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

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

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*.

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*

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

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

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

where m is some very large value of n.

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

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

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

15·4. Unbiasedness. 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 $\gamma(\theta)$.

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

$$E(T_n) = \gamma(\theta), \text{ for all } \theta \in \Theta \qquad \ldots(15{\cdot}3)$$

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

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

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

$$b(\theta) = E(T_n) - \gamma(\theta), \theta \in \Theta \qquad \ldots(15{\cdot}3a)$$

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 $\psi(\gamma(\theta))$.*

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

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

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 $A \subseteq B \;\Rightarrow\; P(A) \leq P(B) \;\Rightarrow\; P(B) \geq P(A) \qquad \ldots(***)$

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

$$P\left[|\psi(T_n) - \psi(\gamma(\theta)| < \varepsilon_1\right] \geq P\left[|T_n - \gamma(\theta)| < \varepsilon\right]$$

$$\Rightarrow \quad P\left[|\psi(T_n) - \psi(\gamma(\theta)| < \varepsilon_1\right] \geq 1 - \eta \;;\; \forall\; n \geq m \qquad \text{[Using (*)]}$$

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

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

15·4·2. Sufficient Conditions for Consistency.

Theorem 15·2. Let $\{T_n\}$ *be a sequence of estimators such that for all* $\theta \in \Theta$,

$$\text{(i)} \quad E_\theta(T_n) \to \gamma(\theta),\ n \to \infty$$

and

$$\text{(ii)} \quad Var_\theta(T_n) \to 0, \text{ as } n \to \infty.$$

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.,
$$P\left[\,|T_n - \gamma(\theta)| < \varepsilon\,\right] > 1 - \eta\ ;\ \forall\ n \geq m\ (\varepsilon, \eta) \qquad \ldots(15{\cdot}4)$$

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

$$P\left[\,|T_n - E_\theta(T_n)| \leq \delta\,\right] \geq 1 - \frac{\mathrm{Var}_\theta(T_n)}{\delta^2} \qquad \ldots(15{\cdot}5)$$

We have

$$|T_n - \gamma(\theta)| = |T_n - E(T_n) + E(T_n) - \gamma(\theta)|$$

Now

$$|T_n - E_\theta(T_n)| \le \delta \Rightarrow |T_n - \gamma(\theta)| \le \delta + |E_\theta(T_n) - \gamma(\theta)| \quad \ldots(15{\cdot}7)$$

Hence, on using (***) of Theorem 15·1, we get

$$P\left[|T_n - \gamma(\theta)| \le \delta + |E_\theta(T_n) - \gamma(\theta)|\right] \ge P\left[|T_n - E_\theta(T_n)| \le \delta\right]$$

$$\ge 1 - \frac{\text{Var}_\theta(T_n)}{\delta^2} \quad \text{[From (15·5)]} \ldots(15{\cdot}8)$$

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

$$|E_\theta(T_n) - \gamma(\theta)| \le \delta_1, \;\forall\; n \ge n_0(\delta_1) \quad \ldots(15{\cdot}9)$$

Also $\text{Var}_\theta(T_n) \to 0$ as $n \to \infty$, (Given).

$$\therefore \qquad \frac{\text{Var}_\theta(T_n)}{\delta^2} \le \eta, \;\forall\; n \ge n_0'(\eta) \quad \ldots(15{\cdot}10)$$

where η is arbitrarily small positive number.

Substituting from (15·9) and (15·10) in (15·8), we get

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

$$\Rightarrow \qquad P\left[|T_n - \gamma(\theta)| \le \varepsilon\right] \ge 1 - \eta \;;\; n \ge m$$

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

$$\Rightarrow \qquad T_n \xrightarrow{p} \gamma(\theta), \text{ as } n \to \infty \qquad \text{[Using (15·4)]}$$

$\Rightarrow$ $\qquad T_n$ 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 $$\text{Var}(T) = E(T^2) - [E(T)]^2 = E(T^2) - \theta^2$$

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

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

Example 15·3. *Show that* $\frac{[\sum x_i (\sum x_i - 1)]}{n(n-1)}$ *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;

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

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

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

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

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

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

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

$\Rightarrow \quad [\sum x_i\,(\sum x_i - 1)]\,/\,[n(n-1)]$ 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* $\exp[-(k+1)\theta]$, $k > 0$, *is* $T(X) = (-k)^X$ *so that*

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

and

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

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

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

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

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 $\bar{x}$ is also normally distributed as $N(\mu, \sigma^2/n)$.

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

Thus as $n \to \infty$,

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

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 :*

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

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,

$$V(Md) = \frac{1}{4n f_1^2} \qquad \text{[c.f. Example 15·5(b)]}$$

Here f_1 = Median ordinate of the parent distribution.

= Modal ordinate of the parent distribution.

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

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

Since and

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

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

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

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

We have seen above :

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

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

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 :*

$$E = \frac{V_1}{V_2} \qquad \text{...(15·12)}$$

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 :

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

Obviously $E_i \leq 1, i = 1, 2, \ldots n$.

For example, in the normal samples, since sample mean $\bar{x}$ is the most efficient estimator of μ [c.f. Remark to Example 15·31], 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 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

$E(X_i) = \mu$, Var $(X_i) = \sigma^2$, (say) ; 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) $$E(t_2) = \frac{1}{2} E(X_1 + X_2) + E(X_3)$$

$$= \frac{1}{2} (\mu + \mu) + \mu \qquad \text{[Using (*)]}$$

$$= 2\mu$$

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

(iii) $$E(t_3) = \mu$$

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

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

$$\Rightarrow \quad 2\mu + \mu + \lambda\mu = 3\mu$$

$$\Rightarrow \quad \lambda\mu = 0 \quad \Rightarrow \quad \lambda = 0$$

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*

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

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

$\Rightarrow \quad T_1$ is an unbiased estimator of μ

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where (as indicated) $g_\theta[t(\mathbf{x})]$ depends on θ and $\mathbf{x}$ only through the value of $t(\mathbf{x})$ and $h(\mathbf{x})$ 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

$$L = L(\mathbf{x}, \theta) = g(\mathbf{x}).h(\theta).\exp\{a(\theta)\psi(\mathbf{x})\} \qquad \ldots(15{\cdot}30)$$

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.

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 :*

$$
\begin{aligned}
L &= \prod_{i=1}^{n} f(x_i, \theta) \\
&= g_1(t_1, \theta) . k(x_1, x_2, \ldots, x_n) \qquad \ldots(15{\cdot}31)
\end{aligned}
$$

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

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

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

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

where

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

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

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

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

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

Hence by Factorisation Theorem,

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