } 0 < \theta < 1 \\ 0 & \text{otherwise .} \end{cases}$$

The density of population is given by

$$f(x/\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ 0 & \text{otherwise.} \end{cases}$$

The joint density of the sample and the parameter is given by

$$\begin{aligned} u(x, \theta) &= h(\theta)\, f(x/\theta) \\ &= 1 \;\left(\frac{1}{\theta}\right) \\ &= \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta < 1 \\ 0 & \text{otherwise .} \end{cases} \end{aligned}$$

The marginal density of the sample is

$$\begin{aligned} g(x) &= \int_x^1 u(x,\theta)\, d\theta \\ &= \int_x^1 \frac{1}{\theta}\, d\theta \\ &= \begin{cases} -\ln x & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

The conditional density of θ given the sample is

$$k(\theta/x) = \frac{u(x,\theta)}{g(x)} = \begin{cases} -\frac{1}{\theta \, \ln x} & \text{if } 0 < x < \theta < 1 \\ \\ 0 & \text{elsewhere .} \end{cases}$$

Since the loss function is quadratic error, therefore the Bayes' estimator of θ is

$$\begin{aligned} \widehat{\theta} &= E[\theta/x] \\ &= \int_x^1 \theta\, k(\theta/x)\, d\theta \\ &= \int_x^1 \theta\, \frac{-1}{\theta \, \ln x}\, d\theta \\ &= -\frac{1}{\ln x} \int_x^1 d\theta \\ &= \frac{x-1}{\ln x}. \end{aligned}$$

Thus, the Bayes' estimator of θ based on one observation X is

$$\widehat{\theta} = \frac{X-1}{\ln X}.$$

Example 15.25. Given θ, the random variable X has a binomial distribution with $n = 2$ and probability of success θ. If the prior density of θ is

$$h(\theta) = \begin{cases} k & \text{if } \frac{1}{2} < \theta < 1 \\ \\ 0 & \text{otherwise,} \end{cases}$$

what is the Bayes' estimate of θ for a squared error loss if $X = 1$?

Answer: Note that θ is uniform on the interval $\left(\frac{1}{2}, 1\right)$, hence $k = 2$. Therefore, the prior density of θ is

$$h(\theta) = \begin{cases} 2 & \text{if } \frac{1}{2} < \theta < 1 \\ \\ 0 & \text{otherwise.} \end{cases}$$

The population density is given by

$$f(x/\theta) = \binom{n}{x} \theta^x \, (1-\theta)^{n-x} = \binom{2}{x} \theta^x \, (1-\theta)^{2-x}, \qquad x = 0, 1, 2.$$

The joint density of the sample and the parameter θ is

$$\begin{aligned} u(x, \theta) &= h(\theta) \, f(x/\theta) \\ &= 2 \binom{2}{x} \theta^x \, (1-\theta)^{2-x} \end{aligned}$$

where $\frac{1}{2} < \theta < 1$ and $x = 0, 1, 2$. The marginal density of the sample is given by

$$g(x) = \int_{\frac{1}{2}}^{1} u(x, \theta) \, d\theta.$$

This integral is easy to evaluate if we substitute $X = 1$ now. Hence

$$\begin{aligned} g(1) &= \int_{\frac{1}{2}}^{1} 2 \binom{2}{1} \theta \, (1-\theta) \, d\theta \\ &= \int_{\frac{1}{2}}^{1} \left(4\theta - 4\theta^2\right) \, d\theta \\ &= 4 \left[\frac{\theta^2}{2} - \frac{\theta^3}{3} \right]_{\frac{1}{2}}^{1} \\ &= \frac{2}{3} \left[3\theta^2 - 2\theta^3 \right]_{\frac{1}{2}}^{1} \\ &= \frac{2}{3} \left[(3-2) - \left(\frac{3}{4} - \frac{2}{8} \right) \right] \\ &= \frac{1}{3}. \end{aligned}$$

Therefore, the posterior density of θ given $x = 1$, is

$$k(\theta/x = 1) = \frac{u(1, \theta)}{g(1)} = 12 \, (\theta - \theta^2),$$

where $\frac{1}{2} < \theta < 1$. Since the loss function is quadratic error, therefore the

