Chapter 16

CRITERIA FOR EVALUATING THE GOODNESS OF ESTIMATORS

We have seen in Chapter 15 that, in general, different parameter estimation methods yield different estimators. For example, if $X \sim UNIF(0, \theta)$ and $X_1, X_2, ..., X_n$ is a random sample from the population X, then the estimator of θ obtained by moment method is

$$\widehat{\theta}_{MM} = 2\overline{X}$$

where as the estimator obtained by the maximum likelihood method is

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

where $\overline{X}$ and $X_{(n)}$ are the sample average and the n^{th} order statistic, respectively. Now the question arises: which of the two estimators is better? Thus, we need some criteria to evaluate the goodness of an estimator. Some well known criteria for evaluating the goodness of an estimator are: (1) Unbiasedness, (2) Efficiency and Relative Efficiency, (3) Uniform Minimum Variance Unbiasedness, (4) Sufficiency, and (5) Consistency.

In this chapter, we shall examine only the first four criteria in details. The concepts of unbiasedness, efficiency and sufficiency were introduced by Sir Ronald Fisher.

16.1. The Unbiased Estimator

Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population with probability density function $f(x;\theta)$. An estimator $\widehat{\theta}$ of θ is a function of the random variables $X_1, X_2, ..., X_n$ which is free of the parameter θ. An estimate is a realized value of an estimator that is obtained when a sample is actually taken.

Definition 16.1. An estimator $\widehat{\theta}$ of θ is said to be an unbiased estimator of θ if and only if

$$E\left(\widehat{\theta}\right) = \theta.$$

If $\widehat{\theta}$ is not unbiased, then it is called a biased estimator of θ.

An estimator of a parameter may not equal to the actual value of the parameter for every realization of the sample $X_1, X_2, ..., X_n$, but if it is unbiased then on an average it will equal to the parameter.

Example 16.1. Let $X_1, X_2, ..., X_n$ be a random sample from a normal population with mean μ and variance $\sigma^2 > 0$. Is the sample mean $\overline{X}$ an unbiased estimator of the parameter μ ?

Answer: Since, each $X_i \sim N(\mu, \sigma^2)$, we have

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

That is, the sample mean is normal with mean μ and variance $\frac{\sigma^2}{n}$. Thus

$$E\left(\overline{X}\right) = \mu.$$

Therefore, the sample mean $\overline{X}$ is an unbiased estimator of μ.

Example 16.2. Let $X_1, X_2, ..., X_n$ be a random sample from a normal population with mean μ and variance $\sigma^2 > 0$. What is the maximum likelihood estimator of σ^2 ? Is this maximum likelihood estimator an unbiased estimator of the parameter σ^2 ?

Answer: In Example 15.13, we have shown that the maximum likelihood estimator of σ^2 is

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

Now, we examine the unbiasedness of this estimator

$$
\begin{aligned}
E\left[\widehat{\sigma^2}\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right] \\
&= E\left[\frac{n-1}{n}\,\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right] \\
&= \frac{n-1}{n}\,E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right] \\
&= \frac{n-1}{n}\,E\left[S^2\right] \\
&= \frac{\sigma^2}{n}\,E\left[\frac{n-1}{\sigma^2}\,S^2\right] \qquad \left(\text{since } \frac{n-1}{\sigma^2}\,S^2\sim\chi^2(n-1)\right) \\
&= \frac{\sigma^2}{n}E\left[\chi^2(n-1)\right] \\
&= \frac{\sigma^2}{n}\,(n-1) \\
&= \frac{n-1}{n}\,\sigma^2 \\
&\neq \sigma^2.
\end{aligned}
$$

Therefore, the maximum likelihood estimator of σ^2 is a biased estimator.

Next, in the following example, we show that the sample variance S^2 given by the expression

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

is an unbiased estimator of the population variance σ^2 irrespective of the population distribution.

Example 16.3. Let $X_1, X_2, ..., X_n$ be a random sample from a population with mean μ and variance $\sigma^2 > 0$. Is the sample variance S^2 an unbiased estimator of the population variance σ^2 ?

Answer: Note that the distribution of the population is not given. However, we are given $E(X_i) = \mu$ and $E[(X_i-\mu)^2] = \sigma^2$. In order to find $E\left(S^2\right)$, we need $E\left(\overline{X}\right)$ and $E\left(\overline{X}^2\right)$. Thus we proceed to find these two expected values. Consider

$$
\begin{aligned}
E\left(\overline{X}\right) &= E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
Var\left(\overline{X}\right) &= Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}.
\end{aligned}
$$

Therefore

$$
E\left(\overline{X}^2\right) = Var\left(\overline{X}\right) + E\left(\overline{X}\right)^2 = \frac{\sigma^2}{n} + \mu^2.
$$

Consider

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

Therefore, the sample variance S^2 is an unbiased estimator of the population variance σ^2.

Example 16.4. Let X be a random variable with mean 2. Let $\widehat{\theta_1}$ and $\widehat{\theta_2}$ be unbiased estimators of the second and third moments, respectively, of X about the origin. Find an unbiased estimator of the third moment of X about its mean in terms of $\widehat{\theta_1}$ and $\widehat{\theta_2}$.

Answer: Since, $\widehat{\theta_1}$ and $\widehat{\theta_2}$ are the unbiased estimators of the second and third moments of X about origin, we get

$$E\left(\widehat{\theta_1}\right) = E(X^2) \qquad \text{and} \qquad E\left(\widehat{\theta_2}\right) = E\left(X^3\right).$$

The unbiased estimator of the third moment of X about its mean is

$$\begin{aligned} E\left[(X-2)^3\right] &= E\left[X^3 - 6X^2 + 12X - 8\right] \\ &= E\left[X^3\right] - 6E\left[X^2\right] + 12E\left[X\right] - 8 \\ &= \widehat{\theta_2} - 6\widehat{\theta_1} + 24 - 8 \\ &= \widehat{\theta_2} - 6\widehat{\theta_1} + 16. \end{aligned}$$

Thus, the unbiased estimator of the third moment of X about its mean is $\widehat{\theta_2} - 6\widehat{\theta_1} + 16$.

Example 16.5. Let $X_1, X_2, ..., X_5$ be a sample of size 5 from the uniform distribution on the interval $(0, \theta)$, where θ is unknown. Let the estimator of θ be $k\,X_{\max}$, where k is some constant and $X_{\max}$ is the largest observation. In order $k\,X_{\max}$ to be an unbiased estimator, what should be the value of the constant k ?

Answer: The probability density function of $X_{\max}$ is given by

$$\begin{aligned} g(x) &= \frac{5!}{4!\,0!}\,\left[F(x)\right]^4\,f(x) \\ &= 5\,\left(\frac{x}{\theta}\right)^4\,\frac{1}{\theta} \\ &= \frac{5}{\theta^5}\,x^4. \end{aligned}$$

If $k\,X_{\max}$ is an unbiased estimator of θ, then

$$\begin{aligned} \theta &= E\left(k\,X_{\max}\right) \\ &= k\,E\left(X_{\max}\right) \\ &= k\int_0^\theta x\,g(x)\,dx \\ &= k\int_0^\theta \frac{5}{\theta^5}\,x^5\,dx \\ &= \frac{5}{6}\,k\,\theta. \end{aligned}$$

Hence,

$$k = \frac{6}{5}.$$

Example 16.6. Let $X_1, X_2, ..., X_n$ be a sample of size n from a distribution with unknown mean $-\infty < \mu < \infty$, and unknown variance $\sigma^2 > 0$. Show that the statistic $\overline{X}$ and $Y = \frac{X_1+2X_2+\cdots+nX_n}{\frac{n\,(n+1)}{2}}$ are both unbiased estimators of μ. Further, show that $Var\left(\overline{X}\right) < Var(Y)$.

