CHAPTER SEVENTEEN

Statistical Inference-I (Theory of Estimation)

LEARNING OBJECTIVES. Upon completion of this chapter, you should be able to :

1. Differentiate between Point Estimate and Interval Estimate.
2. Understand and discuss various characteristics of estimators like Consistency, Unbiasedness, Efficiency and Sufficiency along with their importance in estimation theory.
3. Know and understand about Cramer-Rao Inequality regarding the lower bound to the variance of an unbiased estimator.
4. Distinguish between Minimum Variance Unbiased (MVU) Estimator and Minimum Variance Bound (MVB) Unbised Estimator.
5. Understand the various methods of estimation like methods of (*i*) 'Maximum Likelihood Estimation', (*ii*) 'Minimum Variance', (*iii*) 'Moments' and (*iv*) 'Least Squares'; along with their important properties.
6. Explain what is meant by confidence interval and confidence limits.

CHAPTER OUTLINE

→ Point Estimates

→ Interval estimate

17·1. INTRODUCTION

One of the main objectives of Statistics is to draw inferences about a population from the analysis of a sample drawn from that population. Two important problems in statistical inference are (*i*) estimation and (*ii*) testing of hypothesis.

The theory of estimation was founded by Prof. R.A. Fisher in a series of fundamental papers round about 1930.

Parameter Space. Let us consider a random variable X with *p.d.f.* $f(x, \theta)$. In most common applications, though not always, the functional form of the population distribution is assumed to be known except for the value of some unknown parameter(s) θ which may take any value on a set Θ. This is expressed by writing the *p.d.f.* in the form $f(x, \theta)$, $\theta \in \Theta$. The set Θ, which is the set of all possible values of θ is called the *parameter space*. Such a situation gives rise not to one probability distribution but a family of probability distributions which we write as $\{f(x, \theta), \theta \in \Theta\}$, *e.g.*, if $X \sim N(\mu, \sigma^2)$, then the parameter space $\Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty;\ 0 < \sigma < \infty\}$.

In particular, for $\sigma^2 = 1$, the family of probability distributions is given by :

$$\{N(\mu, 1)\ ;\ \mu \in \Theta\}, \text{ where } \Theta = \{\mu : -\infty < \mu < \infty\}$$

In the following discussion we shall consider a general family of distributions:

$$\{f(x\ ;\theta_1, \theta_2, \ldots, \theta_k) : \theta_i \in \Theta,\ i = 1, 2, \ldots, k\}.$$

Let us consider a random sample $x_1, x_2, \ldots, x_n$ of size n from a population, with probability function $f(x\ ;\theta_1, \theta_2, \ldots, \theta_k)$, where $\theta_1, \theta_2, \ldots, \theta_k$ are the unknown population parameters. There will then always be an infinite number of functions of sample values, called statistics, which may be proposed as estimates of one or more of the parameters.

Evidently, the best estimate would be one that falls nearest to the true value of the parameter to be estimated. In other words, the statistic whose distribution concentrates as closely as possible near the true value of the parameter may be regarded the best estimate. Hence the basic problem of the estimation in the above case, can be formulated as follows :

'We wish to determine the functions of the sample observations :

$$T_1 = \hat{\theta}_1(x_1, x_2, \ldots, x_n),\ T_2 = \hat{\theta}_2(x_1, x_2, \ldots, x_n), \ldots, T_k = \hat{\theta}_k(x_1, x_2, \ldots, x_n),$$

such that their distribution is concentrated as closely as possible near the true value of the parameter. The estimating functions are then referred to as *estimators*.

Definition. *Any function of the random sample $x_1, x_2, \ldots, x_n$ that are being observed, say $T_n(x_1, x_2, \ldots, x_n)$ is called a statistic. Clearly, a statistic is a random variable. If it is used to estimate an unknown parameter θ of the distribution, it is called an estiamtor. A particular value of the estimator, say, $T_n(x_1, x_2, \ldots, x_n)$ is called an estiamte of θ.*

We shall, however, use the terms *estimator* and *estimate*, somewhat loosely, their actual implications being clear from the context.

17·2. CHARACTERISTICS OF ESTIMATORS.

The following are some of the criteria that should be satisfied by a good estimator.

(*i*) Unbiasedness, (*ii*) Consistency, (*iii*) Efficiency, and (*iv*) Sufficiency. We shall now, briefly, explain these terms one by one.

17·2·1. Unbiasedness.

Definition. *An estimator $T_n = T(x_1, x_2, \ldots, x_n)$ is said to be an unbaised estimator of $\gamma(\theta)$ if*

$$E(T_n) = \gamma(\theta), \text{ for all } \theta \in \Theta \quad \ldots (17\cdot1)$$

We have seen in chapter 13 that in sampling from a population with mean μ and variance σ^2, $E(\bar{x}) = \mu$ and $E(s^2) \neq \sigma^2$ but $E(S^2) = \sigma^2$. Hence there is a reason to prefer

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \text{ to the sample variance } s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Remark. If $E(T_n) > \theta$, T_n is said to be positively biased and if $E(T_n) < \theta$, it is said to be negatively biased, the amount of bias $b(\theta)$ being given by $\quad b(\theta) = E(T_n) - \gamma(\theta),\ \theta \in \Theta \quad \ldots(17\cdot1a)$

Example 17·1. *$x_1, x_2, \ldots, x_n$ is a random sample from a normal population $N(\mu, 1)$. Show that $t = \frac{1}{n}\sum_{i=1}^{n} x_i^2$, is an unbiased estimator of $\mu^2 + 1$.*

Solution. (*a*) We are given : $\quad E(x_i) = \mu,\ V(x_i) = 1\ \forall\ i = 1, 2, \ldots, n \quad \ldots (*)$

Now $\quad E(x_i^2) = V(x_i) + \{E(x_i)\}^2 = 1 + \mu^2 \quad$ [From (*)]

$$\therefore \quad E(t) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i^2) = \frac{1}{n}\sum_{i=1}^{n}(1 + \mu^2) = 1 + \mu^2$$

Hence t is an unbiased estimator of $1 + \mu^2$.

Example 17·2. *If T is an unbiased estimator for θ, show that T^2 is a biased estimator for θ^2.*

Solution. Since T is an unbiased estimator for θ, we have $\quad E(T) = \theta$

Also $\quad \text{Var}(T) = E(T^2) - \{E(T)\}^2 = E(T^2) - \theta^2 \Rightarrow E(T^2) = \theta^2 + \text{Var}(T),\ (\text{Var } T > 0).$

Since $E(T^2) \neq \theta^2$, T^2 is a biased estimator for θ^2.

Example 17·3. *Show that $\frac{[\Sigma x_i(\Sigma x_i - 1)]}{n(n-1)}$ is an unbiased estimate of θ^2, for the sample $x_1, x_2, \ldots, x_n$ drawn on X which takes the values 1 or 0 with respective probabilities θ and $(1-\theta)$.*

Solution. Since $x_1, x_2, \ldots, x_n$ is a random sample from Bernoulli population with parameter θ, $T = \sum_{i=1}^{n} x_i \sim B(n, \theta) \quad \Rightarrow \quad E(T) = n\theta$ and $\text{Var}(T) = n\theta(1-\theta)$

$$\therefore \quad E\left\{\frac{\Sigma x_i(\Sigma x_i - 1)}{n(n-1)}\right\} = E\left\{\frac{T(T-1)}{n(n-1)}\right\} = \frac{1}{n(n-1)}\{E(T^2) - E(T)\}$$

$$= \frac{1}{n(n-1)} \left[\text{Var}(T) + \{E(T)\}^2 - E(T) \right]$$

$$= \frac{1}{n(n-1)} \{ n\,\theta\,(1-\theta) + n^2\theta^2 - n\,\theta \} = \frac{n\,\theta^2\,(n-1)}{n(n-1)} = \theta^2$$

$\Rightarrow \quad \{\Sigma x_i\,(\Sigma x_i - 1)\} \,/\, \{n(n-1)\}$ is an unbiased estimator of θ^2.