Bayes' estimate of θ is

$$\begin{aligned}
\widehat{\theta} &= E[\theta/x=1] \\
&= \int_{\frac{1}{2}}^{1} \theta\, k(\theta/x=1)\, d\theta \\
&= \int_{\frac{1}{2}}^{1} 12\,\theta\,(\theta-\theta^2)\, d\theta \\
&= \left[4\,\theta^3 - 3\,\theta^4\right]_{\frac{1}{2}}^{1} \\
&= 1 - \frac{5}{16} \\
&= \frac{11}{16}.
\end{aligned}$$

Hence, based on the sample of size one with $X=1$, the Bayes' estimate of θ is $\frac{11}{16}$, that is

$$\widehat{\theta} = \frac{11}{16}.$$

The following theorem help us to evaluate the Bayes estimate of a sample if the loss function is absolute error loss. This theorem is based the fact that a function $\phi(c) = E\left[\,|X-c|\,\right]$ is minimum if c is the median of X.

Theorem 15.4. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density $f(x/\theta)$, where θ is the unknown parameter to be estimated. If the loss function is absolute error, then the Bayes estimator $\widehat{\theta}$ of the parameter θ is given by

$$\widehat{\theta} = \text{median of } k(\theta/x_1, x_2, ..., x_n)$$

where $k(\theta/x_1, x_2, ..., x_n)$ is the posterior distribution of θ.

The followings are some examples to illustrate the above theorem.

Example 15.26. Given θ, the random variable X has a binomial distribution with $n=3$ and probability of success θ. If the prior density of θ is

$$h(\theta) = \begin{cases} k & \text{if } \frac{1}{2} < \theta < 1 \\ \\ 0 & \text{otherwise,} \end{cases}$$

what is the Bayes' estimate of θ for an *absolute difference error loss* if the sample consists of one observation $x=3$?

Answer: Since, the prior density of θ is

$$h(\theta) = \begin{cases} 2 & \text{if } \frac{1}{2} < \theta < 1 \\ \\ 0 & \text{otherwise}, \end{cases}$$

and the population density is

$$f(x/\theta) = \binom{3}{x} \theta^x (1-\theta)^{3-x},$$

the joint density of the sample and the parameter is given by

$$u(3, \theta) = h(\theta)\, f(3/\theta) = 2\, \theta^3,$$

where $\frac{1}{2} < \theta < 1$. The marginal density of the sample (at $x = 3$) is given by

$$\begin{aligned} g(3) &= \int_{\frac{1}{2}}^{1} u(3, \theta)\, d\theta \\ &= \int_{\frac{1}{2}}^{1} 2\, \theta^3\, d\theta \\ &= \left[\frac{\theta^4}{2} \right]_{\frac{1}{2}}^{1} \\ &= \frac{15}{32}. \end{aligned}$$

Therefore, the conditional density of θ given $X = 3$ is

$$k(\theta/x = 3) = \frac{u(3, \theta)}{g(3)} = \begin{cases} \frac{64}{15}\, \theta^3 & \text{if } \frac{1}{2} < \theta < 1 \\ \\ 0 & \text{elsewhere.} \end{cases}$$

Since, the loss function is absolute error, the Bayes' estimator is the median of the probability density function $k(\theta/x = 3)$. That is

$$\begin{aligned} \frac{1}{2} &= \int_{\frac{1}{2}}^{\widehat{\theta}} \frac{64}{15}\, \theta^3\, d\theta \\ &= \frac{64}{60} \left[\theta^4 \right]_{\frac{1}{2}}^{\widehat{\theta}} \\ &= \frac{64}{60} \left[\left(\widehat{\theta} \right)^4 - \frac{1}{16} \right]. \end{aligned}$$

Solving the above equation for $\widehat{\theta}$, we get

$$\widehat{\theta} = \sqrt[4]{\frac{17}{32}} = 0.8537.$$

Example 15.27. Suppose the prior distribution of θ is uniform over the interval $(2, 5)$. Given θ, X is uniform over the interval $(0, \theta)$. What is the Bayes' estimator of θ for *absolute error loss* if $X = 1$?

