Statistical Inference-I (Theory of Estimation)

15·1. Introduction. 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 $\{f(x, \theta), \theta \in \Theta\}$. For example 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

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

15·2. Characteristics of Estimators. The following are some of the criteria that should be satisfied by a good estimator.

(*i*) *Consistency*

(*ii*) *Unbiasedness*

(*iii*) *Efficiency* and

(*iv*) *Sufficiency*

We shall now, briefly, explain these terms one by one.

23.4 Unbiased Estimator. A statistic t **is called an unbiased** estimator of a parameter θ if the **expectation of t is θ, *i.e.* if** $E(t) = \theta$.

If $E(t) \neq \theta$ then t is called *biased* ***estimator* of θ.**

Bias. Let t be a estimator of the parameter θ. **Then $E(t) - \theta$ is** called bias of t. Obviously the bias of an **unbiased estimator is 0.** In the following Theorems we shall **find some examples of** unbiased and biased estimator.

Unbiasedness is a property associated with finite n. A statistic $T_n = T(x_1, x_2, \ldots, x_n)$, is said to be an unbiased estimator of $\gamma(\theta)$ if

$$E(T_n) = \gamma(\theta), \text{ for all } \theta \in \Theta \qquad \ldots(15\cdot3)$$

We have seen (c.f. § 12·12) that in sampling from a population with mean μ and variance σ^2,

$$E(\bar{x}) = \mu \text{ and } E(s^2) \neq \sigma^2 \text{ but } E(S^2) = \sigma^2.$$

Hence there is a reason to prefer

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

Theorem 1. The sample mean $\bar{x}$ is an unbiased estimator of the population mean. [*W.B.U.T. 2012, 2006*]

Proof. Let $\{x_1, x_2, \ldots x_n\}$ be a random sample drawn from a population of mean μ. The sample mean, $\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$. Since $x_1, x_2, \ldots x_n$ vary over the population and since the sample is random, $\bar{x}_1$ = population mean $= \mu$. In this way $\bar{x}_2 = \bar{x}_3 = \ldots = \bar{x}_n = \mu$.

Now, $E(\bar{x}) = E\left(\frac{x_1 + x_2 + \ldots + x_n}{n}\right) = \frac{1}{n}\{E(x_1) + E(x_2) + \ldots + E(x_n)\}$

$$= \frac{1}{n}(\bar{x}_1 + \bar{x}_2 + \ldots + \bar{x}_n) = \frac{(\mu + \mu + \ldots + \mu)}{n} = \mu.$$

This completes the proof.

15·3. Consistency. 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
$$T_n \xrightarrow{p} \gamma(\theta) \text{ as } n \to \infty \qquad \ldots(15\cdot1)$$

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

$$P\left[|T_n - \gamma(\theta)| < \varepsilon\right] \to 1 \text{ as } n \to \infty \qquad \ldots(15\cdot2)$$

$$\Rightarrow \qquad P\left[|T_n - \gamma(\theta)| < \varepsilon\right] > 1 - \eta \;;\; \forall\, n \geq m \qquad \ldots(15\cdot2a)$$

where m is some very large value of n.

2.3.5 Consistent Estimator.

Let θ be a parameter and t be a statistic. t depends on the size of the sample and so we denote it by t_n when t is computed for samples of size n. If t_n is expected to come closer to θ as n increases t is called a consistent estimator of θ. More precisely t is called consistent estimator of θ if, for arbitrary $\varepsilon > 0$, $\lim_{n\to\infty} P\{|t_n - \theta| > \varepsilon\} = 0$ [or, equivalently $\lim_{n\to\infty} P\{|t_n - \theta| < \varepsilon\} = 1$]

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 $\gamma(\theta)$.

15·4·1. Invariance Property of Consistent Estimators.

Theorem 15·1. *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 $\psi(\gamma(\theta))$.*

15·4·2. Sufficient Conditions for Consistency.

Theorem 15·2. 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 $\gamma(\theta)$.