Answer: First, we show that $\overline{X}$ is an unbiased estimator of μ

$$\begin{aligned} E\left(\overline{X}\right) &= E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mu = \mu. \end{aligned}$$

Hence, the sample mean $\overline{X}$ is an unbiased estimator of the population mean irrespective of the distribution of X. Next, we show that Y is also an unbiased estimator of μ.

$$\begin{aligned} E\left(Y\right) &= E\left(\frac{X_1 + 2X_2 + \cdots + nX_n}{\frac{n\,(n+1)}{2}}\right) \\ &= \frac{2}{n\,(n+1)}\sum_{i=1}^{n} i\,E\left(X_i\right) \\ &= \frac{2}{n\,(n+1)}\sum_{i=1}^{n} i\,\mu \\ &= \frac{2}{n\,(n+1)}\;\mu\;\frac{n\,(n+1)}{2} \\ &= \mu. \end{aligned}$$

Hence, $\overline{X}$ and Y are both unbiased estimator of the population mean irrespective of the distribution of the population. The variance of $\overline{X}$ is given by

$$\begin{aligned} Var\left[\overline{X}\right] &= Var\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \\ &= \frac{1}{n^2}\,Var\left[X_1 + X_2 + \cdots + X_n\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var\left[X_i\right] \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

Similarly, the variance of Y can be calculated as follows:

$$\begin{aligned}
Var\,[Y] &= Var\left[\frac{X_1 + 2X_2 + \cdots + nX_n}{\frac{n\,(n+1)}{2}}\right] \\
&= \frac{4}{n^2\,(n+1)^2}\,Var\,[1\,X_1 + 2\,X_2 + \cdots + n\,X_n] \\
&= \frac{4}{n^2\,(n+1)^2}\,\sum_{i=1}^{n} Var\,[i\,X_i] \\
&= \frac{4}{n^2\,(n+1)^2}\,\sum_{i=1}^{n} i^2\,Var\,[X_i] \\
&= \frac{4}{n^2\,(n+1)^2}\,\sigma^2\,\sum_{i=1}^{n} i^2 \\
&= \sigma^2\,\frac{4}{n^2\,(n+1)^2}\,\frac{n\,(n+1)\,(2n+1)}{6} \\
&= \frac{2}{3}\,\frac{2n+1}{(n+1)}\,\frac{\sigma^2}{n} \\
&= \frac{2}{3}\,\frac{2n+1}{(n+1)}\,Var\left[\,\overline{X}\,\right].
\end{aligned}$$

Since $\frac{2}{3}\,\frac{2n+1}{(n+1)} > 1$ for $n \geq 2$, we see that $Var\left[\,\overline{X}\,\right] < Var[\,Y\,]$. This shows that although the estimators $\overline{X}$ and Y are both unbiased estimator of μ, yet the variance of the sample mean $\overline{X}$ is smaller than the variance of Y.

In statistics, between two unbiased estimators one prefers the estimator which has the minimum variance. This leads to our next topic. However, before we move to the next topic we complete this section with some known disadvantages with the notion of unbiasedness. The first disadvantage is that an unbiased estimator for a parameter may not exist. The second disadvantage is that the property of unbiasedness is not invariant under functional transformation, that is, if $\widehat{\theta}$ is an unbiased estimator of θ and g is a function, then $g(\widehat{\theta})$ may not be an unbiased estimator of $g(\theta)$.

16.2. The Relatively Efficient Estimator

We have seen that in Example 16.6 that the sample mean

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

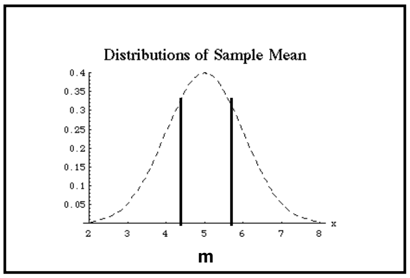
and the statistic

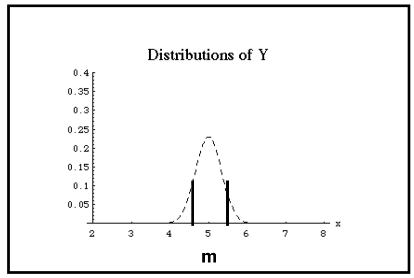
$$Y = \frac{X_1 + 2X_2 + \cdots + nX_n}{1 + 2 + \cdots + n}$$

are both unbiased estimators of the population mean. However, we also seen that

$$Var\left(\overline{X}\right) < Var(Y).$$

The following figure graphically illustrates the shape of the distributions of both the unbiased estimators.





If an unbiased estimator has a smaller variance or dispersion, then it has a greater chance of being close to true parameter θ. Therefore when two estimators of θ are both unbiased, then one should pick the one with the smaller variance.

Definition 16.2. Let $\widehat{\theta}_1$ and $\widehat{\theta}_2$ be two unbiased estimators of θ. The estimator $\widehat{\theta}_1$ is said to be more efficient than $\widehat{\theta}_2$ if

$$Var\left(\widehat{\theta}_1\right) < Var\left(\widehat{\theta}_2\right).$$

The ratio η given by

$$\eta\left(\widehat{\theta}_1, \widehat{\theta}_2\right) = \frac{Var\left(\widehat{\theta}_2\right)}{Var\left(\widehat{\theta}_1\right)}$$

is called the relative efficiency of $\widehat{\theta}_1$ with respect to $\widehat{\theta}_2$.

Example 16.7. Let X_1, X_2, X_3 be a random sample of size 3 from a population with mean μ and variance $\sigma^2 > 0$. If the statistics $\overline{X}$ and Y given by

$$Y = \frac{X_1 + 2X_2 + 3X_3}{6}$$

are two unbiased estimators of the population mean μ, then which one is more efficient?

Answer: Since $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, we get

$$\begin{aligned}
E\left(\overline{X}\right) &= E\left(\frac{X_1 + X_2 + X_3}{3}\right) \\
&= \frac{1}{3}\,\left(E(X_1) + E(X_2) + E(X_3)\right) \\
&= \frac{1}{3}\,3\mu \\
&= \mu
\end{aligned}$$

and

$$\begin{aligned}
E(Y) &= E\left(\frac{X_1 + 2X_2 + 3X_3}{6}\right) \\
&= \frac{1}{6}\,\left(E(X_1) + 2E(X_2) + 3E(X_3)\right) \\
&= \frac{1}{6}\,6\mu \\
&= \mu.
\end{aligned}$$

Therefore both $\overline{X}$ and Y are unbiased. Next we determine the variance of both the estimators. The variances of these estimators are given by

$$\begin{aligned}
Var\left(\overline{X}\right) &= Var\left(\frac{X_1 + X_2 + X_3}{3}\right) \\
&= \frac{1}{9}\,\left[Var(X_1) + Var(X_2) + Var(X_3)\right] \\
&= \frac{1}{9}\,3\sigma^2 \\
&= \frac{12}{36}\,\sigma^2
\end{aligned}$$

and

$$\begin{aligned}
Var(Y) &= Var\left(\frac{X_1 + 2X_2 + 3X_3}{6}\right) \\
&= \frac{1}{36}\,\left[Var(X_1) + 4Var(X_2) + 9Var(X_3)\right] \\
&= \frac{1}{36}\,14\sigma^2 \\
&= \frac{14}{36}\,\sigma^2.
\end{aligned}$$

Therefore

$$\frac{12}{36}\,\sigma^2 = Var\left(\overline{X}\right) < Var(Y) = \frac{14}{36}\,\sigma^2.$$

Hence, $\overline{X}$ is more efficient than the estimator Y. Further, the relative efficiency of $\overline{X}$ with respect to Y is given by

$$\eta\left(\overline{X}, Y\right) = \frac{14}{12} = \frac{7}{6}.$$

Example 16.8. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population with density

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

where $\theta > 0$ is a parameter. Are the estimators X_1 and $\overline{X}$ unbiased? Given, X_1 and $\overline{X}$, which one is more efficient estimator of θ ?

Answer: Since the population X is exponential with parameter θ, that is $X \sim EXP(\theta)$, the mean and variance of it are given by

$$E(X) = \theta \qquad \text{and} \qquad Var(X) = \theta^2.$$

Since $X_1, X_2, ..., X_n$ is a random sample from X, we see that the statistic $X_1 \sim EXP(\theta)$. Hence, the expected value of X_1 is θ and thus it is an unbiased estimator of the parameter θ. Also, the sample mean is an unbiased estimator of θ since

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

Next, we compute the variances of the unbiased estimators X_1 and $\overline{X}$. It is easy to see that

$$Var\left(X_1\right) = \theta^2$$

and

$$\begin{aligned} Var\left(\overline{X}\right) &= Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var\left(X_i\right) \\ &= \frac{1}{n^2}\, n\theta^2 \\ &= \frac{\theta^2}{n}. \end{aligned}$$

Hence
$$\frac{\theta^2}{n} = Var\left(\overline{X}\right) < Var\left(X_1\right) = \theta^2.$$
Thus $\overline{X}$ is more efficient than X_1 and the relative efficiency of $\overline{X}$ with respect to X_1 is
$$\eta(\overline{X},\, X_1) = \frac{\theta^2}{\frac{\theta^2}{n}} = n.$$

Example 16.9. Let X_1, X_2, X_3 be a random sample of size 3 from a population with density

$$f(x;\lambda) = \begin{cases} \frac{\lambda^x\, e^{-\lambda}}{x!} & \text{if } x = 0, 1, 2, ..., \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where λ is a parameter. Are the estimators given by

$$\widehat{\lambda_1} = \frac{1}{4}\,\left(X_1 + 2X_2 + X_3\right) \qquad \text{and} \qquad \widehat{\lambda_2} = \frac{1}{9}\,\left(4X_1 + 3X_2 + 2X_3\right)$$

unbiased? Given, $\widehat{\lambda_1}$ and $\widehat{\lambda_2}$, which one is more efficient estimator of λ ? Find an unbiased estimator of λ whose variance is smaller than the variances of $\widehat{\lambda_1}$ and $\widehat{\lambda_2}$.