Answer: Since, the prior density of θ is

$$h(\theta) = \begin{cases} \frac{1}{3} & \text{if } 2 < \theta < 5 \\ \\ 0 & \text{otherwise }, \end{cases}$$

and the population density is

$$f(x/\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ \\ 0 & \text{elsewhere,} \end{cases}$$

the joint density of the sample and the parameter is given by

$$u(x, \theta) = h(\theta)\, f(x/\theta) = \frac{1}{3\theta},$$

where $2 < \theta < 5$ and $0 < x < \theta$. The marginal density of the sample (at $x = 1$) is given by

$$\begin{aligned} g(1) &= \int_1^5 u(1, \theta)\, d\theta \\ &= \int_1^2 u(1, \theta)\, d\theta + \int_2^5 u(1, \theta)\, d\theta \\ &= \int_2^5 \frac{1}{3\theta}\, d\theta \\ &= \frac{1}{3}\, \ln\left(\frac{5}{2}\right). \end{aligned}$$

Therefore, the conditional density of θ given the sample $x = 1$, is

$$\begin{aligned} k(\theta/x = 1) &= \frac{u(1, \theta)}{g(1)} \\ &= \frac{1}{\theta\, \ln\left(\frac{5}{2}\right)}. \end{aligned}$$

Since, the loss function is absolute error, the Bayes estimate of θ is the median of $k(\theta/x = 1)$. Hence

$$\frac{1}{2} = \int_2^{\widehat{\theta}} \frac{1}{\theta \, \ln\left(\frac{5}{2}\right)} \, d\theta$$

$$= \frac{1}{\ln\left(\frac{5}{2}\right)} \, \ln\left(\frac{\widehat{\theta}}{2}\right).$$

Solving for $\widehat{\theta}$, we get

$$\widehat{\theta} = \sqrt{10} = 3.16.$$

Example 15.28. What is the basic principle of Bayesian estimation?

Answer: The basic principle behind the Bayesian estimation method consists of choosing a value of the parameter θ for which the observed data have as high a posterior probability $k(\theta/x_1, x_2, ..., x_n)$ of θ as possible subject to a loss function.

15.4. Review Exercises

1. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with a probability density function

$$f(x;\theta) = \begin{cases} \frac{1}{2\theta} & \text{if} \quad -\theta < x < \theta \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < \theta$ is a parameter. Using the moment method find an estimator for the parameter θ.

2. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with a probability density function

$$f(x;\theta) = \begin{cases} (\theta+1)\, x^{-\theta-2} & \text{if} \quad 1 < x < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < \theta$ is a parameter. Using the moment method find an estimator for the parameter θ.

3. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with a probability density function

$$f(x;\theta) = \begin{cases} \theta^2 \, x \, e^{-\theta x} & \text{if} \quad 0 < x < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < \theta$ is a parameter. Using the moment method find an estimator for the parameter θ.

4. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with a probability density function

$$f(x;\theta) = \begin{cases} \theta\, x^{\theta-1} & \text{if} \quad 0 < x < 1 \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < \theta$ is a parameter. Using the maximum likelihood method find an estimator for the parameter θ.

5. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with a probability density function

$$f(x;\theta) = \begin{cases} (\theta+1)\, x^{-\theta-2} & \text{if} \quad 1 < x < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < \theta$ is a parameter. Using the maximum likelihood method find an estimator for the parameter θ.

6. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with a probability density function

$$f(x;\theta) = \begin{cases} \theta^2\, x\, e^{-\theta\, x} & \text{if} \quad 0 < x < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < \theta$ is a parameter. Using the maximum likelihood method find an estimator for the parameter θ.

7. Let X_1, X_2, X_3, X_4 be a random sample from a distribution with density function

$$f(x;\beta) = \begin{cases} \frac{1}{\beta}\, e^{\frac{-(x-4)}{\beta}} & \text{for} \quad x > 4 \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $\beta > 0$. If the data from this random sample are 8.2, 9.1, 10.6 and 4.9, respectively, what is the maximum likelihood estimate of β?

8. Given θ, the random variable X has a binomial distribution with $n = 2$ and probability of success θ. If the prior density of θ is

$$h(\theta) = \begin{cases} k & \text{if } \frac{1}{2} < \theta < 1 \\ \\ 0 & \text{otherwise,} \end{cases}$$

what is the Bayes' estimate of θ for a squared error loss if the sample consists of $x_1 = 1$ and $x_2 = 2$.