Example 17·4. *Let X be distributed in the Poisson form with parameter θ. Show that only unbiased estimator of* $\exp\{-(k+1)\theta\}$, $k > 0$, *is* $T(X) = (-k)^X$ *so that* $T(x) > 0$ *if x is even and* $T(x) < 0$ *if x is odd.*

Solution. $E\{T(X)\} = E\{(-k)^X\}, k > 0 = \sum\limits_{x=0}^{\infty} (-k)^x \left(\dfrac{e^{-\theta}\,\theta^x}{x!} \right)$

$$= e^{-\theta} \sum_{x=0}^{\infty} \left\{ \frac{(-k\theta)^x}{x!} \right\} = e^{-\theta} \cdot e^{-k\theta} = e^{-(1+k)\theta}$$

$\Rightarrow \quad T(X) = (-k)^X$ is an unbiased estimator for $\exp\{-(1+k)\,\theta\}, k > 0.$

17·2·2. Consistency

Definition. *An estimator* $T_n = T(x_1, x_2, \ldots, x_n)$, *based on a random sample of size n, is said to be consistent estimator of* $\gamma(\theta)$, $\theta \in \Theta$, *the parameter space, if* T_n *converges to* $\gamma(\theta)$ *in probability, i.e., if* $T_n \xrightarrow{p} \gamma(\theta)$ *as* $n \to \infty$. *In other words,* T_n *is a consistent estimator of* $\gamma(\theta)$ *if for every* $\varepsilon > 0$, $\eta > 0$, *there exists a positive integer* $n \geq m\,(\varepsilon, \eta)$ *such that*

$$P\{|T_n - \gamma(\theta)| < \varepsilon\} \to 1 \text{ as } n \to \infty \quad \Rightarrow \quad P\{|T_n - \gamma(\theta)| < \varepsilon\} > 1 - \eta\,;\ \forall\, n \geq m \quad \ldots (17\cdot2)$$

where m is some very large value of n.

Remarks. 1. If $X_1, X_2, \ldots, X_n$ is a random sample from population with finite mean $EX_i = \mu < \infty$, then by Khinchine's weak law of large numbers (W.L.L.N), we have

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{p} E(X_i) = \mu\,, \text{ as } n \to \infty.$$

Hence sample mean ($\overline{X}_n$) is always a consistent estimator of the population mean (μ).

2. Obviously consistency is a property concerning the behaviour of an estimator for indefinitely large values of the sample size n, *i.e.,* as $n \to \infty$. Nothing is regarded of its behaviour for finite n.

Moreover, if there exists a consistent estimator, say, T_n of $\gamma(\theta)$, then infinitely many such estimators can be constructed, *e.g.,*

$$T_n' = \left(\frac{n-a}{n-b} \right) T_n = \left[\frac{1-(a/n)}{1-(b/n)} \right] T_n \to T_n \xrightarrow{p} \gamma(\theta), \text{ as } n \to \infty$$

and hence, for different values of a and b, T_n' is also consistent for $\gamma(\theta)$.

Invariance Property of Consistent Estimators.

Theorem 17·1. *If* T_n *is a consistent estimator of* $\gamma(\theta)$ *and* $\psi\{\gamma(\theta)\}$ *is a continuous function of* $\gamma(\theta)$, *then* $\psi(T_n)$ *is a consistent estimator of* $\psi\{\gamma(\theta)\}$.

Proof. Since T_n is a consistent estimator of $\gamma(\theta)$, $T_n \xrightarrow{p} \gamma(\theta)$ as $n \to \infty$, *i.e.,* for every $\varepsilon > 0$, $\eta > 0$, $\exists$ a positive integer $n \geq m\,(\varepsilon, \eta)$ such that

$$P\{|T_n - \gamma(\theta)| < \varepsilon\} > 1 - \eta\,, \forall\, n \geq m \quad \ldots(*)$$

Since $\psi(\cdot)$ is a continuous function, for every $\varepsilon > 0$, however small, $\exists$ a positive number ε_1 such that $| \psi(T_n) - \psi\{\gamma(\theta)\} | < \varepsilon_1$, whenever $| T_n - \gamma(\theta) | < \varepsilon$, *i.e.*,

$$| T_n - \gamma(\theta) | < \varepsilon \;\Rightarrow\; | \psi(T_n) - \psi\{\gamma(\theta)\} | < \varepsilon_1 \qquad \ldots(**)$$

For two events A and B, if $A \Rightarrow B$, then

$$A \subseteq B \;\Rightarrow\; P(A) \le P(B) \quad \text{or} \quad P(B) \ge P(A) \qquad \ldots(***)$$

From (**) and (***), we get

$$P\big[\, | \psi(T_n) - \psi\{\gamma(\theta)\} | < \varepsilon_1 \big] \ge P\big[\, | T_n - \gamma(\theta) | < \varepsilon \big]$$

$$P\big[\, | \psi(T_n) - \psi\{\gamma(\theta)\} | < \varepsilon_1 \big] \ge 1 - \eta \,;\, \forall\, n \ge m \qquad \text{[Using (*)]}$$

$\Rightarrow$

$\Rightarrow$ $\psi(T_n) \xrightarrow{p} \psi\{\gamma(\theta)\}$, as $n \to \infty$ or $\psi(T_n)$ is a consistent estimator of $\gamma(\theta)$.

Sufficient Conditions for Consistency.

Theorem 17·2. *Let $\{T_n\}$ be a sequence of estimators such that for all $\theta \in \Theta$,*

(i) $E_\theta(T_n) \to \gamma(\theta)$, $n \to \infty$ and (ii) $Var_\theta(T_n) \to 0$, as $n \to \infty$.

Then T_n is a consistent estimator of $\gamma(\theta)$.

Proof. We have to prove that T_n is a consistent estimator of $\gamma(\theta)$

i.e.,

$$T_n \xrightarrow{p} \gamma(\theta), \text{ as } n \to \infty$$

i.e.,

$$P\big[\, | T_n - \gamma(\theta) | < \varepsilon \big] > 1 - \eta \,;\, \forall\, n \ge m(\varepsilon, \eta) \qquad \ldots(17·3)$$

where ε and η are arbitrarily small positive numbers and m is some large value of n.

Applying Chebychev's inequality to the statistic T_n, we get

$$P\big[\, | T_n - E_\theta(T_n) | \le \delta \big] \ge 1 - \frac{Var_\theta(T_n)}{\delta^2} \qquad \ldots(17·4)$$

We have

$$| T_n - \gamma(\theta) | = | T_n - E(T_n) + E(T_n) - \gamma(\theta) | \le | T_n - E_\theta(T_n) | + | E_\theta(T_n) - \gamma(\theta) | \qquad \ldots(17·5)$$

Now $| T_n - E_\theta(T_n) | \le \delta \;\Rightarrow\; | T_n - \gamma(\theta) | \le \delta + | E_\theta(T_n) - \gamma(\theta) |$ ...(17·6)

Hence, on using (***) of Theorem 17·1, we get

$$P\{\, | T_n - \gamma(\theta) | \le \delta + | E_\theta(T_n) - \gamma(\theta) | \,\} \ge P\{\, | T_n - E_\theta(T_n) | \le \delta \}$$

$$\ge 1 - \frac{Var_\theta(T_n)}{\delta^2} \qquad \text{[From (17·4)]} \qquad \ldots(17·7)$$

We are given : $E_\theta(T_n) \to \gamma(\theta)\ \forall\, \theta \in \Theta$ as $n \to \infty$

Hence, for every $\delta_1 > 0$, $\exists$ a positive integer $n \ge n_0(\delta_1)$ such that

$$| E_\theta(T_n) - \gamma(\theta) | \le \delta_1,\ \forall\, n \ge n_0(\delta_1) \qquad \ldots(17·8)$$

Also $Var_\theta(T_n) \to 0$ as $n \to \infty$, (Given) $\therefore$ $\dfrac{Var_\theta(T_n)}{\delta^2} \le \eta,\ \forall\, n \ge n_0'(\eta)$, ...(17·9)

where η is arbitrarily small positive number.