Example 15·1. *$x_1, x_2, \ldots x_n$ 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

$$E(x_i) = \mu,\ V(x_i) = 1\ \forall\ i = 1, 2, \ldots, n$$

Now $$E(x_i^2) = V(x_i) + \{E(x_i)\}^2 = 1 + \mu^2$$

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

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

Example 15·2. *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

$$E(T) = \theta$$

Also $\quad \text{Var}(T) = E(T^2) - [E(T)]^2 = E(T^2) - \theta^2$

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

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

Example 15·3. *Show that $\dfrac{[\Sigma x_i (\Sigma x_i - 1)]}{n(n-1)}$ 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,

$$T = \sum_{i=1}^{n} x_i \sim B(n, \theta)$$

$$\Rightarrow \quad E(T) = n\theta \quad \text{and} \quad \text{Var}(T) = n\,\theta\,(1 - \theta)$$

$$E\left[\frac{\Sigma x_i (\Sigma x_i - 1)}{n(n-1)}\right] = E\left[\frac{T(T-1)}{n(n-1)}\right]$$

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

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

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

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

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

Example 15·4. *Let X be distributed in the Poissòn form with parameter θ. Show that the only unbiased estimator of $\exp\left[-(k+1)\theta\right]$, $k > 0$, is $T(X) = (-k)^X$ so that*

$$T(x) > 0 \text{ if } x \text{ is even}$$

and

$$T(x) < 0 \text{ if } x \text{ is odd.}$$

[**Delhi Univ. B.Sc. (Stat. Hons.), 1993, 1988**]

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

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

$\Rightarrow$ $T(X) = (-k)^X$ is an unbiased estimator for $\exp\left[-(1+k)\theta\right], k > 0$.

Example 15·5. (*a*) *Prove that in sampling from a $N(\mu, \sigma^2)$ population, the sample mean is a consistent estimator of μ.*

Solution. In sampling from a $N(\mu, \sigma^2)$ population, the sample mean $\overline{x}$ is also normally distributed as $N(\mu, \sigma^2/n)$.

$$\Rightarrow \qquad E(\overline{x}) = \mu \quad \text{and} \quad V(\overline{x}) = \sigma^2/n$$

Thus as $n \to \infty$,

$$E(\overline{x}) = \mu \quad \text{and} \quad V(\overline{x}) = 0$$

Hence by Theorem 15·2, $\overline{x}$ is a consistent estimator for μ.

Example 15·6. *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 :*

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

[*Delhi Univ. B.Sc. (Stat. Hons.), 1988*]

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

$$T = \sum_{i=1}^{n} x_i \sim B(n, p)$$

$$\Rightarrow \qquad E(T) = np \quad \text{and} \quad \text{Var}(T) = npq$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{T}{n}$$

$$\therefore \qquad E(\bar{X}) = \frac{1}{n}E(T) = \frac{1}{n}\cdot np = p$$

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

Since $E(\bar{X}) \to p$ and $\text{Var}(\bar{X}) \to 0$, as $n \to \infty$; $\bar{X}$ is a consistent estimator of p.

Also $\dfrac{\Sigma x_i}{n}\left(1-\dfrac{\Sigma x_i}{n}\right) = \bar{X}\,(1-\bar{X})$, 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 15·1), $\bar{X}\,(1-\bar{X})$ is a consistent estimator of $p\,(1-p)$.

15·5. 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 15·5*a*].

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

$$V(Md) = \frac{1}{4n f_1^2} \qquad [c.f. \text{ Example } 15\cdot5(b)]$$

Here $\quad f_1$ = Median ordinate of the parent distribution.

= Modal ordinate of the parent distribution.

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

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

Since and $\left.\begin{array}{l} E(Md) = \mu \\ V(Md) \to 0 \end{array}\right\}$, 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

$$V(T_1) < V(T_2), \text{ for all } n \qquad \ldots(15{\cdot}11)$$

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

We have seen above :

For all n, $\quad V(\bar{x}) = \dfrac{\sigma^2}{n}$

and for large n, $\quad V(Md) = \dfrac{\pi\sigma^2}{2n} = 1{\cdot}57\,\dfrac{\sigma^2}{n}$

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.