9. Suppose two observations were taken of a random variable X which yielded the values 2 and 3. The density function for X is

$$f(x/\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 < x < \theta \\ \\ 0 & \text{otherwise,} \end{cases}$$

and prior distribution for the parameter θ is

$$h(\theta) = \begin{cases} 3\,\theta^{-4} & \text{if } \theta > 1 \\ \\ 0 & \text{otherwise.} \end{cases}$$

If the loss function is quadratic, then what is the Bayes' estimate for θ?

10. The Pareto distribution is often used in study of incomes and has the *cumulative density function*

$$F(x; \alpha, \theta) = \begin{cases} 1 - \left(\frac{\alpha}{x}\right)^{\theta} & \text{if } \alpha \leq x \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < \alpha < \infty$ and $1 < \theta < \infty$ are parameters. Find the maximum likelihood estimates of α and θ based on a sample of size 5 for value 3, 5, 2, 7, 8.

11. The Pareto distribution is often used in study of incomes and has the *cumulative density function*

$$F(x; \alpha, \theta) = \begin{cases} 1 - \left(\frac{\alpha}{x}\right)^{\theta} & \text{if } \alpha \leq x \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < \alpha < \infty$ and $1 < \theta < \infty$ are parameters. Using moment methods find estimates of α and θ based on a sample of size 5 for value 3, 5, 2, 7, 8.

12. Suppose one observation was taken of a random variable X which yielded the value 2. The density function for X is

$$f(x/\mu) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(x-\mu)^2} \qquad -\infty < x < \infty,$$

and prior distribution of μ is

$$h(\mu) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\mu^2} \qquad -\infty < \mu < \infty.$$

If the loss function is quadratic, then what is the Bayes' estimate for μ?

13. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with probability density

$$f(x) = \begin{cases} \frac{1}{\theta} & \text{if } 2\theta \leq x \leq 3\theta \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$. What is the maximum likelihood estimator of θ?

14. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with probability density

$$f(x) = \begin{cases} 1 - \theta^2 & \text{if } 0 \leq x \leq \frac{1}{1-\theta^2} \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$. What is the maximum likelihood estimator of θ?

15. Given θ, the random variable X has a binomial distribution with $n = 3$ and probability of success θ. If the prior density of θ is

$$h(\theta) = \begin{cases} k & \text{if } \frac{1}{2} < \theta < 1 \\ \\ 0 & \text{otherwise,} \end{cases}$$

what is the Bayes' estimate of θ for a *absolute difference error loss* if the sample consists of one observation $x = 1$?

16. Suppose the random variable X has the cumulative density function $F(x)$. Show that the expected value of the random variable $(X - c)^2$ is minimum if c equals the expected value of X.

17. Suppose the continuous random variable X has the cumulative density function $F(x)$. Show that the expected value of the random variable $|X - c|$ is minimum if c equals the median of X (that is, $F(c) = 0.5$).

18. Eight independent trials are conducted of a given system with the following results: S, F, S, F, S, S, S, S where S denotes the success and F denotes the failure. What is the maximum likelihood estimate of the probability of successful operation p ?

19. What is the maximum likelihood estimate of β if the 5 values $\frac{4}{5}$, $\frac{2}{3}$, 1, $\frac{3}{2}$, $\frac{5}{4}$ were drawn from the population for which $f(x; \beta) = \frac{1}{2}\,(1 + \beta)^5 \, \left(\frac{x}{2}\right)^{\beta}$?

20. If a sample of five values of X is taken from the population for which $f(x;t) = 2(t-1)t^x$, what is the maximum likelihood estimator of t ?

21. A sample of size n is drawn from a gamma distribution

$$f(x;\beta) = \begin{cases} \frac{x^3\, e^{-\frac{x}{\beta}}}{6\beta^4} & \text{if } 0 < x < \infty \\ \\ 0 & \text{otherwise.} \end{cases}$$

What is the maximum likelihood estimator of β ?