Substituting from (17·8) and (17·9) in (17·7), we get

$$P\big[\, | T_n - \gamma(\theta) | \le \delta + \delta_1 \big] \ge 1 - \eta \,;\, n \ge m(\delta_1, \eta)$$

$\Rightarrow$

$$P\big[\, | T_n - \gamma(\theta) | \le \varepsilon \big] \ge 1 - \eta \,;\, n \ge m,$$

where $m = \max(n_0, n_0')$ and $\varepsilon = \delta + \delta_1 > 0$.

$$\Rightarrow \qquad T_n \xrightarrow{p} \gamma(\theta) \text{ as } n \to \infty$$

[Using

$\therefore$ T_n is a consistent estimator of $\gamma(\theta)$.

Example 17·5. *(a) Prove that in sampling from a $N(\mu, \sigma^2)$ population, the sample m is a consistent estimator of μ.*

(b) Prove that for Cauchy's distribution not sample mean but sample median consistent estimator of the population mean.

Solution. In sampling from a $N(\mu, \sigma^2)$ population, the sample mean $\bar{x}$ is also normally distributed as $N(\mu, \sigma^2/n)$, *i.e.*, $E(\bar{x}) = \mu$ and $V(\bar{x}) = \sigma^2/n$

Thus as $n \to \infty$, $\quad E(\bar{x}) = \mu$ and $V(\bar{x}) = 0$.

Hence by *Theorem 17·2*, $\bar{x}$ is a consistent estimator for μ.

(*b*) The Cauchy's population is given by the probability function :

$$dF(x) = \frac{1}{\pi} \cdot \frac{dx}{1 + (x - \mu)^2}, -\infty \le x \le \infty$$

The mean of the distribution, if we conventionally agree to assume that it exis at $x = \mu$. If $\bar{x}$, the sample mean is taken as an estimator of μ, then the samp distribution of $\bar{x}$ is given by:

$$dF(\bar{x}) = \frac{1}{\pi} \cdot \frac{d\bar{x}}{1 + (\bar{x} - \mu)^2}; -\infty < \bar{x} < \infty,$$

because in Cauchy's distribution, the distribution of $\bar{x}$ is same as the distribution o

Since in this case, the distribution of $\bar{x}$ is same as distribution of any single sam observation, it does not increase in accuracy with increasing n. In other words

$$E(\bar{x}) = \mu \quad \text{but} \quad V(\bar{x}) = V(x) \ne 0, \text{ as } n \to \infty$$

Hence by *Theorem 17·2*, $\bar{x}$ is not a consistent estimator of μ in this case.

Consideration of symmetry of (*) is enough to show that the sample median an unbiased estimate of the population mean, which of course is same as population median. Therefore $\qquad E(Md) = \mu$.

For large n, the sampling distribution of median is asymptotically normal is given by $\qquad dF \propto \exp\{-2n f_1^2 (x - \mu)^2\} dx,$

where f_1 is the median ordinate of the parent population. *i.e.*,

$$dF \propto \exp\left\{-\frac{(x - \mu)^2}{1/(2nf_1^2)}\right\}$$

But f_1= Median ordinate of (*) = Modal ordinate of (*) $\qquad$ [Because of symm

$$= [f(x)]_{x = \mu} = \frac{1}{\pi}$$

Hence, from (***), the variance of the sampling distribution of median is :

$$V(Md) = \frac{1}{4n f_1^2} = \frac{1}{4n(1/\pi)^2} = \frac{\pi^2}{4n} \to 0 \text{ as } n \to \infty$$

Hence from (**) and (****), using *Theorem 17·2*, we conclude that for Cau distribution, median is a consistent estimator for μ.

Example 17·6. *If $X_1, X_2, \ldots, X_n$ are random observations on a Bernoulli variate X taking the value 1 with probability p and the value 0 with probability $(1-p)$, show that :*

$$\frac{\Sigma x_i}{n}\left(1-\frac{\Sigma x_i}{n}\right) \text{ is a consistent estimator of } p(1-p).$$

Solution. Since $X_1, X_2, \ldots, X_n$ are i.i.d Bernoulli variates with parameter 'p',

$$T = \sum_{i=1}^{n} x_i \sim B(n, p) \Rightarrow E(T) = np \quad \text{and} \quad Var(T) = npq \qquad \ldots(i)$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{T}{n} \Rightarrow E(\bar{X}) = \frac{1}{n}E(T) = \frac{1}{n}\cdot np = p \qquad \text{[From (i)]}$$

and $$Var(\bar{X}) = Var\left(\frac{T}{n}\right) = \frac{1}{n^2}\cdot Var(T) = \frac{pq}{n} \to 0 \text{ as } n \to \infty. \qquad \text{[From (i)]}$$

Since $E(\bar{X}) \to p$ and $Var(\bar{X}) \to 0$, as $n \to \infty$; $\bar{X}$ is a consistent estimator of p. Also $\frac{\Sigma x_i}{n}\left(1-\frac{\Sigma x_i}{n}\right) = \bar{X}(1-\bar{X})$, being a polynomial in $\bar{X}$, is a continuous function of $\bar{X}$.

Since $\bar{X}$ is consistent estimator of p, by the invariance property of consistent estimators (*Theorem* 17·1), $\bar{X}(1-\bar{X})$ is a consistent estimator of $p(1-p)$.

17·2·3. Efficient Estimators. *Efficiency*. Even if we confine ourselves to unbiased estimates, there will, in general, exist more than one consistent estimator of a parameter. For example, in sampling from a normal population $N(\mu, \sigma^2)$, when σ^2 is known, sample mean $\bar{x}$ is an unbiased and consistent estimator of μ [*c.f.* Example 17·5*a*].

From symmetry it follows immediately that sample median (Md) is an unbiased estimate of μ, which is same as the population median. Also for large n,

$$V(Md) = \frac{1}{4nf_1^2} \qquad \text{[c.f. Example 17·5(b)]}$$

Here f_1 = Median ordinate of the parent distribution.

= Modal ordinate of the parent distribution.

$$= \left[\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-(x-\mu)^2/2\sigma^2\right\}\right]_{x=\mu} = \frac{1}{\sigma\sqrt{2\pi}}$$

$$\therefore \quad V(Md) = \frac{1}{4n}\cdot 2\pi\sigma^2 = \frac{\pi\sigma^2}{2n}$$

Since and

$$\left.\begin{aligned} E(Md) &= \mu \\ V(Md) &\to 0 \end{aligned}\right\}, \text{ as } n \to \infty$$

median is also an unbiased and consistent estimator of μ.

Thus, there is a necessity of some further criterion which will enable us to choose between the estimators with the common property of consistency. Such a criterion which is based on the variances of the sampling distribution of estimators is usually known as *efficiency*.

If, of the two consistent estimators T_1, T_2 of a certain parameter θ, we have

$$V(T_1) < V(T_2), \text{ for all } n \qquad \ldots(17\cdot10)$$

then T_1 is more efficient than T_2 for all samples sizes.

We have seen above :

For all n, $V(\bar{x}) = \frac{\sigma^2}{n}$ and for large n, $V(Md) = \frac{\pi\sigma^2}{2n} = 1{\cdot}57\frac{\sigma^2}{n}$

Since $V(\bar{x}) < V(Md)$, we conclude that for normal distribution, sample mean is more efficient estimator for μ than the sample median, for large samples at least.

Most Efficient Estimator. *If in a class of consistent estimators for a parameter, there exists one whose sampling variance is less than that of any such estimator, it is called the most efficient estimator. Whenever such an estimator exists, it provides a criterion for measurement of efficiency of the other estimators.*

Efficiency *(Definition) If T_1 is the most efficient estimator with variance V_1 and T_2 is any other estimator with variance V_2, then the efficiency E of T_2 is defined as :*

$$E = \frac{V_1}{V_2} \qquad \ldots(17{\cdot}11)$$

Obviously, E cannot exceed unity.