15·5·1. 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 *(Def.) 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 :*

$$E = \frac{V_1}{V_2} \qquad \ldots(15{\cdot}12)$$

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 :

$$E_i = \frac{\text{Var } T}{\text{Var } T_i}; \; i = 1, 2, \ldots, n \qquad \ldots(15{\cdot}12a)$$

Example 15·7. *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) $t_1 = \dfrac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$

(ii) $t_2 = \dfrac{X_1 + X_2}{2} + X_3$, (iii) $t_3 = \dfrac{2X_1 + X_2 + \lambda X_3}{3}$

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

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

(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.

(ii) $$\begin{aligned} E(t_2) &= \tfrac{1}{2} E(X_1 + X_2) + E(X_3) \\ &= \tfrac{1}{2}(\mu + \mu) + \mu \qquad \text{[Using (*)]} \\ &= 2\mu \end{aligned}$$

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

(iii) $$\begin{aligned} E(t_3) &= \mu \\ \Rightarrow \quad \tfrac{1}{3} E(2X_1 + X_2 + \lambda X_3) &= \mu \\ \Rightarrow \quad 2E(X_1) + E(X_2) + \lambda E(X_3) &= 3\mu \\ \Rightarrow \quad 2\mu + \mu + \lambda\mu &= 3\mu \\ \Rightarrow \quad \lambda\mu = 0 \quad &\Rightarrow \quad \lambda = 0 \end{aligned}$$

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

Example 15·8. *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*

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

(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,

$$E(X_i) = \mu,\ \text{Var}(X_i) = \sigma^2 \text{ and } \text{Cov}(X_i, X_j) = 0,\ (i \neq j = 1, 2, \ldots, n) \quad \ldots(*)$$

(*i*) $E(T_1) = E(X_1) + E(X_2) - E(X_3) = \mu + \mu - \mu = \mu$

$\Rightarrow$ T_1 is an unbiased estimator of μ

$E(T_2) = 2E(X_1) + 3E(X_3) - 4E(X_2) = 2\mu + 3\mu - 4\mu = \mu$

$\Rightarrow$ T_2 is an unbiased estimator of μ.

(*ii*) We are given : $E(T_3) = \mu$

$$\Rightarrow \quad \frac{1}{3}[\lambda E(X_1) + E(X_2) + E(X_3) = \mu$$

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

(*iii*) With $\lambda = 1$, $T_3 = \frac{1}{3}(X_1 + X_2 + X_3) = \bar{X}$

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 $\text{Var}(T_3)$ is minimum, T_3 is the best estimator in the sense of minimum variance.

15·5·2. 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 $\gamma(\theta)$.

More precisely, T is MVUE of $\gamma(\theta)$ if

$$E_\theta(T) = \gamma(\theta) \text{ for all } \theta \in \Theta \qquad \ldots(15\cdot13)$$

and

$$\text{Var}_\theta(T) \le \text{Var}_\theta(T') \text{ for all } \theta \in \Theta \qquad \ldots(15\cdot14)$$

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

15·6. Sufficiency. An estimator is said to be sufficient for a parameter, if it contains all the information in the sample regarding the parameter. More precisely, 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 $$T = t(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \ldots + x_n \sim B(n, p)$$

$$\therefore \qquad P(T = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

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

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

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

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

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

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*

$$L = g_\theta[t(x)].h(x) \qquad \ldots(15\cdot 29)$$

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.

4. Invariance Property of Sufficient Estimator.

If T is a sufficient estimator for the parameter θ and if $\psi(T)$ is a one to one function of T, then $\psi(T)$ is sufficient for $\psi(\theta)$.