Answer: Since each $X_i \sim POI(\lambda)$, we get

$$E\left(X_i\right) = \lambda \qquad \text{and} \qquad Var\left(X_i\right) = \lambda.$$

It is easy to see that

$$\begin{aligned} E\left(\widehat{\lambda_1}\right) &= \frac{1}{4}\left(E\left(X_1\right) + 2E\left(X_2\right) + E\left(X_3\right)\right) \\ &= \frac{1}{4}\; 4\lambda \\ &= \lambda, \end{aligned}$$

and
$$\begin{aligned} E\left(\widehat{\lambda_2}\right) &= \frac{1}{9}\left(4E\left(X_1\right) + 3E\left(X_2\right) + 2E\left(X_3\right)\right) \\ &= \frac{1}{9}\; 9\lambda \\ &= \lambda. \end{aligned}$$

Thus, both $\widehat{\lambda_1}$ and $\widehat{\lambda_2}$ are unbiased estimators of λ. Now we compute their variances to find out which one is more efficient. It is easy to note that

$$\begin{aligned} Var\left(\widehat{\lambda_1}\right) &= \frac{1}{16}\left(Var\left(X_1\right) + 4Var\left(X_2\right) + Var\left(X_3\right)\right) \\ &= \frac{1}{16}\, 6\lambda \\ &= \frac{6}{16}\lambda \\ &= \frac{486}{1296}\lambda, \end{aligned}$$

and

$$\begin{aligned} Var\left(\widehat{\lambda_2}\right) &= \frac{1}{81}\left(16Var\left(X_1\right) + 9Var\left(X_2\right) + 4Var\left(X_3\right)\right) \\ &= \frac{1}{81}\, 29\lambda \\ &= \frac{29}{81}\lambda \\ &= \frac{464}{1296}\lambda, \end{aligned}$$

Since,

$$Var\left(\widehat{\lambda_2}\right) < Var\left(\widehat{\lambda_1}\right),$$

the estimator $\widehat{\lambda_2}$ is efficient than the estimator $\widehat{\lambda_1}$. We have seen in section 16.1 that the sample mean is always an unbiased estimator of the population mean irrespective of the population distribution. The variance of the sample mean is always equals to $\frac{1}{n}$ times the population variance, where n denotes the sample size. Hence, we get

$$Var\left(\overline{X}\right) = \frac{\lambda}{3} = \frac{432}{1296}\,\lambda.$$

Therefore, we get

$$Var\left(\overline{X}\right) < Var\left(\widehat{\lambda_2}\right) < Var\left(\widehat{\lambda_1}\right).$$

Thus, the sample mean has even smaller variance than the two unbiased estimators given in this example.

In view of this example, now we have encountered a new problem. That is how to find an unbiased estimator which has the smallest variance among all unbiased estimators of a given parameter. We resolve this issue in the next section.

16.3. The Uniform Minimum Variance Unbiased Estimator

Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population with probability density function $f(x;\theta)$. Recall that an estimator $\widehat{\theta}$ of θ is a function of the random variables $X_1, X_2, ..., X_n$ which does depend on θ.

Definition 16.3. An unbiased estimator $\widehat{\theta}$ of θ is said to be a uniform minimum variance unbiased estimator of θ if and only if

$$Var\left(\widehat{\theta}\right) \leq Var\left(\widehat{T}\right)$$

for any unbiased estimator $\widehat{T}$ of θ.

If an estimator $\widehat{\theta}$ is unbiased then the mean of this estimator is equal to the parameter θ, that is

$$E\left(\widehat{\theta}\right) = \theta$$

and the variance of $\widehat{\theta}$ is

$$\begin{aligned} Var\left(\widehat{\theta}\right) &= E\left[\left(\widehat{\theta} - E\left(\widehat{\theta}\right)\right)^2\right] \\ &= E\left[\left(\widehat{\theta} - \theta\right)^2\right]. \end{aligned}$$

This variance, if exists, is a function of the unbiased estimator $\widehat{\theta}$ and it has a minimum in the class of all unbiased estimators of θ. Therefore we have an alternative definition of the uniform minimum variance unbiased estimator.

Definition 16.4. An unbiased estimator $\widehat{\theta}$ of θ is said to be a uniform minimum variance unbiased estimator of θ if it minimizes the variance $E\left[\left(\widehat{\theta} - \theta\right)^2\right]$.

Example 16.10. Let $\widehat{\theta_1}$ and $\widehat{\theta_2}$ be unbiased estimators of θ. Suppose $Var\left(\widehat{\theta_1}\right) = 1$, $Var\left(\widehat{\theta_2}\right) = 2$ and $Cov\left(\widehat{\theta_1}, \widehat{\theta_2}\right) = \frac{1}{2}$. What are the values of c_1 and c_2 for which $c_1\widehat{\theta_1} + c_2\widehat{\theta_2}$ is an unbiased estimator of θ with minimum variance among unbiased estimators of this type?

Answer: We want $c_1\widehat{\theta_1} + c_2\widehat{\theta_2}$ to be a minimum variance unbiased estimator of θ. Then

$$\begin{aligned} & E\left[c_1\widehat{\theta_1} + c_2\widehat{\theta_2}\right] = \theta \\ \Rightarrow \quad & c_1\, E\left[\widehat{\theta_1}\right] + c_2\, E\left[\widehat{\theta_2}\right] = \theta \\ \Rightarrow \quad & c_1\,\theta + c_2\,\theta = \theta \\ \Rightarrow \quad & c_1 + c_2 = 1 \\ \Rightarrow \quad & c_2 = 1 - c_1. \end{aligned}$$

Therefore

$$
\begin{aligned}
Var\left[c_1\widehat{\theta}_1 + c_2\widehat{\theta}_2\right] &= c_1^2\, Var\left[\widehat{\theta}_1\right] + c_2^2\, Var\left[\widehat{\theta}_2\right] + 2\, c_1\, c_2\, Cov\left(\widehat{\theta}_1, \widehat{\theta}_1\right)\\
&= c_1^2 + 2c_2^2 + c_1c_2\\
&= c_1^2 + 2(1-c_1)^2 + c_1(1-c_1)\\
&= 2(1-c_1)^2 + c_1\\
&= 2 + 2c_1^2 - 3c_1.
\end{aligned}
$$

Hence, the variance $Var\left[c_1\widehat{\theta}_1 + c_2\widehat{\theta}_2\right]$ is a function of c_1. Let us denote this function by $\phi(c_1)$, that is

$$\phi(c_1) := Var\left[c_1\widehat{\theta}_1 + c_2\widehat{\theta}_2\right] = 2 + 2c_1^2 - 3c_1.$$

Taking the derivative of $\phi(c_1)$ with respect to c_1, we get

$$\frac{d}{dc_1}\phi(c_1) = 4c_1 - 3.$$

Setting this derivative to zero and solving for c_1, we obtain

$$4c_1 - 3 = 0 \qquad \Rightarrow \quad c_1 = \frac{3}{4}.$$

Therefore

$$c_2 = 1 - c_1 = 1 - \frac{3}{4} = \frac{1}{4}.$$

In Example 16.10, we saw that if $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are any two unbiased estimators of θ, then $c\,\widehat{\theta}_1 + (1-c)\,\widehat{\theta}_2$ is also an unbiased estimator of θ for any $c \in \mathbb{R}$. Hence given two estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$,

$$\mathcal{C} = \left\{\widehat{\theta} \mid \widehat{\theta} = c\,\widehat{\theta}_1 + (1-c)\,\widehat{\theta}_2,\;\; c \in \mathbb{R}\right\}$$

forms an uncountable class of unbiased estimators of θ. When the variances of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are known along with the their covariance, then in Example 16.10 we were able to determine the minimum variance unbiased estimator in the class $\mathcal{C}$. If the variances of the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are not known, then it is very difficult to find the minimum variance estimator even in the class of estimators $\mathcal{C}$. Notice that $\mathcal{C}$ is a subset of the class of all unbiased estimators and finding a minimum variance unbiased estimator in this class is a difficult task.

One way to find a uniform minimum variance unbiased estimator for a parameter is to use the Cramér-Rao lower bound or the Fisher information inequality.