If $T, T_1, T_2, \ldots, T_n$ are all estimators of $\gamma(\theta)$ and $\text{Var}(T)$ is minimum, then the efficiency E_i of T_i, $(i = 1, 2, \ldots, n)$ is defined as :

$$E_i = \frac{\text{Var } T}{\text{Var } T_i}; i = 1, 2, \ldots, n \qquad \ldots(17{\cdot}11a)$$

Obviously $E_i \le 1; i = 1, 2, \ldots n$. For example, in the normal samples, since sample mean $\bar{x}$ is the most efficient estimator of μ [*c.f. Remark to Example 17· 31*], the efficiency E of Md for such samples, (for large n), is :

$$E = \frac{V(\bar{x})}{V(Md)} = \frac{\sigma^2/n}{\pi\sigma^2/(2n)} = \frac{2}{\pi} = 0{\cdot}637.$$

Example 17·7. *A random sample $(X_1, X_2, X_3, X_4, X_5)$ of size 5 is drawn from a normal population with unknown mean μ. Consider the following estimators to estimate μ :*

(i) $t_1 = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$, (ii) $t_2 = \frac{X_1 + X_2}{2} + X_3$, (iii) $t_3 = \frac{2X_1 + X_2 + \lambda X_3}{3}$

where λ is such that t_3 is an unbiased estimator of μ.

Find λ. Are t_1 and t_2 unbiased ? State giving reasons, the estimator which is best among t_1, t_2 and t_3.

Solution. We are given :

$$E(X_i) = \mu, \text{Var}(X_i) = \sigma^2, \text{(say)} ; \text{Cov}(X_i, X_j) = 0, (i \ne j = 1, 2, \ldots, n) \qquad \ldots(*)$$

(i) $E(t_1) = \frac{1}{5}\sum_{i=1}^{5} E(X_i) = \frac{1}{5}\sum_{i=1}^{5} \mu = \frac{1}{5} \cdot 5\mu = \mu \Rightarrow t_1$ is an unbiased estimator of μ [Using (*)]

(ii) $E(t_2) = \frac{1}{2} E(X_1 + X_2) + E(X_3) = \frac{1}{2}(\mu + \mu) + \mu = 2\mu$

$\Rightarrow$ t_2 is *not* an unbiased estimator of μ.

(iii) $E(t_3) = \mu \Rightarrow \frac{1}{3}E(2X_1 + X_2 + \lambda X_3) = \mu$

($\because t_3$ is unbiased estimator of μ)

$\therefore$ $2E(X_1) + E(X_2) + \lambda E(X_3) = 3\mu \quad \therefore \quad 2\mu + \mu + \lambda\mu = 3\mu \Rightarrow \lambda = 0$

Using (*), we get

$$V(t_1) = \frac{1}{25}\{V(X_1) + V(X_2) + V(X_3) + V(X_4) + V(X_5)\} = \frac{1}{5}\sigma^2$$

$$V(t_2) = \frac{1}{4}\{V(X_1) + V(X_2)\} + V(X_3) = \frac{1}{2}\sigma^2 + \sigma^2 = \frac{3}{2}\sigma^2$$

$$V(t_3) = \frac{1}{9}\{4V(X_1) + V(X_2)\} = \frac{1}{9}(4\sigma^2 + \sigma^2) = \frac{5}{9}\sigma^2 \qquad (\because \lambda = 0)$$

Since $V(t_1)$ is least, t_1 is the best estimator (in the sense of least variance) of μ.

Example 17·8. *X_1, X_2, and X_3 is a random sample of size 3 from a population with mean value μ and variance σ^2. T_1, T_2, T_3 are the estimators used to estimate mean value μ, where*

$$T_1 = X_1 + X_2 - X_3, \quad T_2 = 2X_1 + 3X_3 - 4X_2, \quad \text{and} \quad T_3 = \tfrac{1}{3}(\lambda X_1 + X_2 + X_3)/3.$$

(i) Are T_1 and T_2 unbiased estimators ?

(ii) Find the value of λ such that T_3 is unbiased estimator for μ.

(iii) With this value of λ is T_3 a consistent estimator ?

(iv) Which is the best estimator ?

Solution. Since X_1, X_2, X_3 is a random sample from a population with mean μ and variance σ^2, $E(X_i) = \mu$, Var $(X_i) = \sigma^2$ and Cov $(X_i, X_j) = 0, (i \neq j = 1, 2, \ldots, n)$ ...(*)

(i) We have [On using (*)],

$$E(T_1) = E(X_1) + E(X_2) - E(X_3) = \mu \Rightarrow T_1 \text{ is an unbiased estimator of } \mu$$

$$E(T_2) = 2E(X_1) + 3E(X_3) - 4E(X_2) = \mu \Rightarrow T_2 \text{ is an unbiased estimator of } \mu.$$

(ii) We are given : $E(T_3) = \mu \Rightarrow \frac{1}{3}\left\{\lambda E(X_1) + E(X_2) + E(X_3)\right\} = \mu$

$$\Rightarrow \frac{1}{3}(\lambda\mu + \mu + \mu) = \mu \Rightarrow \lambda + 2 = 3 \Rightarrow \lambda = 1.$$

(iii) With $\lambda = 1$, $T_3 = \frac{1}{3}(X_1 + X_2 + X_3) = \overline{X}$. Since sample mean is a consistent estimator of population mean μ, by Weak Law of Large Numbers, T_3 is a consistent estimator of μ.

(iv) We have [on using (*)] :

$$\text{Var}(T_1) = \text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) = 3\sigma^2$$

$$\text{Var}(T_2) = 4\,\text{Var}(X_1) + 9\,\text{Var}(X_3) + 16\,\text{Var}(X_2) = 29\,\sigma^2$$

$$\text{Var}(T_3) = \frac{1}{9}[\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3)] = \frac{1}{3}\sigma^2 \qquad (\because \lambda = 1)$$

Since Var(T_3) is minimum, T_3 is the best estimator of μ in the sense of minimum variance.

Definition., Minimum Variance Unbiased (M.V.U.) Estimators.

If a statistic $T = T(x_1, x_2, \ldots, x_n)$, based on sample of size n is such that :

(i) T is unbiased for $\gamma(\theta)$, for all $\theta \in \Theta$ and

(ii) It has the smallest variance among the class of all unbiased estimators of $\gamma(\theta)$, then T is called the minimum variance unbiased estimator (MVUE) of $\gamma(\theta)$. ...(17·12)

More precisely, T is MVUE of $\gamma(\theta)$ if ...(17·13)

$$E_\theta(T) = \gamma(\theta) \text{ for all } \theta \in \Theta \qquad ...(17{\cdot}14)$$

and

$$\text{Var}_\theta(T) \leq \text{Var}_\theta(T') \text{ for all } \theta \in \Theta$$

where T' is any other unbiased estimator of $\gamma(\theta)$.

If $T = t(x_1, x_2, \ldots, x_n)$ is an estimator of a parameter θ, based on a sample $x_1, x_2, \ldots, x_n$ of size n from the population with density $f(x, \theta)$ such that the conditional distribution of $x_1, x_2, \ldots, x_n$ given T, is independent of θ, then T is sufficient estimator for θ.

Illustration. Let $x_1, x_2, \ldots, x_n$ be a random sample from a Bernoulli population with parameter 'p', $0 < p < 1$, i.e.,

$$x_i = \begin{cases} 1, \text{ with probability } p \\ 0, \text{ with probability } q = (1-p) \end{cases}$$

Then
$$T = t(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \ldots + x_n \sim B(n, p)$$

$$\therefore \quad P(T = k) = \binom{n}{k} p^k (1-p)^{n-k};\ k = 0, 1, 2, \ldots, n$$

The conditional distribution of $(x_1, x_2, \ldots, x_n)$ given T is :

$$P(x_1 \cap x_2 \cap \ldots \cap x_n \mid T = k) = \frac{P(x_1 \cap x_2 \cap \ldots \cap x_n \cap T = k)}{P(T = k)}$$

$$= \begin{cases} \dfrac{p^k (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}} = \dfrac{1}{\binom{n}{k}} \\ 0, \text{ if } \sum\limits_{i=1}^{n} x_i \neq k \end{cases}$$

Since this does not depend on 'p', $T = \sum\limits_{i=1}^{n} x_i$, is sufficient for '$p$'.