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

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

$$= g_1(t_1, \theta).\, k(x_1, x_2, \ldots, x_n) \qquad \ldots(15\cdot 31)$$

where $g_1(t_1, \theta)$ is the p.d.f. of 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 15·14. *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

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

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

$$= g_\theta\,[t(\mathbf{x})].\, h(\mathbf{x})$$

where

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

$$t(\mathbf{x}) = [t_1(\mathbf{x}), t_2(\mathbf{x})] = (\Sigma x_i, \Sigma x_i^2) \text{ and } h(\mathbf{x}) = 1$$

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

Example 15·16. *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 $t_1 = \prod_{i=1}^{n} X_i$, is sufficient for θ.

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

$$= \theta^n \left(\prod_{i=1}^{n} x_i\right)^\theta . \frac{1}{\left(\prod_{i=1}^{n} x_i\right)}$$

$$= g(t_1, \theta).\, h(x_1, x_2, \ldots, x_n), \text{ (say).}$$

Hence by Factorisation Theorem,

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

15·7. Cramer-Rao Inequality

Theorem 15·8. *If t is an unbiased estimator for $\gamma(\theta)$, a function of parameter θ, then*

$$Var(t) \geq \frac{\left[\frac{d}{d\theta} \cdot \gamma(\theta)\right]^2}{E\left[\frac{\partial}{\partial\theta}\log L\right]^2} = \frac{[\gamma'(\theta)]^2}{I(\theta)} \qquad \ldots(15\cdot32)$$

where $I(\theta)$ is the information on θ, supplied by the sample.

In other words, Cramer-Rao inequality provides a lower bound $[\gamma'(\theta)]^2/I(\theta)$, to the variance of an unbiased estimator of $\gamma(\theta)$.

Remarks. 1. An unbiased estimator t of $\gamma(\theta)$ for which Cramer-Rao lower bound in (15·32) is attained is called a *minimum variance bound* (*MVB*) *estimator.*

2. We have :

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log L\right)^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\log L\right] \qquad \ldots(15\cdot38)$$

and

$$I(\theta) = n\left[\frac{\partial}{\partial\theta}\log f(x, \theta)\right]^2 = -n\left[\frac{\partial^2}{\partial\theta^2}\log f\right] \qquad \ldots(15\cdot38a)$$

15·10. Methods of Estimation. So far we have been discussing the requisites of a good estimator. Now we shall briefly outline some of the important methods for obtaining such estimators. Commonly used methods are

(*i*) *Method of Maximum Likelihood Estimation.*

15·11. Method of Maximum Likelihood Estimation. From theoretical point of view, the most general method of estimation known is the method of *Maximum Likelihood Estimators* (M.L.E.) which was initially formulated by C.F. Gauss but as a general method of estimation was first introduced by Prof. R.A. Fisher and later on developed by him in a series of papers. Before introducing the method we will first define *Likelihood Function.*

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

$$L = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta). \qquad \ldots(15\cdot53)$$

L 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 $L(\theta)$ for variations in parameter *i.e.*, we wish to find $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ so that

$$L(\hat{\theta}) > L(\theta) \qquad \forall\ \theta \in \Theta$$

$$i.e., \qquad L(\hat{\theta}) = \operatorname{Sup} L(\theta)\ \forall\ \theta \in \Theta.$$

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

$$\frac{\partial L}{\partial \theta} = 0 \quad \text{and} \quad \frac{\partial^2 L}{\partial \theta^2} < 0 \qquad \ldots(15{\cdot}54)$$

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 (15·54) can be rewritten as

$$\frac{1}{L} \cdot \frac{\partial L}{\partial \theta} = 0 \quad \Rightarrow \quad \frac{\partial \log L}{\partial \theta} = 0, \qquad \ldots(15{\cdot}54a)$$

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 :

$$\frac{\partial}{\partial \theta_i} \log L = \frac{\partial}{\partial \theta_i} \log L\,(\theta_1, \theta_2, \ldots, \theta_k) = 0\ ;\ i = 1, 2, \ldots, k \qquad \ldots(15{\cdot}54b)$$

Equations (15·54a) and (15·54b) 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

$$\left(\frac{\partial^2 \log L}{\partial \theta_i\, \partial \theta_j}\right)_{\theta = \hat{\theta}} \text{ should be negative definite.}$$

15·11·1. Properties of Maximum Likelihood Estimators.