Theorem 16.1. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population X with probability density $f(x;\theta)$, where θ is a scalar parameter. Let $\widehat{\theta}$ be any unbiased estimator of θ. Suppose the likelihood function $L(\theta)$ is a differentiable function of θ and satisfies

$$\begin{aligned} \frac{d}{d\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h(x_1,...,x_n)\, L(\theta)\, dx_1\cdots dx_n \\ = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h(x_1,...,x_n)\, \frac{d}{d\theta}L(\theta)\, dx_1\cdots dx_n \end{aligned} \tag{1}$$

for any $h(x_1, ..., x_n)$ with $E(h(X_1, ..., X_n)) < \infty$. Then

$$Var\left(\widehat{\theta}\right) \geq \frac{1}{E\left[\left(\frac{\partial \ln L(\theta)}{\partial\theta}\right)^2\right]}. \tag{CR1}$$

Proof: Since $L(\theta)$ is the joint probability density function of the sample $X_1, X_2, ..., X_n$,

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} L(\theta)\, dx_1\cdots dx_n = 1. \tag{2}$$

Differentiating (2) with respect to θ we have

$$\frac{d}{d\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} L(\theta)\, dx_1\cdots dx_n = 0$$

and use of (1) with $h(x_1, ..., x_n) = 1$ yields

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{d}{d\theta}L(\theta)\; dx_1\cdots dx_n = 0. \tag{3}$$

Rewriting (3) as

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{dL(\theta)}{d\theta}\,\frac{1}{L(\theta)}\; L(\theta)\; dx_1\cdots dx_n = 0$$

we see that

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{d\ln L(\theta)}{d\theta}\; L(\theta)\; dx_1\cdots dx_n = 0.$$

Hence
$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \theta \, \frac{d \ln L(\theta)}{d\theta} \, L(\theta) \; dx_1 \cdots dx_n = 0. \tag{4}$$

Since $\widehat{\theta}$ is an unbiased estimator of θ, we see that

$$E\left(\widehat{\theta}\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \widehat{\theta} \, L(\theta) \, dx_1 \cdots dx_n = \theta. \tag{5}$$

Differentiating (5) with respect to θ, we have

$$\frac{d}{d\theta} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \widehat{\theta} \, L(\theta) \, dx_1 \cdots dx_n = 1.$$

Again using (1) with $h(X_1, ..., X_n) = \widehat{\theta}$, we have

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \widehat{\theta} \, \frac{d}{d\theta} \, L(\theta) \; dx_1 \cdots dx_n = 1. \tag{6}$$

Rewriting (6) as

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \widehat{\theta} \, \frac{dL(\theta)}{d\theta} \, \frac{1}{L(\theta)} \, L(\theta) \; dx_1 \cdots dx_n = 1$$

we have
$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \widehat{\theta} \, \frac{d \ln L(\theta)}{d\theta} \, L(\theta) \; dx_1 \cdots dx_n = 1. \tag{7}$$

From (4) and (7), we obtain

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\widehat{\theta} - \theta\right) \; \frac{d \ln L(\theta)}{d\theta} \, L(\theta) \; dx_1 \cdots dx_n = 1. \tag{8}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}
1 &= \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\widehat{\theta} - \theta\right) \; \frac{d \ln L(\theta)}{d\theta} \, L(\theta) \; dx_1 \cdots dx_n \right)^2 \\
&\leq \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\widehat{\theta} - \theta\right)^2 \; L(\theta) \; dx_1 \cdots dx_n \right) \\
&\quad \cdot \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{d \ln L(\theta)}{d\theta} \right)^2 \; L(\theta) \; dx_1 \cdots dx_n \right) \\
&= Var\left(\widehat{\theta}\right) \; E\left[\left(\frac{\partial \ln L(\theta)}{\partial\theta} \right)^2 \right].
\end{aligned}$$

Therefore

$$Var\left(\widehat{\theta}\right) \geq \frac{1}{E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)^2\right]}$$

and the proof of theorem is now complete.

If $L(\theta)$ is twice differentiable with respect to θ, the inequality (CR1) can be stated equivalently as

$$Var\left(\widehat{\theta}\right) \geq \frac{-1}{E\left[\frac{\partial^2 \ln L(\theta)}{\partial \theta^2}\right]}. \tag{CR2}$$

The inequalities (CR1) and (CR2) are known as Cramér-Rao lower bound for the variance of $\widehat{\theta}$ or the Fisher information inequality. The condition (1) interchanges the order on integration and differentiation. Therefore any distribution whose range depend on the value of the parameter is not covered by this theorem. Hence distribution like the uniform distribution may not be analyzed using the Cramér-Rao lower bound.

If the estimator $\widehat{\theta}$ is minimum variance in addition to being unbiased, then equality holds. We state this as a theorem without giving a proof.

Theorem 16.2. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a population X with probability density $f(x;\theta)$, where θ is a parameter. If $\widehat{\theta}$ is an unbiased estimator of θ and

$$Var\left(\widehat{\theta}\right) = \frac{1}{E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)^2\right]},$$

then $\widehat{\theta}$ is a uniform minimum variance unbiased estimator of θ. The converse of this is not true.

Definition 16.5. An unbiased estimator $\widehat{\theta}$ is called an efficient estimator if it satisfies Cramér-Rao lower bound, that is

$$Var\left(\widehat{\theta}\right) = \frac{1}{E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)^2\right]}.$$

In view of the above theorem it is easy to note that an efficient estimator of a parameter is always a uniform minimum variance unbiased estimator of a parameter. However, not every uniform minimum variance unbiased estimator of a parameter is efficient. In other words not every uniform minimum variance unbiased estimators of a parameter satisfy the Cramér-Rao lower bound

$$Var\left(\widehat{\theta}\right) \geq \frac{1}{E\left[\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right)^2\right]}.$$

Example 16.11. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with density function

$$f(x;\theta) = \begin{cases} 3\theta\, x^2 e^{-\theta x^3} & \text{if } 0 < x < \infty \\ \\ 0 & \text{otherwise.} \end{cases}$$

What is the Cramér-Rao lower bound for the variance of unbiased estimator of the parameter θ ?

Answer: Let $\widehat{\theta}$ be an unbiased estimator of θ. Cramér-Rao lower bound for the variance of $\widehat{\theta}$ is given by

$$Var\left(\widehat{\theta}\right) \geq \frac{-1}{E\left[\frac{\partial^2 \ln L(\theta)}{\partial \theta^2}\right]},$$

where $L(\theta)$ denotes the likelihood function of the given random sample $X_1, X_2, ..., X_n$. Since, the likelihood function of the sample is

$$L(\theta) = \prod_{i=1}^{n} 3\theta\, x_i^2 e^{-\theta x_i^3}$$

we get

$$\ln L(\theta) = n \ln\theta + \sum_{i=1}^{n} \ln\left(3x_i^2\right) - \theta \sum_{i=1}^{n} x_i^3.$$

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} x_i^3,$$

and

$$\frac{\partial^2 \ln L(\theta)}{\partial \theta^2} = -\frac{n}{\theta^2}.$$

Hence, using this in the Cramér-Rao inequality, we get

$$Var\left(\widehat{\theta}\right) \geq \frac{\theta^2}{n}.$$

Thus the Cramér-Rao lower bound for the variance of the unbiased estimator of θ is $\frac{\theta^2}{n}$.

Example 16.12. Let $X_1, X_2, ..., X_n$ be a random sample from a normal population with unknown mean μ and known variance $\sigma^2 > 0$. What is the maximum likelihood estimator of μ? Is this maximum likelihood estimator an efficient estimator of μ?

Answer: The probability density function of the population is

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

Thus

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

and hence

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

Taking the derivative of $\ln L(\mu)$ with respect to μ, we get

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

Setting this derivative to zero and solving for μ, we see that $\widehat{\mu} = \overline{X}$.

The variance of $\overline{X}$ is given by

$$\begin{aligned} Var\left(\overline{X}\right) &= Var\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

Next we determine the Cramér-Rao lower bound for the estimator $\overline{X}$. We already know that

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

and hence

$$\frac{d^2\ln L(\mu)}{d\mu^2} = -\frac{n}{\sigma^2}.$$

Therefore

$$E\left(\frac{d^2\ln L(\mu)}{d\mu^2}\right) = -\frac{n}{\sigma^2}$$

and

$$-\frac{1}{E\left(\frac{d^2 \ln L(\mu)}{d\mu^2}\right)} = \frac{\sigma^2}{n}.$$

Thus

$$Var\left(\overline{X}\right) = -\frac{1}{E\left(\frac{d^2 \ln L(\mu)}{d\mu^2}\right)}$$

and $\overline{X}$ is an efficient estimator of μ. Since every efficient estimator is a uniform minimum variance unbiased estimator, therefore $\overline{X}$ is a uniform minimum variance unbiased estimator of μ.

Example 16.13. Let $X_1, X_2, ..., X_n$ be a random sample from a normal population with known mean μ and unknown variance $\sigma^2 > 0$. What is the maximum likelihood estimator of σ^2? Is this maximum likelihood estimator a uniform minimum variance unbiased estimator of σ^2?