Theorem 15·7. FACTORIZATION THEOREM (Neymann). The necessary and sufficient condition for a distribution to admit sufficient statistic is provided by the 'factorization theorem' due to Neymann.

Statement *$T = t(x)$ is sufficient for θ if and only if the joint density function L (say), of the sample values can be expressed in the form :*

$$L = g_\theta[t(x)].h(x) \qquad \ldots(17\text{·}29)$$

where (as indicated) $g_\theta[t(x)]$ depends on θ and x only through the value of $t(x)$ and $h(x)$ is independent of θ.

Remarks 1. It should be clearly understood that by 'a function independent of θ' we not only mean that it does not involve θ but also that its domain does not contain θ. For example, the function :

$$f(x) = \frac{1}{2a},\ a - \theta < x < a + \theta;\ -\infty < \theta < \infty$$

depends on θ.

2. It should be noted that the original sample $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, is always a sufficient statistic.

3. The most general form of the distributions admitting sufficient statistic is *Koopman's form* and is given by :

$$L = L(\mathbf{x}, \theta) = g(\mathbf{x}).h(\theta).\exp\{a(\theta)\psi(\mathbf{x})\} \qquad \ldots(17\text{·}30)$$

where $h(\theta)$ and $a(\theta)$ are functions of the parameter θ only and $g(\mathbf{x})$ and $\psi(\mathbf{x})$ are the functions of the sample observations only.

Equation (17·30) represents the famous *exponential family of distributions,* of which most of the common distributions like the binomial, the Poisson and the normal with unknown mean and variance, are the members.

4. Invariance Property of Sufficient Estimator. *If T is a sufficient estimator for the parameter θ and if ψ (T) is a one to one function of T, then ψ (T) is sufficient for ψ(θ).*

5. Fisher-Neyman Criterion. A statistic $t_1 = t(x_1, x_2, \ldots, x_n)$ *is sufficient estimator of parameter θ if and only if the likelihood function (joint p.d.f. of the sample) can be expressed as :*

$$L = \prod_{i=1}^{n} f(x_i, \theta) = g_1(t_1, \theta) . k(x_1, x_2, \ldots, x_n) \quad \ldots(17\cdot$$

where $g_1(t_1, \theta)$ is the p.d.f. of the statistic t_1 and $k(x_1, x_2, \ldots, x_n)$ is a function of sample observations only, independent of θ.

Note that this method requires the working out of the *p.d.f.* (*p.m.f.*) of the statistic $t_1 = t(x_1, x_2, \ldots, x_n)$, which is not always easy.

Example 17·13. *Let $x_1, x_2, \ldots, x_n$ be a random sample from a uniform population on [0, θ]. Find a sufficient estimator for θ.*

Solution. We are given : $f_\theta(x_i) = \begin{cases} \dfrac{1}{\theta}, & 0 \le x_i \le \theta \\ 0, & \text{otherwise} \end{cases}$

Let $k(a, b) = \left.\begin{matrix} 1, \text{ if } a \le b \\ 0, \text{ if } a > b \end{matrix}\right\}$, then $f_\theta(x_i) = \dfrac{k(0, x_i)\, k(x_i, \theta)}{\theta}$,

$$L = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \left[\frac{k(0, x_i)\, k(x_i, \theta)}{\theta}\right] = \frac{k(0, \min\limits_{1 \le i \le n} x_i) \cdot k(\max\limits_{1 \le i \le n} x_i, \theta)}{\theta^n} = g_\theta\{t(x)\}\, h(x)$$

where $g_\theta[t(x)] = \dfrac{k\{t(x), \theta\}}{\theta^n}$, $t(x) = \max\limits_{1 \le i \le n} x_i$ and $h(x) = k(0, \min\limits_{1 \le i \le n} x_i)$

Hence by Factorization theorem, $T = \max\limits_{1 \le i \le n} x_i$, is sufficient statistic for θ.

Aliter. We have $L = \prod\limits_{i=1}^{n} f(x_i, \theta) = \dfrac{1}{\theta^n}$; $0 < x_i < \theta$ ...

If $t = \max(x_1, x_2, \ldots, x_n) = x_{(n)}$, then *p.d.f.* of t is given by :

$$g(t, \theta) = n\{F(x_n)\}^{n-1} . f(x_{(n)}) \quad \ldots$$

We have $F(x) = P(X \le x) = \int_0^x f(x, \theta)\, dx = \int_0^x \frac{1}{\theta} . dx = \frac{x}{\theta}$

$$\therefore \quad g(t, \theta) = n\left\{\frac{x_{(n)}}{\theta}\right\}^{n-1}\left(\frac{1}{\theta}\right) = \frac{n}{\theta^n}\left[x_{(n)}\right]^{n-1} \quad \text{[From (}$$

Rewriting (*), $L = \dfrac{n\,[x_{(n)}]^{n-1}}{\theta^n} \cdot \dfrac{1}{n\,[x_{(n)}]^{n-1}} = g(t, \theta) . h(x_1, x_2, \ldots, x_n)$

Hence by Fisher – Neymann criterion, the statistic $t = x_{(n)}$, is sufficient estimator for θ.

Example 17·14. *Let $x_1, x_2, \ldots, x_n$ be a random sample from $N(\mu, \sigma^2)$ population. Find sufficient estimators for μ and σ^2.*

Solution. Let us write $\theta = (\mu, \sigma^2)$; $-\infty < \mu < \infty$, $0 < \sigma^2 < \infty$.

Then
$$L = \prod_{i=1}^{n} f_\theta(x_i) = \left\{\frac{1}{\sigma\sqrt{2\pi}}\right\}^n . \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 - 2\mu\sum x_i + n\mu^2\right)\right\}$$

$$= g_\theta[t(x)] . h(x)$$

$$g_\theta[t(x)] = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\left\{t_2(x) - 2\mu t_1(x) + n\mu^2\right\}\right],$$

where $t(x) = \{t_1(x), t_2(x)\} = (\Sigma x_i, \Sigma x_i^2)$ and $h(x) = 1$

Thus $t_1(x) = \Sigma x_i$ is sufficient for μ and $t_2(x) = \Sigma x_i^2$, is sufficient for σ^2.

Example 17·15. *Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with p.d.f. :*

$$f(x, \theta) = e^{-(x-\theta)}, \theta < x < \infty ; -\infty < \theta < \infty$$

Obtain sufficient statistic for θ.

Solution. Here

$$L = \sum_{i=1}^{n} f(x_i, \theta) = \sum_{i=1}^{n} \{e^{-(x_i-\theta)}\} = \exp\left(-\sum_{i=1}^{n} x_i\right) \times \exp(n\theta) \qquad \ldots(*)$$

Let $Y_1, Y_2, \ldots, Y_n$ denote the orderstatistics of the random sample such that $Y_1 < Y_2 < \ldots < Y_n$. The *p.d.f.* of the smallest observation Y_1 is given by :

$$g_1(y_1, \theta) = n[1 - F(y_1)]^{n-1} f(y_1, \theta),$$

where $F(\cdot)$ is the distribution function corresponding to *p.d.f.* $f(\cdot)$.

Now $$F(x) = \int_\theta^x e^{-(x-\theta)}\,dx = \left|\frac{e^{-(x-\theta)}}{-1}\right|_\theta^x = 1 - e^{-(x-\theta)}$$

$$\therefore \quad g_1(y_1, \theta) = n\left[e^{-(y_1-\theta)}\right]^{n-1} . e^{-(y_1-\theta)} = \begin{cases} n\, e^{-n(y_1-\theta)}, & \theta < y_1 < \infty \\ 0, & \text{otherwise} \end{cases}$$

Thus the likelihood function (*) of $X_1, X_2, \ldots, X_n$ may be expressed as

$$L = e^{n\theta} \exp\left(-\sum_{i=1}^{n} x_i\right) = n \exp\{-n(y_1 - \theta)\}\left\{\frac{\exp(-\sum_{i=1}^{n} x_i)}{n \exp(-ny_1)}\right\}$$

$$= g_1(\min x_i, \theta)\left\{\frac{\exp(-\sum_{i=1}^{n} x_i)}{n \exp(-n \min x_i)}\right\}$$

Hence by Fisher-Neymann criterion, the first order statistic $Y_1 = \min(X_1, X_2, \ldots, X_n)$ is a sufficient statistic for θ.