We make the following assumptions, known as the *Regularity Conditions* :

(*i*) The first and second order derivatives, *viz.*, $\dfrac{\partial \log L}{\partial \theta}$ and $\dfrac{\partial^2 \log L}{\partial \theta^2}$ 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

$$\left|\frac{\partial}{\partial \theta} \log L\right| < F_1(x) \text{ and } \left|\frac{\partial^2}{\partial \theta^2} \log L\right| < F_2(x)$$

where $F_1(x)$ and $F_2(x)$ are integrable functions over $(-\infty, \infty)$.

(*ii*) The third order derivative $\dfrac{\partial^2}{\partial \theta^3} \log L$ 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,

$$E\left(-\frac{\partial^2}{\partial \theta^2} \log L\right) = \int_{-\infty}^{\infty} \left(-\frac{\partial^2}{\partial \theta^2} \log L\right) L dx$$
$$= I(\theta),$$

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 15·11. (*Cramer-Rao Theorem*). *"With probability approaching unity as $n \to \infty$, the likelihood equation $\dfrac{\partial}{\partial \theta} \log L = 0$, 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 15·31],

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

$\text{MLE}(\sigma^2) = s^2$ (sample variance), which is consistent but not unbiased estimator of σ^2.

Theorem 15·12. (*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 15·13. (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 $N\left(\theta_0, \frac{I}{I(\theta_0)}\right)$ as $n \to \infty$.

Remark. Variance of M.L.E. is given by

$$V(\hat{\theta}) = \frac{1}{I(\theta)} = \frac{1}{\left[E\left(-\frac{\partial^2}{\partial\theta^2}\log L\right)\right]} \qquad \ldots(15·55)$$

Theorem 15·14. *If M.L.E. exists, it is the most efficient in the class of such estimators.*

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

Theorem 15·16. *If for a given population with p.d.f. $f(x, \theta)$, an MVB estimator T exists for θ, then the likelihood equation will have a solution equal to the estimator T.*

Theorem 15·17. (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 15·31. *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

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

or
$$\sum_{i=1}^{n} (x_i - \mu) = 0 \;\Rightarrow\; \sum_{i=1}^{n} x_i - n\mu = 0$$

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

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

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

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

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

$$\frac{\partial}{\partial\mu} \log L = 0 \text{ and } \frac{\partial}{\partial\sigma^2} \log L = 0, \text{ thus giving}$$

$$\hat{\mu} = \bar{x} \qquad \text{[From (*)]}$$

and
$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2 \qquad \text{[From (**)]}$$

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

Important Note. It may be pointed out here that though

$$\left.\begin{array}{l} E(\hat{\mu}) = E(\bar{x}) = \mu \\ E(\hat{\sigma}^2) = E(s^2) \neq \sigma^2 \end{array}\right\} \qquad (c.f. \text{ § } 12{\cdot}12)$$

Hence the maximum likelihood estimators (M.L.Es.) need not necessarily be unbiased.

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 15·32. *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. **[*Burdwan Univ. B.Sc. (Maths. Hons.), 1991*]**

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 :

$$\frac{d}{d\alpha} \log L = \frac{d}{d\alpha}\left[\log 2 - 2 \log \alpha + \log(\alpha - x)\right] = 0$$

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

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 15·33. (*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

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

$$\therefore \qquad \log L = -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\log\lambda - \sum_{i=1}^{n}\log(x_i\,!)$$

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

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 the estimate is given by

$$\frac{1}{V(\hat{\lambda})} = E\left[-\frac{\partial^2}{\partial\lambda^2}(\log L)\right] \qquad [c.f.\ (15{\cdot}55)]$$

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

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

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

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

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

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