Answer: Let us write $\theta = \sigma^2$. Then

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

and

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

Differentiating $\ln L(\theta)$ with respect to θ, we have

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

Setting this derivative to zero and solving for θ, we see that

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2.$$

Next we show that this estimator is unbiased. For this we consider

$$\begin{aligned}
E\left(\widehat{\theta}\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right) \\
&= \frac{\sigma^2}{n}\,E\left(\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right) \\
&= \frac{\theta}{n}\,E(\,\chi^2(n)\,) \\
&= \frac{\theta}{n}\,n = \theta.
\end{aligned}$$

Hence $\widehat{\theta}$ is an unbiased estimator of θ. The variance of $\widehat{\theta}$ can be obtained as follows:

$$\begin{aligned}
Var\left(\widehat{\theta}\right) &= Var\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right) \\
&= \frac{\sigma^4}{n}\,Var\left(\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right) \\
&= \frac{\theta^2}{n^2}\,Var(\,\chi^2(n)\,) \\
&= \frac{\theta^2}{n^2}\;4\;\frac{n}{2} \\
&= \frac{2\theta^2}{n} = \frac{2\sigma^4}{n}.
\end{aligned}$$

Finally we determine the Cramér-Rao lower bound for the variance of $\widehat{\theta}$. The second derivative of $\ln L(\theta)$ with respect to θ is

$$\frac{d^2\ln L(\theta)}{d\theta^2} = \frac{n}{2\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Hence

$$\begin{aligned}
E\left(\frac{d^2\ln L(\theta)}{d\theta^2}\right) &= \frac{n}{2\theta^2} - \frac{1}{\theta^3}E\left(\sum_{i=1}^{n}(X_i-\mu)^2\right) \\
&= \frac{n}{2\theta^2} - \frac{\theta}{\theta^3}E\left(\,\chi^2(n)\,\right) \\
&= \frac{n}{2\theta^2} - \frac{n}{\theta^2} \\
&= -\frac{n}{2\theta^2}
\end{aligned}$$

Thus

$$-\frac{1}{E\left(\frac{d^2\ln L(\theta)}{d\theta^2}\right)} = \frac{\theta^2}{n} = \frac{2\sigma^4}{n}.$$

Therefore

$$Var\left(\widehat{\theta}\right) = -\frac{1}{E\left(\frac{d^2\ln L(\theta)}{d\theta^2}\right)}.$$

Hence $\widehat{\theta}$ is an efficient estimator of θ. Since every efficient estimator is a uniform minimum variance unbiased estimator, therefore $\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ is a uniform minimum variance unbiased estimator of σ^2.

Example 16.14. Let $X_1, X_2, ..., X_n$ be a random sample of size n from a normal population known mean μ and variance $\sigma^2 > 0$. Show that $S^2 =$

$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is an unbiased estimator of σ^2. Further, show that S^2 can not attain the Cramér-Rao lower bound.

Answer: From Example 16.2, we know that S^2 is an unbiased estimator of σ^2. The variance of S^2 can be computed as follows:

$$\begin{aligned}
Var\left(S^2\right) &= Var\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right)\\
&= \frac{\sigma^4}{(n-1)^2}\,Var\left(\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2\right)\\
&= \frac{\sigma^4}{(n-1)^2}\,Var(\,\chi^2(n-1)\,)\\
&= \frac{\sigma^4}{(n-1)^2}\;2\;(n-1)\\
&= \frac{2\sigma^4}{n-1}.
\end{aligned}$$

Next we let $\theta = \sigma^2$ and determine the Cramér-Rao lower bound for the variance of S^2. The second derivative of $\ln L(\theta)$ with respect to θ is

$$\frac{d^2\ln L(\theta)}{d\theta^2} = \frac{n}{2\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Hence

$$\begin{aligned}
E\left(\frac{d^2\ln L(\theta)}{d\theta^2}\right) &= \frac{n}{2\theta^2} - \frac{1}{\theta^3}E\left(\sum_{i=1}^{n}(X_i-\mu)^2\right)\\
&= \frac{n}{2\theta^2} - \frac{\theta}{\theta^3}E\left(\,\chi^2(n)\,\right)\\
&= \frac{n}{2\theta^2} - \frac{n}{\theta^2}\\
&= -\frac{n}{2\theta^2}
\end{aligned}$$

Thus

$$-\frac{1}{E\left(\frac{d^2\ln L(\theta)}{d\theta^2}\right)} = \frac{\theta^2}{n} = \frac{2\sigma^4}{n}.$$

Hence

$$\frac{2\sigma^4}{n-1} = Var\left(S^2\right) > -\frac{1}{E\left(\frac{d^2\ln L(\theta)}{d\theta^2}\right)} = \frac{2\sigma^4}{n}.$$

This shows that S^2 can not attain the Cramér-Rao lower bound.

The disadvantages of Cramér-Rao lower bound approach are the followings: (1) Not every density function $f(x;\theta)$ satisfies the assumptions of Cramér-Rao theorem and (2) not every allowable estimator attains the Cramér-Rao lower bound. Hence in any one of these situations, one does not know whether an estimator is a uniform minimum variance unbiased estimator or not.

16.4. Sufficient Estimator

In many situations, we can not easily find the distribution of the estimator $\widehat{\theta}$ of a parameter θ even though we know the distribution of the population. Therefore, we have no way to know whether our estimator $\widehat{\theta}$ is unbiased or biased. Hence, we need some other criteria to judge the quality of an estimator. Sufficiency is one such criteria for judging the quality of an estimator.

Recall that an estimator of a population parameter is a function of the sample values that does not contain the parameter. An estimator summarizes the information found in the sample about the parameter. If an estimator summarizes just as much information about the parameter being estimated as the sample does, then the estimator is called a sufficient estimator.

Definition 16.6. Let $X \sim f(x;\theta)$ be a population and let $X_1, X_2, ..., X_n$ be a random sample of size n from this population X. An estimator $\widehat{\theta}$ of the parameter θ is said to be a sufficient estimator of θ if the conditional distribution of the sample given the estimator $\widehat{\theta}$ does not depend on the parameter θ.

Example 16.15. If $X_1, X_2, ..., X_n$ is a random sample from the distribution with probability density function

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

where $0 < \theta < 1$. Show that $Y = \sum_{i=1}^n X_i$ is a sufficient statistic of θ.

Answer: First, we find the distribution of the sample. This is given by

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

Since, each $X_i \sim BER(\theta)$, we have

$$Y = \sum_{i=1}^n X_i \sim BIN(n, \theta).$$

Therefore, the probability density function of Y is given by

$$g(y) = \binom{n}{y} \theta^y \, (1-\theta)^{n-y}.$$

Further, since each $X_i \sim BER(\theta)$, the space of each X_i is given by

$$R_{X_i} = \{\, 0, 1 \,\}.$$

Therefore, the space of the random variable $Y = \sum_{i=1}^{n} X_i$ is given by

$$R_Y = \{\, 0, 1, 2, 3, 4, ..., n \,\}.$$

Let A be the event $(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$ and B denotes the event $(Y = y)$. Then $A \subset B$ and therefore $A \bigcap B = A$. Now, we find the conditional density of the sample given the estimator Y, that is

$$\begin{aligned}
f(x_1, x_2, ..., x_n/Y = y) &= P\left(X_1 = x_1, X_2 = x_2, ..., X_n = x_n \,/ Y = y\right) \\
&= P(A/B) \\
&= \frac{P(A \bigcap B)}{P(B)} \\
&= \frac{P(A)}{P(B)} \\
&= \frac{f(x_1, x_2, ..., x_n)}{g(y)} \\
&= \frac{\theta^y (1-\theta)^{n-y}}{\binom{n}{y} \theta^y \, (1-\theta)^{n-y}} \\
&= \frac{1}{\binom{n}{y}}.
\end{aligned}$$

Hence, the conditional density of the sample given the statistic Y is independent of the parameter θ. Therefore, by definition Y is a sufficient statistic.

Example 16.16. If $X_1, X_2, ..., X_n$ is a random sample from the distribution with probability density function

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

where $-\infty < \theta < \infty$. What is the maximum likelihood estimator of θ ? Is this maximum likelihood estimator sufficient estimator of θ ?

Answer: We have seen in Chapter 15 that the maximum likelihood estimator of θ is $Y = X_{(1)}$, that is the first order statistic of the sample. Let us find the probability density of this statistic, which is given by

$$\begin{aligned} g(y) &= \frac{n!}{(n-1)!} \, [F(y)]^0 \, f(y) \, [1 - F(y)]^{n-1} \\ &= n \, f(y) \, [1 - F(y)]^{n-1} \\ &= n \, e^{-(y-\theta)} \, \left[1 - \left\{1 - e^{-(y-\theta)}\right\}\right]^{n-1} \\ &= n \, e^{n\theta} \, e^{-ny}. \end{aligned}$$

The probability density of the random sample is

$$\begin{aligned} f(x_1, x_2, ..., x_n) &= \prod_{i=1}^{n} e^{-(x_i - \theta)} \\ &= e^{n\theta} \, e^{-n\,\overline{x}}, \end{aligned}$$

where $n\overline{x} = \displaystyle\sum_{i=1}^{n} x_i$. Let A be the event $(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$ and B denotes the event $(Y = y)$. Then $A \subset B$ and therefore $A \bigcap B = A$. Now, we find the conditional density of the sample given the estimator Y, that is

$$\begin{aligned} f(x_1, x_2, ..., x_n / Y = y) &= P\left(X_1 = x_1, X_2 = x_2, ..., X_n = x_n \, / Y = y\right) \\ &= P(A/B) \\ &= \frac{P(A \bigcap B)}{P(B)} \\ &= \frac{P(A)}{P(B)} \\ &= \frac{f(x_1, x_2, ..., x_n)}{g(y)} \\ &= \frac{e^{n\theta} \, e^{-n\,\overline{x}}}{n \, e^{n\theta} \, e^{-n\,y}} \\ &= \frac{e^{-n\,\overline{x}}}{n \, e^{-n\,y}}. \end{aligned}$$

Hence, the conditional density of the sample given the statistic Y is independent of the parameter θ. Therefore, by definition Y is a sufficient statistic.