Example 17·16. *Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with p.d.f :*

$$f(x, \theta) = \theta x^{\theta - 1} ; 0 < x < 1, \ \theta > 0.$$

Show that

$$t_1 = \prod_{i=1}^{n} X_i, \text{ is sufficient for } \theta.$$

Solution.
$$L(x, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \theta^n \prod_{i=1}^{n} (x_i^{\theta-1})$$

$$= \theta^n\left(\prod_{i=1}^{n} x_i\right)^\theta . \frac{1}{\left(\prod_{i=1}^{n} x_i\right)} = g(t_1, \theta) . h(x_1, x_2, \ldots, x_n), \text{ (say)}.$$

Hence by Factorisation Theorem, $t_1 = \prod_{i=1}^{n} (X_i)$, is sufficient estimator for θ.

Example 17·17. *Let $X_1, X_2, \ldots, X_n$ be a random sample from Cauchy population:*

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

Examine if there exists a sufficient statistic for θ.

Solution. $L(\mathbf{x}, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \frac{1}{\pi^n} \cdot \prod_{i=1}^{n} \left\{ \frac{1}{1 + (x_i - \theta)^2} \right\} \neq g(t_1, \theta) \cdot h(x_1, x_2, \ldots, x_n)$

Hence by Factorisation Theorem, there is no single statistic, which alone, is sufficient estimator of θ.

However, $\quad L(\mathbf{x}, \theta) = k_1(X_1, X_2, \ldots, X_n, \theta) \cdot k_2(X_1, X_2, \ldots, X_n)$

$\Rightarrow \quad$ The whole set $(X_1, X_2, \ldots, X_n)$ is jointly sufficient for θ.

17·6. METHODS OF ESTIMATION

So far we have been discussing the requisites of a good estimator. Now we shal briefly outline some of the important methods for obtaining such estimato Commonly used methods are :

(i) *Method of Maximum Likelihood Estimation.*
(ii) *Method of Minimum Variance.*
(iii) *Method of Moments.*
(iv) *Method of Least Squares.*
(v) *Method of Minimum Chi-square.*
(vi) *Method of Inverse Probability.*

In the following sections, we shall discuss briefly the first four methods only.

17·6·1. Method of Maximum Likelihood Estimation. From theoretical point o view, the most general method of estimation known is the method of *Maximu Likelihood Estimators* (M.L.E.) which was initially formulated by C.F. Gauss but as general method of estimation was first introduced by Prof. R.A. Fisher and later o develoed by him in a series of papers. Before introducing the method we will fir define *Likelihood Function.*

Likelihood Function. *Definition.* Let $x_1, x_2, \ldots, x_n$ be a random sample of size from a population with density function $f(x, \theta)$. Then the likelihood function of th sample values $x_1, x_2, \ldots, x_n$, usually denoted by $L = L(\theta)$ is their joint density functio given by :

$$L = f(x_1, \theta)\, f(x_2, \theta) \ldots f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta). \qquad \ldots(17\cdot5)$$

gives the relative likelihood that the random variables assume a particular set of values $x_1, x_2, \ldots, x_n$. For a given sample $x_1, x_2, \ldots, x_n$, L becomes a function of the variable θ, the parameter.

The principle of maximum likelihood consists in finding an estimator for the unknown parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$, say, which maximises the likelihood function for variations in parameter, *i.e.*, we wish to find $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ so that

$$L(\hat{\theta}) > L(\theta) \quad \forall\, \theta \in \Theta, \text{ i.e., } L(\hat{\theta}) = \text{Sup } L(\theta) \ \forall\, \theta \in \Theta.$$

Thus if there exists a function $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$ of the sample values which maximises L for variations in θ, then $\hat{\theta}$ is to be taken as an estimator of θ. $\hat{\theta}$ is usually called *Maximum Likelihood Estimator* (M.L.E.). Thus $\hat{\theta}$ is the solution, if any, of

$$\frac{\partial L}{\partial \theta} = 0 \quad \text{and} \quad \frac{\partial^2 L}{\partial \theta^2} < 0 \qquad \ldots(17{\cdot}54)$$

Since $L > 0$, and $\log L$ is a non-decreasing function of L; L and $\log L$ attain their extreme values (maxima or minima) at the same value of $\hat{\theta}$. The first of the two equations in (17·54) can be rewritten as :

$$\frac{1}{L} \cdot \frac{\partial L}{\partial \theta} = 0 \quad \Rightarrow \quad \frac{\partial \log L}{\partial \theta} = 0, \qquad \ldots(17{\cdot}54a)$$

a form which is much more convenient from practical point of view.

If θ is vector valued parameter, then $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$, is given by the solution of simultaneous equations :

$$\frac{\partial}{\partial \theta_i} \log L = \frac{\partial}{\partial \theta_i} \log L(\theta_1, \theta_2, \ldots, \theta_k) = 0 \ ; \ i = 1, 2, \ldots, k \qquad \ldots(17{\cdot}54b)$$

The above equations (17·54 *a*) and (17·54 *b*) are usually referred to as the *Likelihood Equations* for estimating the parameters.

Remark. For the solution $\hat{\theta}$ of the likelihood equations, we have to see that the second derivative of L *w.r. to* θ is negative. If θ is vector valued, then for L to be maximum, the matrix of derivatives $\left(\frac{\partial^2}{\partial \theta_i \, \partial \theta_j} \log L \right)_{\theta = \hat{\theta}}$ should be negative definite.

Properties of Maximum Likelihood Estimators. We make the following assumptions, known as the *Regularity Conditions* :

(*i*) The first and second order derivatives, *viz.*, $\frac{\partial \log L}{\partial \theta}$ and $\frac{\partial^2 \log L}{\partial \theta^2}$ exist and are continuous functions of θ in a range R (including the true value θ_0 of the parameter) for almost all x. For every θ in R, $\left| \frac{\partial}{\partial \theta} \log L \right| < F_1(x)$ and $\left| \frac{\partial^2}{\partial \theta^2} \log L \right| < F_2(x)$ where $F_1(x)$ and $F_2(x)$ are integrable functions over $(-\infty, \infty)$.

(*ii*) The third order derivative $\frac{\partial^3}{\partial \theta^3} \log L$ exists such that $\left| \frac{\partial^3}{\partial \theta^3} . \log L \right| < M(x)$, where $E[M(x)] < K$, a positive quantity.

(*iii*) For every θ in R,

$$E\left(-\frac{\partial^2}{\partial\theta^2}\log L\right) = \int_{-\infty}^{\infty}\left(-\frac{\partial^2}{\partial\theta^2}\log L\right) L\, dx = I(\theta), \text{ is finite and non-zero.}$$

(*iv*) The range of integration is independent of θ. But if the range of integration depends on θ, then $f(x, \theta)$ vanishes at the extremes depending on θ.

This assumption is to make the differentiation under the integral sign valid.

Under the above assumptions M.L.E. possesses a number of important properties, which will be stated in the form of theorems.

Theorem 17·11. (*Cramer-Rao Theorem*). *"With probability approaching unity as $n \to \infty$, the likelihood equation $\frac{\partial}{\partial\theta}\log L = 0$, has a solution which converges in probability to the true value θ_0". In other words M.L.E.'s are consistent.*

Remark. MLE's are always consistent estimators but need not be unbiased. For example in sampling from $N(\mu, \sigma^2)$ population, [c.f. Example 17·31],

MLE(μ) = $\bar{x}$ (sample mean), which is both unbiased and consistent estimator of μ.

MLE(σ^2) = s^2 (sample variance), which is consistent but not unbiased estimator of σ^2.