22. The probability density function of the random variable X is defined by

$$f(x;\lambda) = \begin{cases} 1 - \frac{2}{3}\lambda + \lambda\sqrt{x} & \text{if } 0 \le x \le 1 \\ \\ 0 & \text{otherwise.} \end{cases}$$

What is the maximum likelihood estimate of the parameter λ based on two independent observations $x_1 = \frac{1}{4}$ and $x_2 = \frac{9}{16}$?

23. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density function $f(x;\sigma) = \frac{\sigma}{2}\, e^{-\sigma|x-\mu|}$. What is the maximum likelihood estimator of σ ?

24. Suppose $X_1, X_2, ...$ are independent random variables, each with probability of success p and probability of failure $1-p$, where $0 \le p \le 1$. Let N be the number of observation needed to obtain the first success. What is the maximum likelihood estimator of p in term of N ?

25. Let X_1, X_2, X_3 and X_4 be a random sample from the discrete distribution X such that

$$P(X = x) = \begin{cases} \frac{\theta^{2x}\;\, e^{-\theta^2}}{x!} & \text{for } x = 0, 1, 2, ..., \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$. If the data are 17, 10, 32, 5, what is the maximum likelihood estimate of θ ?

26. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population with a probability density function

$$f(x;\alpha,\lambda) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1} e^{-\lambda\, x} & \text{if} \quad 0 < x < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where α and λ are parameters. Using the moment method find the estimators for the parameters α and λ.

27. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population distribution with the probability density function

$$f(x;p) = \binom{10}{x} p^x \, (1-p)^{10-x}$$

for $x = 0, 1, ..., 10$, where p is a parameter. Find the Fisher information in the sample about the parameter p.

28. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population distribution with the probability density function

$$f(x;\theta) = \begin{cases} \theta^2 \, x \, e^{-\theta \, x} & \text{if} \quad 0 < x < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < \theta$ is a parameter. Find the Fisher information in the sample about the parameter θ.

29. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population distribution with the probability density function

$$f(x;\mu,\sigma^2) = \begin{cases} \frac{1}{x \, \sigma \, \sqrt{2\,\pi}} \, e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}, & \text{if } 0 < x < \infty \\ \\ 0 & \text{otherwise ,} \end{cases}$$

where $-\infty < \mu < \infty$ and $0 < \sigma^2 < \infty$ are unknown parameters. Find the Fisher information matrix in the sample about the parameters μ and σ^2.

30. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population distribution with the probability density function

$$f(x;\mu,\lambda) = \begin{cases} \sqrt{\frac{\lambda}{2\pi}} \, x^{-\frac{3}{2}} \; e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}}, & \text{if } 0 < x < \infty \\ \\ 0 & \text{otherwise ,} \end{cases}$$

where $0 < \mu < \infty$ and $0 < \lambda < \infty$ are unknown parameters. Find the Fisher information matrix in the sample about the parameters μ and λ.

31. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with a probability density function

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\,\theta^\alpha} \, x^{\alpha-1} \, e^{-\frac{x}{\theta}} & \text{if} \quad 0 < x < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha > 0$ and $\theta > 0$ are parameters. Using the moment method find estimators for parameters α and β.

32. Let $X_1, X_2, ..., X_n$ be a random sample of size$n$ from a distribution with a probability density function

$$f(x;\theta) = \frac{1}{\pi\,[1+(x-\theta)^2]}, \qquad -\infty < x < \infty,$$

where $0 < \theta$ is a parameter. Using the maximum likelihood method find an estimator for the parameter θ.

33. Let $X_1, X_2, ..., X_n$ be a random sample of size$n$ from a distribution with a probability density function

$$f(x;\theta) = \frac{1}{2}\, e^{-|x-\theta|}, \qquad -\infty < x < \infty,$$

where $0 < \theta$ is a parameter. Using the maximum likelihood method find an estimator for the parameter θ.

34. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population distribution with the probability density function

$$f(x;\lambda) = \begin{cases} \frac{\lambda^x\, e^{-\lambda}}{x!} & \text{if } x = 0, 1, ..., \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $\lambda > 0$ is an unknown parameter. Find the Fisher information matrix in the sample about the parameter λ.

35. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population distribution with the probability density function

$$f(x;p) = \begin{cases} (1-p)^{x-1} p & \text{if } x = 1, ..., \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < p < 1$ is an unknown parameter. Find the Fisher information matrix in the sample about the parameter p.