We have seen that to verify whether an estimator is sufficient or not one has to examine the conditional density of the sample given the estimator. To compute this conditional density one has to use the density of the estimator. The density of the estimator is not always easy to find. Therefore, verifying the sufficiency of an estimator using this definition is not always easy. The following *factorization theorem* of Fisher and Neyman helps to decide when an estimator is sufficient.

Theorem 16.3. Let $X_1, X_2, ..., X_n$ denote a random sample with probability density function $f(x_1, x_2, ..., x_n; \theta)$, which depends on the population parameter θ. The estimator $\widehat{\theta}$ is sufficient for θ if and only if

$$f(x_1, x_2, ..., x_n; \theta) = \phi(\widehat{\theta}, \theta)\, h(x_1, x_2, ..., x_n)$$

where ϕ depends on $x_1, x_2, ..., x_n$ only through $\widehat{\theta}$ and $h(x_1, x_2, ..., x_n)$ does not depend on θ.

Now we give two examples to illustrate the factorization theorem.

Example 16.17. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density function

$$f(x; \lambda) = \begin{cases} \frac{\lambda^x\, e^{-\lambda}}{x!} & \text{if } x = 0, 1, 2, ..., \infty \\ \\ 0 & \text{elsewhere,} \end{cases}$$

where $\lambda > 0$ is a parameter. Find the maximum likelihood estimator of λ and show that the maximum likelihood estimator of λ is sufficient estimator of the parameter λ.

Answer: First, we find the density of the sample or the likelihood function of the sample. The likelihood function of the sample is given by

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n} f(x_i; \lambda) \\ &= \prod_{i=1}^{n} \frac{\lambda^{x_i}\, e^{-\lambda}}{x_i!} \\ &= \frac{\lambda^{n\overline{X}} e^{-n\lambda}}{\prod_{i=1}^{n} (x_i!)}. \end{aligned}$$

Taking the logarithm of the likelihood function, we get

$$\ln L(\lambda) = n\overline{x}\ \ln\lambda - n\lambda - \ln\prod_{i=1}^{n}(x_i!).$$

Therefore

$$\frac{d}{d\lambda}\ln L(\lambda) = \frac{1}{\lambda}\,n\overline{x} - n.$$

Setting this derivative to zero and solving for λ, we get

$$\lambda = \overline{x}.$$

The second derivative test assures us that the above λ is a maximum. Hence, the maximum likelihood estimator of λ is the sample mean $\overline{X}$. Next, we show that $\overline{X}$ is sufficient, by using the Factorization Theorem of Fisher and Neyman. We factor the joint density of the sample as

$$\begin{aligned} L(\lambda) &= \frac{\lambda^{n\overline{x}}e^{-n\lambda}}{\prod_{i=1}^{n}(x_i!)} \\ &= \left[\lambda^{n\overline{x}}\,e^{-n\lambda}\right]\ \frac{1}{\prod_{i=1}^{n}(x_i!)} \\ &= \phi(\overline{X},\lambda)\,h\,(x_1,x_2,...,x_n)\,. \end{aligned}$$

Therefore, the estimator $\overline{X}$ is a sufficient estimator of λ.

Example 16.18. Let $X_1, X_2, ..., X_n$ be a random sample from a normal distribution with density function

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

where $-\infty < \mu < \infty$ is a parameter. Find the maximum likelihood estimator of μ and show that the maximum likelihood estimator of μ is a sufficient estimator.

Answer: We know that the maximum likelihood estimator of μ is the sample mean $\overline{X}$. Next, we show that this maximum likelihood estimator $\overline{X}$ is a sufficient estimator of μ. The joint density of the sample is given by

$$
\begin{aligned}
f(x_1, x_2, ..., x_n; \mu) \\
&= \prod_{i=1}^{n} f(x_i; \mu) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}(x_i-\mu)^2} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}\left[(x_i-\overline{x})+(\overline{x}-\mu)\right]^2} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}\left[(x_i-\overline{x})^2+2(x_i-\overline{x})(\overline{x}-\mu)+(\overline{x}-\mu)^2\right]} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}\left[(x_i-\overline{x})^2+(\overline{x}-\mu)^2\right]} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{n}{2}(\overline{x}-\mu)^2} \, e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2}
\end{aligned}
$$

Hence, by the Factorization Theorem, $\overline{X}$ is a sufficient estimator of the population mean.

Note that the probability density function of the Example 16.17 which is

$$
f(x;\lambda) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{if } x = 0, 1, 2, ..., \infty \\ \\ 0 & \text{elsewhere ,} \end{cases}
$$

can be written as

$$
f(x;\lambda) = e^{\{x \ln\lambda - \ln x! - \lambda\}}
$$

for $x = 0, 1, 2, ...$ This density function is of the form

$$
f(x;\lambda) = e^{\{K(x)A(\lambda)+S(x)+B(\lambda)\}}.
$$

Similarly, the probability density function of the Example 16.12, which is

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

can also be written as

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

This probability density function is of the form

$$f(x;\mu) = e^{\{K(x)A(\mu)+S(x)+B(\mu)\}}.$$

We have also seen that in both the examples, the sufficient estimators were the sample mean $\overline{X}$, which can be written as $\frac{1}{n}\sum_{i=1}^{n} X_i$.

Our next theorem gives a general result in this direction. The following theorem is known as the Pitman-Koopman theorem.

Theorem 16.4. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with probability density function of the exponential form

$$f(x;\theta) = e^{\{K(x)A(\theta)+S(x)+B(\theta)\}}$$

on a support free of θ. Then the statistic $\sum_{i=1}^{n} K(X_i)$ is a sufficient statistic for the parameter θ.

Proof: The joint density of the sample is

$$\begin{aligned} f(x_1, x_2, ..., x_n; \theta) &= \prod_{i=1}^{n} f(x_i;\theta) \\ &= \prod_{i=1}^{n} e^{\{K(x_i)A(\theta)+S(x_i)+B(\theta)\}} \\ &= e^{\left\{\sum_{i=1}^{n} K(x_i)A(\theta) + \sum_{i=1}^{n} S(x_i) + n\,B(\theta)\right\}} \\ &= e^{\left\{\sum_{i=1}^{n} K(x_i)A(\theta) + n\,B(\theta)\right\}}\, e^{\left[\sum_{i=1}^{n} S(x_i)\right]}. \end{aligned}$$

Hence by the Factorization Theorem the estimator $\sum_{i=1}^{n} K(X_i)$ is a sufficient statistic for the parameter θ. This completes the proof.

Example 16.19. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density function

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

where $\theta > 0$ is a parameter. Using the Pitman-Koopman Theorem find a sufficient estimator of θ.

Answer: The Pitman-Koopman Theorem says that if the probability density function can be expressed in the form of

$$f(x;\theta) = e^{\{K(x)A(\theta)+S(x)+B(\theta)\}}$$

then $\sum_{i=1}^{n} K(X_i)$ is a sufficient statistic for θ. The given population density can be written as

$$\begin{aligned} f(x;\theta) &= \theta\, x^{\theta-1} \\ &= e^{\{\ln\left[\theta\, x^{\theta-1}\right]} \\ &= e^{\{\ln\theta + (\theta-1)\ln x\}}. \end{aligned}$$

Thus,

$$\begin{aligned} K(x) &= \ln x & \qquad A(\theta) &= \theta \\ S(x) &= -\ln x & \qquad B(\theta) &= \ln\theta. \end{aligned}$$

Hence by Pitman-Koopman Theorem,

$$\begin{aligned} \sum_{i=1}^{n} K(X_i) &= \sum_{i=1}^{n} \ln X_i \\ &= \ln \prod_{i=1}^{n} X_i. \end{aligned}$$

Thus $\ln \prod_{i=1}^{n} X_i$ is a sufficient statistic for θ.

Remark 16.1. Notice that $\displaystyle\prod_{i=1}^{n} X_i$ is also a sufficient statistic of θ, since knowing $\ln\left(\displaystyle\prod_{i=1}^{n} X_i\right)$, we also know $\displaystyle\prod_{i=1}^{n} X_i$.

Example 16.20. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density function

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

where $0 < \theta < \infty$ is a parameter. Find a sufficient estimator of θ.