Theorem 17·12. (*Hazoor Bazar's Theorem*). *Any consistent solution of the likelihood equation provides a maximum of the likelihood with probability tending to unity as the sample size (n) tends to infinity.*

Theorem 17·13. (ASYMPTOTIC NORMALITY OF MLE'S). *A consistent solution of the likelihood equation is asymptotically normally distributed about the true value θ_0.* Thus, $\hat{\theta}$ is asymptotically $N\left(\theta_0, \frac{1}{I(\theta_0)}\right)$, as $n \to \infty$.

Remark. Variance of M.L.E. is given by : $V(\hat{\theta}) = \dfrac{1}{I(\theta)} = \dfrac{1}{\left\{E\left(-\frac{\partial^2}{\partial\theta^2}\log L\right)\right\}}$ ...(17·55)

Theorem 17·14. *If M.L.E. exists, it is the most efficient in the class of such estimators.*

Theorem 17·15. *If a sufficient estimator exists, it is a function of the Maximum Likelihood Estimator.*

Proof. If $t = t(x_1, x_2, \ldots, x_n)$ is a sufficient estimator of θ, then Likelihood Function can be written as (*c.f.* Theorem 17·7) : $L = g(t, \theta)\, h(x_1, x_2, x_3, \ldots, x_n \mid t)$, where $g(t, \theta)$ is the density function of t and θ and $h(x_1, x_2, \ldots, x_n \mid t)$ is the density function of the sample, given t, and is independent of θ.

$$\therefore \quad \log L = \log g(t, \theta) + \log h(x_1, x_2, \ldots, x_n \mid t)$$

Differentiating *w.r. to* θ, we get : $\dfrac{\partial \log L}{\partial\theta} = \dfrac{\partial}{\partial\theta}\log g(t, \theta) = \psi(t, \theta)$, (say), ...(17·56)

which is a function of t and θ only.

M.L.E. of θ is given by $\dfrac{\partial \log L}{\partial\theta} = 0 \quad \Rightarrow \quad \psi(t, \theta) = 0$

$$\therefore \quad \hat{\theta} = \eta(t) = \text{Some function of sufficient statistic}$$

$$\Rightarrow \quad t = \xi(\hat{\theta}) = \text{Some function of M.L.E.}$$

Hence the theorem.

Remark. This theorem is quite helpful in finding if a sufficient estimator exists or not. If $\frac{\partial}{\partial\theta}\log L$ can be expressed in the form (17·56), *i.e.*, as a function of a statistic and parameter alone, then the statistic is regarded as a sufficient estimator of the parameter. If $\frac{\partial}{\partial\theta}\log L$ cannot be expressed in the form (17·56), no sufficient estimator exists in that case.

Theorem 17·16. *If for a given population with p.d.f. $f(x, \theta)$, an MVB estimator T exits for θ, then likelihood equation will have a solution equal to the estimator T.*

Proof. Since T is an MVB estimator of θ, we have [*c.f.* (17·40)],

$$\frac{\partial}{\partial\theta}\log L = \frac{T-\theta}{\lambda(\theta)} = (T-\theta)\,A(\theta)$$

MLE for θ is the solution of the likelihood equation :

$$\frac{\partial}{\partial\theta}\log L = 0 \quad\Rightarrow\quad \hat{\theta} = T, \text{ as required.}$$

Theorem 17·17. (INVARIANCE PROPERTY OF MLE). *If T is the MLE of θ and $\psi(\theta)$ is one to one function of θ, then $\psi(T)$ is the MLE of $\psi(\theta)$.*

Example 17·31. *In random sampling from normal population $N(\mu, \sigma^2)$, find the maximum likelihood estimators for*

(*i*) *μ when σ^2 is known,* (*ii*) *σ^2 when μ is known, and*

(*iii*) *the simultaneous estimation of μ and σ^2.*

Solution. $X \sim N(\mu, \sigma^2)$, then

$$L = \prod_{i=1}^{n}\left[\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(x_i-\mu)^2\right\}\right] = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\sum_{i=1}^{n}(x_i-\mu)^2/2\sigma^2\right\}$$

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

Case (*i*). When σ^2 is known, the likelihood equation for estimating μ is :

$$\frac{\partial}{\partial\mu}\log L = 0 \quad\Rightarrow\quad -\frac{1}{2\sigma^2}\sum_{i=1}^{n}2(x_i-\mu)(-1) = 0$$

$$\Rightarrow\quad \sum_{i=1}^{n}(x_i-\mu) = 0 \quad\Rightarrow\quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}x_i = \bar{x} \qquad \ldots(*)$$

Hence M.L.E. for μ is the sample mean $\bar{x}$.

Case (*ii*). When μ is known, the likelihood equation for estimating σ^2 is :

$$\frac{\partial}{\partial\sigma^2}\log L = 0 \quad\Rightarrow\quad -\frac{n}{2}\times\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\Rightarrow\quad n - \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \quad\Rightarrow\quad \boxed{\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2} \qquad \ldots(**)$$

Case (*iii*). The likelihood equations for simultaneous estimation of μ and σ^2 are :

$$\frac{\partial}{\partial\mu}\log L = 0 \text{ and } \frac{\partial}{\partial\sigma^2}\log L = 0, \text{ thus giving} \quad \hat{\mu} = \bar{x} \qquad \text{[From (*)]}$$

and

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = s^2, \text{ the sample variance.}$$

Important Note. It may be pointed out here that though

$$E(\hat{\mu}) = E(\bar{x}) = \mu,\ E(\hat{\sigma^2}) = E(s^2) \neq \sigma^2$$

Hence the maximum likelihood estimators (M.L.Es.) need not necessarily be unbiased. Another illustration is given in *Example* 17·32.

Remark. Since M.L.E. is the most efficient, we conclude that in sampling from a normal population, the sample mean $\bar{x}$ is the most efficient estimator of the population mean μ.

Example 17·32. *Prove that the maximum likelihood estimate of the parameter α of a population having density function : $\frac{2}{\alpha^2}(\alpha - x),\ 0 < x < \alpha$, for a sample of unit size is $2x$, x being the sample value. Show also that the estimate is biased.*

Solution. For a random sample of unit size ($n = 1$), the likelihood function is :

$$L(\alpha) = f(x, \alpha) = \frac{2}{\alpha^2}(\alpha - x)\ ;\ 0 < x < \alpha$$

Likelihood equation gives : $\frac{d}{d\alpha} \log L = \frac{d}{d\alpha}\{\log 2 - 2 \log \alpha + \log(\alpha - x)\} = 0$

$$\Rightarrow \quad -\frac{2}{\alpha} + \frac{1}{\alpha - x} = 0 \quad \Rightarrow \quad 2(\alpha - x) - \alpha = 0 \quad \Rightarrow \quad \alpha = 2x$$

Hence MLE of α is given by : $\hat{\alpha} = 2x.$

$$E(\hat{\alpha}) = E(2X) = 2\int_0^{\alpha} x f(x, \alpha)\, dx = \frac{4}{\alpha^2}\int_0^{\alpha} x(\alpha - x)\, dx = \frac{4}{\alpha^2}\left|\frac{\alpha x^2}{2} - \frac{x^3}{3}\right|_0^{\alpha} = \frac{2}{3}\alpha$$

Since $E(\hat{\alpha}) \neq \alpha$, $\hat{\alpha} = 2x$ is not an unbiased estimate of α.