Answer: First, we rewrite the population density in the exponential form. That is

$$\begin{aligned} f(x;\theta) &= \frac{1}{\theta}\, e^{-\frac{x}{\theta}} \\ &= e^{\ln\left[\frac{1}{\theta}\, e^{-\frac{x}{\theta}}\right]} \\ &= e^{-\ln\theta - \frac{x}{\theta}}. \end{aligned}$$

Hence

$$\begin{aligned} K(x) &= x & A(\theta) &= -\frac{1}{\theta} \\ S(x) &= 0 & B(\theta) &= -\ln\theta. \end{aligned}$$

Hence by Pitman-Koopman Theorem,

$$\sum_{i=1}^{n} K(X_i) = \sum_{i=1}^{n} X_i = n\,\overline{X}.$$

Thus, $n\overline{X}$ is a sufficient statistic for θ. Since knowing $n\overline{X}$, we also know $\overline{X}$, the estimator $\overline{X}$ is also a sufficient estimator of θ.

Example 16.21. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with density function

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

where $-\infty < \theta < \infty$ is a parameter. Can Pitman-Koopman Theorem be used to find a sufficient statistic for θ?

Answer: No. We can not use Pitman-Koopman Theorem to find a sufficient statistic for θ since the domain where the population density is nonzero is not free of θ.

Next, we present the connection between the maximum likelihood estimator and the sufficient estimator. If there is a sufficient estimator for the parameter θ and if the maximum likelihood estimator of this θ is unique, then the maximum likelihood estimator is a function of the sufficient estimator. That is

$$\widehat{\theta}_{\mathrm{ML}} = \psi(\widehat{\theta}_{\mathrm{S}}),$$

where ψ is a real valued function, $\widehat{\theta}_{\mathrm{ML}}$ is the maximum likelihood estimator of θ, and $\widehat{\theta}_{\mathrm{S}}$ is the sufficient estimator of θ.

Similarly, a connection can be established between the uniform minimum variance unbiased estimator and the sufficient estimator of a parameter θ. If there is a sufficient estimator for the parameter θ and if the uniform minimum variance unbiased estimator of this θ is unique, then the uniform minimum variance unbiased estimator is a function of the sufficient estimator. That is

$$\widehat{\theta}_{\mathrm{MVUE}} = \eta(\widehat{\theta}_{\mathrm{S}}),$$

where η is a real valued function, $\widehat{\theta}_{\mathrm{MVUE}}$ is the uniform minimum variance unbiased estimator of θ, and $\widehat{\theta}_{\mathrm{S}}$ is the sufficient estimator of θ.

Finally, we may ask "If there are sufficient estimators, why are not there necessary estimators?" In fact, there are. Dynkin (1951) gave the following definition.

Definition 16.7. An estimator is said to be a necessary estimator if it can be written as a function of every sufficient estimators.

16.5. Consistent Estimator

Let $X_1, X_2, ..., X_n$ be a random sample from a population X with density $f(x;\theta)$. Let $\widehat{\theta}$ be an estimator of θ based on the sample of size n. Obviously the estimator depends on the sample size n. In order to reflect the dependency of $\widehat{\theta}$ on n, we denote $\widehat{\theta}$ as $\widehat{\theta}_n$.

Definition 16.7. Let $X_1, X_2, ..., X_n$ be a random sample from a population X with density $f(x;\theta)$. A sequence of estimators $\{\widehat{\theta}_n\}$ of θ is said to be consistent for θ if and only if the sequence $\{\widehat{\theta}_n\}$ converges in probability to θ, that is, for any $\epsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\widehat{\theta}_n - \theta\right| \geq \epsilon\right) = 0.$$

Note that consistency is actually a concept relating to a sequence of estimators $\{\widehat{\theta}_n\}_{n=n_o}^{\infty}$ but we usually say "consistency of $\widehat{\theta}_n$" for simplicity. Further, consistency is a large sample property of an estimator.

The following theorem states that if the mean squared error goes to zero as n goes to infinity, then $\{\widehat{\theta}_n\}$ converges in probability to θ.

Theorem 16.5. Let $X_1, X_2, ..., X_n$ be a random sample from a population X with density $f(x;\theta)$ and $\{\widehat{\theta}_n\}$ be a sequence of estimators of θ based on the sample. If the variance of $\widehat{\theta}_n$ exists for each n and is finite and

$$\lim_{n\to\infty} E\left(\left(\widehat{\theta}_n - \theta\right)^2\right) = 0$$

then, for any $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\widehat{\theta}_n - \theta\right| \geq \epsilon\right) = 0.$$

Proof: By Markov Inequality (see Theorem 13.8) we have

$$P\left(\left(\widehat{\theta_n} - \theta\right)^2 \geq \epsilon^2\right) \leq \frac{E\left(\left(\widehat{\theta_n} - \theta\right)^2\right)}{\epsilon^2}$$

for all $\epsilon > 0$. Since the events

$$\left(\widehat{\theta_n} - \theta\right)^2 \geq \epsilon^2 \qquad \text{and} \qquad |\widehat{\theta_n} - \theta| \geq \epsilon$$

are same, we see that

$$P\left(\left(\widehat{\theta_n} - \theta\right)^2 \geq \epsilon^2\right) = P\left(|\widehat{\theta_n} - \theta| \geq \epsilon\right) \leq \frac{E\left(\left(\widehat{\theta_n} - \theta\right)^2\right)}{\epsilon^2}$$

for all $n \in \mathbb{N}$. Hence if

$$\lim_{n\to\infty} E\left(\left(\widehat{\theta_n} - \theta\right)^2\right) = 0$$

then

$$\lim_{n\to\infty} P\left(|\widehat{\theta_n} - \theta| \geq \epsilon\right) = 0$$

and the proof of the theorem is complete.

Let

$$B\left(\widehat{\theta}, \theta\right) = E\left(\widehat{\theta}\right) - \theta$$

be the biased. If an estimator is unbiased, then $B\left(\widehat{\theta}, \theta\right) = 0$. Next we show that

$$E\left(\left(\widehat{\theta} - \theta\right)^2\right) = Var\left(\widehat{\theta}\right) + \left[B\left(\widehat{\theta}, \theta\right)\right]^2. \tag{1}$$

To see this consider

$$\begin{aligned}
E\left(\left(\widehat{\theta} - \theta\right)^2\right) &= E\left(\left(\widehat{\theta}^2 - 2\,\widehat{\theta}\,\theta + \theta^2\right)^2\right) \\
&= E\left(\widehat{\theta}^2\right) - 2E\left(\widehat{\theta}\right)\,\theta + \theta^2 \\
&= E\left(\widehat{\theta}^2\right) - E\left(\widehat{\theta}\right)^2 + E\left(\widehat{\theta}\right)^2 - 2E\left(\widehat{\theta}\right)\,\theta + \theta^2 \\
&= Var\left(\widehat{\theta}\right) + E\left(\widehat{\theta}\right)^2 - 2E\left(\widehat{\theta}\right)\,\theta + \theta^2 \\
&= Var\left(\widehat{\theta}\right) + \left[E\left(\widehat{\theta}\right) - \theta\right]^2 \\
&= Var\left(\widehat{\theta}\right) + \left[B\left(\widehat{\theta}, \theta\right)\right]^2.
\end{aligned}$$

In view of (1), we can say that if

$$\lim_{n\to\infty} Var\left(\widehat{\theta}_n\right) = 0 \tag{2}$$

and

$$\lim_{n\to\infty} B\left(\widehat{\theta}_n, \theta\right) = 0 \tag{3}$$

then

$$\lim_{n\to\infty} E\left(\left(\widehat{\theta}_n - \theta\right)^2\right) = 0.$$

In other words, to show a sequence of estimators is consistent we have to verify the limits (2) and (3).

Example 16.22. Let $X_1, X_2, ..., X_n$ be a random sample from a normal population X with mean μ and variance $\sigma^2 > 0$. Is the likelihood estimator

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

of σ^2 a consistent estimator of σ^2?

Answer: Since $\widehat{\sigma^2}$ depends on the sample size n, we denote $\widehat{\sigma^2}$ as $\widehat{\sigma^2}_n$. Hence

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

The variance of $\widehat{\sigma^2}_n$ is given by

$$\begin{aligned}
Var\left(\widehat{\sigma^2}_n\right) &= Var\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right)\\
&= \frac{1}{n^2}\,Var\left(\sigma^2\,\frac{(n-1)S^2}{\sigma^2}\right)\\
&= \frac{\sigma^4}{n^2}\,Var\left(\frac{(n-1)S^2}{\sigma^2}\right)\\
&= \frac{\sigma^4}{n^2}\,Var\left(\chi^2(n-1)\right)\\
&= \frac{2(n-1)\sigma^4}{n^2}\\
&= \left[\frac{1}{n} - \frac{1}{n^2}\right] 2\,\sigma^4.
\end{aligned}$$

Hence

$$\lim_{n\to\infty} Var\left(\widehat{\theta}_n\right) = \lim_{n\to\infty} \left[\frac{1}{n} - \frac{1}{n^2}\right] 2\,\sigma^4 = 0.$$

The biased $B\left(\widehat{\theta}_n, \theta\right)$ is given by

$$\begin{aligned} B\left(\widehat{\theta}_n, \theta\right) &= E\left(\widehat{\theta}_n\right) - \sigma^2 \\ &= E\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right) - \sigma^2 \\ &= \frac{1}{n}\,E\left(\sigma^2\,\frac{(n-1)S^2}{\sigma^2}\right) - \sigma^2 \\ &= \frac{\sigma^2}{n} E\left(\chi^2(n-1)\right) - \sigma^2 \\ &= \frac{(n-1)\sigma^2}{n} - \sigma^2 \\ &= -\frac{\sigma^2}{n}. \end{aligned}$$

Thus

$$\lim_{n\to\infty} B\left(\widehat{\theta}_n, \theta\right) = -\lim_{n\to\infty} \frac{\sigma^2}{n} = 0.$$

Hence $\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ is a consistent estimator of σ^2.

In the last example we saw that the likelihood estimator of variance is a consistent estimator. In general, if the density function $f(x;\theta)$ of a population satisfies some mild conditions, then the maximum likelihood estimator of θ is consistent. Similarly, if the density function $f(x;\theta)$ of a population satisfies some mild conditions, then the estimator obtained by moment method is also consistent.

Since consistency is a large sample property of an estimator, some statisticians suggest that consistency should not be used alone for judging the goodness of an estimator; rather it should be used along with other criteria.

16.6. Review Exercises

1. Let T_1 and T_2 be estimators of a population parameter θ based upon the same random sample. If $T_i \sim N\left(\theta, \sigma_i^2\right)$ $i = 1, 2$ and if $T = bT_1 + (1-b)T_2$, then for what value of b, T is a minimum variance unbiased estimator of θ ?

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

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

where $0 < \theta$ is a parameter. What is the expected value of the maximum likelihood estimator of θ ? Is this estimator unbiased?

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

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

where $0 < \theta$ is a parameter. Is the maximum likelihood estimator an efficient estimator of θ?

4. A random sample $X_1, X_2, ..., X_n$ of size n is selected from a normal distribution with variance σ^2. Let S^2 be the unbiased estimator of σ^2, and T be the maximum likelihood estimator of σ^2. If $20T - 19S^2 = 0$, then what is the sample size?

5. Suppose X and Y are independent random variables each with density function

$$f(x) = \begin{cases} 2\,x\,\theta^2 & \text{for } 0 < x < \frac{1}{\theta} \\ \\ 0 & \text{otherwise.} \end{cases}$$

If $k\,(X + 2Y)$ is an unbiased estimator of θ^{-1}, then what is the value of k?

6. An object of length c is measured by two persons using the same instrument. The instrument error has a normal distribution with mean 0 and variance 1. The first person measures the object 25 times, and the average of the measurements is $\bar{X} = 12$. The second person measures the objects 36 times, and the average of the measurements is $\bar{Y} = 12.8$. To estimate c we use the weighted average $a\,\bar{X} + b\,\bar{Y}$ as an estimator. Determine the constants a and b such that $a\,\bar{X} + b\,\bar{Y}$ is the minimum variance unbiased estimator of c and then calculate the minimum variance unbiased estimate of c.

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

$$f(x) = \begin{cases} 3\,\theta\,x^2\,e^{-\theta\,x^3} & \text{for } 0 < x < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is an unknown parameter. Find a sufficient statistics for θ.

8. Let $X_1, X_2, ..., X_n$ be a random sample from a Weibull distribution with probability density function

$$f(x) = \begin{cases} \frac{\beta}{\theta^\beta}\, x^{\beta-1}\, e^{-\left(\frac{x}{\theta}\right)^\beta} & \text{if } x > 0 \\ \\ 0 & \text{otherwise ,} \end{cases}$$

where $\theta > 0$ and $\beta > 0$ are parameters. Find a sufficient statistics for θ with β known, say $\beta = 2$. If β is unknown, can you find a single sufficient statistics for θ?

9. Let X_1, X_2 be a random sample of size 2 from population with probability density

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

where $\theta > 0$ is an unknown parameter. If $Y = \sqrt{X_1 X_2}$, then what should be the value of the constant k such that kY is an unbiased estimator of the parameter θ ?

10. Let $X_1, X_2, ..., X_n$ be a random sample from a population with probability density function

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

where $\theta > 0$ is an unknown parameter. If $\overline{X}$ denotes the sample mean, then what should be value of the constant k such that $k\overline{X}$ is an unbiased estimator of θ ?

11. Let $X_1, X_2, ..., X_n$ be a random sample from a population with probability density function

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

where $\theta > 0$ is an unknown parameter. If X_{med} denotes the sample median, then what should be value of the constant k such that kX_{med} is an unbiased estimator of θ ?

12. What do you understand by an unbiased estimator of a parameter θ? What is the basic principle of the maximum likelihood estimation of a parameter θ? What is the basic principle of the Bayesian estimation of a parameter θ? What is the main difference between Bayesian method and likelihood method.

13. Let $X_1, X_2, ..., X_n$ be a random sample from a population X with density function

$$f(x;\theta) = \begin{cases} \frac{\theta}{(1+x)^{\theta+1}} & \text{for } 0 \leq x < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is an unknown parameter. What is a sufficient statistic for the parameter θ?

14. Let $X_1, X_2, ..., X_n$ be a random sample from a population X with density function

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

where θ is an unknown parameter. What is a sufficient statistic for the parameter θ?

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

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

where $-\infty < \theta < \infty$ is a parameter. What is the maximum likelihood estimator of θ? Find a sufficient statistics of the parameter θ.

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

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

where $-\infty < \theta < \infty$ is a parameter. Are the estimators $X_{(1)}$ and $\overline{X} - 1$ are unbiased estimators of θ? Which one is more efficient than the other?

17. Let $X_1, X_2, ..., X_n$ be a random sample from a population X with density function

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

where $\theta > 1$ is an unknown parameter. What is a sufficient statistic for the parameter θ?

18. Let $X_1, X_2, ..., X_n$ be a random sample from a population X with density function

$$f(x;\theta) = \begin{cases} \theta\,\alpha\,x^{\alpha-1} e^{-\theta x^{\alpha}} & \text{for } 0 \leq x < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ and $\alpha > 0$ are parameters. What is a sufficient statistic for the parameter θ for a fixed α?

19. Let $X_1, X_2, ..., X_n$ be a random sample from a population X with density function

$$f(x;\theta) = \begin{cases} \frac{\theta\,\alpha^{\theta}}{x^{(\theta+1)}} & \text{for } \alpha < x < \infty \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ and $\alpha > 0$ are parameters. What is a sufficient statistic for the parameter θ for a fixed α?

20. Let $X_1, X_2, ..., X_n$ be a random sample from a population X with density function

$$f(x;\theta) = \begin{cases} \binom{m}{x} \theta^{x} (1-\theta)^{m-x} & \text{for } x = 0, 1, 2, ..., m \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < \theta < 1$ is parameter. Show that $\frac{\overline{X}}{m}$ is a uniform minimum variance unbiased estimator of θ for a fixed m.

21. Let $X_1, X_2, ..., X_n$ be a random sample from a population X with density function

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

where $\theta > 1$ is parameter. Show that $-\frac{1}{n}\sum_{i=1}^{n} \ln(X_i)$ is a uniform minimum variance unbiased estimator of $\frac{1}{\theta}$.

22. Let $X_1, X_2, ..., X_n$ be a random sample from a uniform population X on the interval $[0, \theta]$, where $\theta > 0$ is a parameter. Is the likelihood estimator $\widehat{\theta} = X_{(n)}$ of θ a consistent estimator of θ?

23. Let $X_1, X_2, ..., X_n$ be a random sample from a population $X \sim POI(\lambda)$, where $\lambda > 0$ is a parameter. Is the estimator $\overline{X}$ of λ a consistent estimator of λ?

24. Let $X_1, X_2, ..., X_n$ be a random sample from a population X having the probability density function

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

where $\theta > 0$ is a parameter. Is the estimator $\widehat{\theta} = \frac{\overline{X}}{1-\overline{X}}$ of θ, obtained by the moment method, a consistent estimator of θ?

25. Let $X_1, X_2, ..., X_n$ be a random sample from a population X having the probability density function

$$f(x;p) = \begin{cases} \binom{n}{x}\, p^x\, (1-p)^{n-x}, & \text{if } x = 0, 1, 2, ..., n \\ \\ 0 & \text{otherwise,} \end{cases}$$

where $0 < p < 1$ is a parameter and m is a fixed positive integer. What is the maximum likelihood estimator for p. Is this maximum likelihood estimator for p is an efficient estimator?

26. Let $X_1, X_2, ..., X_n$ be a random sample from a population X having the probability density function

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

where $\theta > 0$ is a parameter. Find an estimator for θ using the moment method.

27. A box contains 50 red and blue balls out of which θ are red. A sample of 30 balls is to be selected without replacement. If X denotes the number of red balls in the sample, then find an estimator for θ using the moment method.