Example 17·33. (*a*) *Find the maximum likelihood estimate for the parameter λ of a Poisson distribution on the basis of a sample of size n. Also find its variance.*

(*b*) *Show that the sample mean $\bar{x}$, is sufficient for estimating the parameter λ of the Poisson distribution.*

Solution. The probability function of the Poisson distribution with parameter λ is given by :

$$P(X = x) = f(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}\ ;\ x = 0, 1, 2, \ldots$$

Likelihood function of random sample $x_1, x_2, \ldots, x_n$ of n observations from this population is :

$$L = \prod_{i=1}^{n} f(x_i, \lambda) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{x_1!\, x_2! \ldots x_n!}$$

$$\therefore \quad \log L = -n\lambda + n\bar{x}\log\lambda - \sum_{i=1}^{n}\log(x_i!)$$

The likelihood equation for estimating λ is :

$$\frac{\partial}{\partial\lambda}\log L = 0 \quad \Rightarrow \quad -n + \frac{n\bar{x}}{\lambda} = 0 \quad \Rightarrow \quad \lambda = \bar{x}$$

Thus the M.L.E. for λ is the sample mean $\bar{x}$. The variance of estimate is given by :

$$\frac{1}{V(\hat{\lambda})} = E\left\{-\frac{\partial^2}{\partial\lambda^2}(\log L)\right\} \qquad [c.f. (17·55)]$$

$$= E\left\{-\frac{\partial}{\partial\lambda}\left(-n+\frac{n\bar{x}}{\lambda}\right)\right\} = E\left\{-\left(-\frac{n\bar{x}}{\lambda^2}\right)\right\} = \frac{n}{\lambda^2}E(\bar{x}) = \frac{n}{\lambda} \qquad [\because E(\bar{x}) = \lambda]$$

$$V(\hat{\lambda}) = \lambda/n$$

(b) For the Poisson distribution with parameter λ, we have

$$\therefore \frac{\partial}{\partial\lambda}\log L = -n+\frac{n\bar{x}}{\lambda} = n\left(\frac{\bar{x}}{\lambda}-1\right) = \psi(\bar{x}, \lambda), \text{ a function of } \bar{x} \text{ and } \lambda \text{ only.}$$

Hence (*c.f.* Remark to *Theorem 17·15*), $\bar{x}$ is sufficient for estimating λ.

Example 17·34. *Let $x_1, x_2, \ldots, x_n$ denote random sample of size n from a uniform population with p.d.f. :*

$$f(x, \theta) = 1\,;\, \theta-\tfrac{1}{2} \le x \le \theta+\tfrac{1}{2}, -\infty < \theta < \infty$$

Obtain M.L.E. for θ.

Solution. Here

$$L = L(\theta; x_1, x_2, \ldots, x_n) = \begin{cases} 1, \theta-\frac{1}{2} \le x_i \le \theta+\frac{1}{2} \\ 0, \text{elsewhere} \end{cases}$$

If $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ is the ordered sample, then

$$\theta-\tfrac{1}{2} \le x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)} \le \theta+\tfrac{1}{2}$$

Thus L attains the maximum if

$$\theta-\tfrac{1}{2} \le x_{(1)} \text{ and } x_{(n)} \le \theta+\tfrac{1}{2} \quad\Rightarrow\quad \theta \le x_{(1)}+\tfrac{1}{2} \text{ and } x_{(n)}-\tfrac{1}{2} \le \theta$$

Hence every *statistic* $t = t(x_1, x_2, \ldots, x_n)$ such that

$$x_{(n)}-\tfrac{1}{2} \le t\,(x_1, x_2, \ldots, x_n) \le x_{(1)}+\tfrac{1}{2}, \text{ provides an M.L.E. for } \theta.$$

Remark. This example illustrates that M.L.E. for a parameter need not be unique.

Example 17·35. *Find the M.L.E. of the parameters α and λ, (λ being large), of the distribution :*

$$f(x; \alpha, \lambda) = \frac{1}{\Gamma(\lambda)}\left(\frac{\lambda}{\alpha}\right)^{\lambda} e^{-\lambda x/\alpha} x^{\lambda-1}\,;\, 0 \le x < \infty, \lambda > 0$$

You may use that for large values of λ,

$$\psi(\lambda) = \frac{\partial}{\partial\lambda}\log\Gamma(\lambda) = \log\lambda - \frac{1}{2\lambda} \quad\text{and}\quad \psi'(\lambda) = \frac{1}{\lambda}+\frac{1}{2\lambda^2}. \qquad \ldots (*)$$

Solution. Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from the given population.

Then

$$L = \prod_{i=1}^{n} f(x_i; \alpha, \lambda) = \left(\frac{1}{\Gamma(\lambda)}\right)^n \cdot \left(\frac{\lambda}{\alpha}\right)^{n\lambda} \cdot \exp\left(-\frac{\lambda}{\alpha}\sum_{i=1}^{n} x_i\right) \cdot \prod_{i=1}^{n} (x_i^{\lambda-1})$$

$$\therefore \log L = -n\log\Gamma(\lambda) + n\lambda(\log\lambda - \log\alpha) - \frac{\lambda}{\alpha}\sum_{i=1}^{n} x_i + (\lambda-1)\sum_{i=1}^{n}\log x_i$$

If G is the geometric mean of $x_1, x_2, \ldots, x_n$, then

$$\log G = \frac{1}{n}\sum_{i=1}^{n}\log x_i \quad\Rightarrow\quad n\log G = \sum_{i=1}^{n}\log x_i$$

$$\therefore \log L = -n\log\Gamma(\lambda) + n\lambda(\log\lambda - \log\alpha) - \frac{\lambda}{\alpha}n\bar{x} + (\lambda-1).\,n\log G,$$

where G is independent of λ and α.

Example 17·37. *(a) Let $x_1, x_2, \ldots, x_n$ be a random sample from the uniform distribution with p.d.f. :*

$$f(x, \theta) = \begin{cases} \dfrac{1}{\theta}, & 0 < x < \infty, \theta > 0 \\ 0, & \text{elsewhere} \end{cases}$$

Obtain the maximum likelihood estimator for θ.

(b) Obtain the M.L.Es. for α and β for the rectangular population :

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{1}{\beta - \alpha}, & \alpha < x < \beta \\ 0, & \text{elsewhere} \end{cases}$$

Solution. (*a*) Here $\quad L = \prod_{i=1}^{n} f(x_i, \theta) = \dfrac{1}{\theta} \cdot \dfrac{1}{\theta} \cdots \dfrac{1}{\theta} = \left(\dfrac{1}{\theta}\right)^n$ ...(*)

Likelihood equation, *viz.*, $\dfrac{\partial}{\partial \theta} \log L = 0$, gives

$$\frac{\partial}{\partial \theta}(-n \log \theta) = 0 \quad \Rightarrow \quad \frac{-n}{\theta} = 0 \quad \text{or} \quad \hat{\theta} = \infty, \text{ obviously an absurd result.}$$

In this case we locate M.L.E. as follows : We have to choose θ so that L in (*) is maximum. Now L is maximum if θ is minimum.

Let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ be the *ordered* random sample of n independent observations from the given population so that $0 \le x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)} \le \theta \quad \Rightarrow \quad \theta \ge x_{(n)}$

Since the minimum value of θ consistent with the sample is $x_{(n)}$, the largest sample observation, $\hat{\theta} = x_{(n)}$.

$\therefore$ M.L.E. for $\theta = x_{(n)}$ = The largest sample observation.

(*b*) Here $\quad L = \left(\dfrac{1}{\beta - \alpha}\right)^n \quad \Rightarrow \quad \log L = -n \log (\beta - \alpha)$ ...(**)

The likelihood equations for α and β give

$$\left.\begin{aligned} \frac{\partial}{\partial \alpha} \log L = 0 &= \frac{n}{\beta - \alpha} \\ \text{and} \quad \frac{\partial}{\partial \beta} \log L = 0 &= \frac{-n}{\beta - \alpha} \end{aligned}\right\}$$

Each of these equations gives $\beta - \alpha = \infty$, an obviously negative result. So, we find M.L.Es for α and β by some other means.

Now L in (**) is maximum if $(\beta - \alpha)$ is minimum, *i.e.*, if β takes the minimum possible value and α takes the maximum possible value.

As in part (*a*), if $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ is an *ordered* random sample from this population, then $\alpha \le x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)} \le \beta$. Thus $\beta \ge x_{(n)}$ and $\alpha \le x_{(1)}$. Hence the minimum possible value of β consistent with the sample is $x_{(n)}$ and the maximum possible value of α consistent with the sample is $x_{(1)}$. Hence L is maximum if $\beta = x_{(n)}$ and $\alpha = x_{(1)}$.