17·1. INTRODUCTION

One of the main objectives of Statistics is to draw inferences about a popula from the analysis of a sample drawn from that population. Two important proble 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 fundamental papers round about 1930.

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

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, wi probability function $f(x ; \theta_1, \theta_2, \ldots, \theta_k)$, where $\theta_1, \theta_2, \ldots, \theta_k$ are the unknown populatio parameters. There will then always be an infinite number of functions of samp values, called statistics, which may be proposed as estimates of one or more of th parameters.

Evidently, the best estimate would be one that falls nearest to the true value of th parameter to be estimated. In other words, the statistic whose distributi concentrates as closely as possible near the true value of the parameter may b regarded the best estimate. Hence the basic problem of the estimation in the abo 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*

$$E(T_n) = \gamma(\theta), \text{ for all } \theta \in \Theta \qquad \ldots (17·1)$$

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 $\quad b(\theta) = E(T_n) - \gamma(\theta), \theta \in \Theta \quad \ldots(17·1a)$

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 $\text{Var}(T) = n\theta(1 - \theta)$

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

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

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

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

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

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

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

$\Rightarrow$ $T(X) = (-k)^X$ 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 sa… to be consistent estimator of* $\gamma(\theta)$, $\theta \in \Theta$, *the parameter space, if* T_n *converges to* $\gamma(\theta)$ *i… probability, i.e., if* $T_n \xrightarrow{p} \gamma(\theta)$ *as* $n \to \infty$. *In other words,* T_n *is a consistent estimator… $\gamma(\theta)$ if for every* $\varepsilon > 0$, $\eta > 0$, *there exists a positive integer* $n \geq m\,(\varepsilon, \eta)$ *such that*

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

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 m… $EX_i = \mu < \infty$, 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).

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

Moreover, if there exists a consistent estimator, say, T_n of $\gamma(\theta)$, then infinitely many su… 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.,*

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

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

$$A \subseteq B \;\Rightarrow\; P(A) \leq P(B) \quad \text{or} \quad P(B) \geq P(A) \quad \ldots(***)$$

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

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

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

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

i.e., $P[|T_n - \gamma(\theta)| < \varepsilon] > 1 - \eta\,;\; \forall\, n \geq m\,(\varepsilon, \eta)$ ...(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

$$P[|T_n - E_\theta(T_n)| \leq \delta] \geq 1 - \frac{\text{Var}_\theta(T_n)}{\delta^2} \quad \ldots(17\cdot4)$$

We have

$$|T_n - \gamma(\theta)| = |T_n - E(T_n) + E(T_n) - \gamma(\theta)| \leq |T_n - E_\theta(T_n)| + |E_\theta(T_n) - \gamma(\theta)| \quad \ldots(17\cdot5)$$

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

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

$$P\{|T_n - \gamma(\theta)| \leq \delta + |E_\theta(T_n) - \gamma(\theta)|\} \geq P\{|T_n - E_\theta(T_n)| \leq \delta\}$$

$$\geq 1 - \frac{\text{Var}_\theta(T_n)}{\delta^2} \quad \text{[From (17·4)]} \quad \ldots(17\cdot7)$$

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 $n \geq n_0(\delta_1)$ such that

$$|E_\theta(T_n) - \gamma(\theta)| \leq \delta_1,\; \forall\, n \geq n_0(\delta_1) \quad \ldots(17\cdot8)$$

Also $\text{Var}_\theta(T_n) \to 0$ as $n \to \infty$, (Given) $\therefore$ $\dfrac{\text{Var}_\theta(T_n)}{\delta^2} \leq \eta\,,\; \forall\, n \geq n_0'(\eta)$, ...(17·9)

where η is arbitrarily small positive number.

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

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

$\Rightarrow$ $P[|T_n - \gamma(\theta)| \leq \varepsilon] \geq 1 - \eta\,;\; n \geq m,$

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

$\Rightarrow \quad T_n \xrightarrow{p} \gamma(\theta)$ as $n \to \infty$ [Using (...)]

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

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

(b) Prove that for Cauchy's distribution not sample mean but sample median [is a] 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[ts] at $x = \mu$. If $\bar{x}$, the sample mean is taken as an estimator of μ, then the sampl[ing] 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 of [...]

Since in this case, the distribution of $\bar{x}$ is same as distribution of any single sam[ple] 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 $M[d]$ [is] an unbiased estimate of the population mean, which of course is same as [the] population median. Therefore $\qquad E(Md) = \mu$.

For large n, the sampling distribution of median is asymptotically normal a[nd] 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.,*

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

But f_1 = Median ordinate of (*) = Modal ordinate of (*) $\qquad$ [Because of symme[try]

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

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

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

Hence from (**) and (****), using *Theorem 17·2*, we conclude that for Cauch[y] 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',

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

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

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

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 $\frac{\Sigma x_i}{n}\left(1 - \frac{\Sigma x_i}{n}\right) = \bar{X}(1-\bar{X})$, 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 $\quad f_1$ = 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 $\quad E(Md) = \mu$
and $\quad V(Md) \to 0$ $\Big\}$, 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

$$V(T_1) < V(T_2), \text{ for all } n \qquad \text{...(17·10)}$$

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 m... 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 paramete... exists one whose sampling variance is less than that of any such estimator, it is called th... efficient estimator. Whenever such an estimator exists, it provides a criterion for measu... of efficiency of the other estimators.*

> **Efficiency** *(Definition) If T_1 is the most efficient estimator with variance V_1 and any other estimator with variance V_2, then the efficiency E of T_2 is defined as :*
>
> $$E = \frac{V_1}{V_2}$$ ...(17
>
> *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, the efficiency E_i of T_i, $(i = 1, 2, \ldots, n)$ is defined as :

$$E_i = \frac{\text{Var } T}{\text{Var } T_i}\,;\ i = 1, 2, \ldots, n$$ ...(17

Obviously $E_i \le 1$; $i = 1, 2, \ldots n$. For example, in the normal samples, since sa... mean $\bar{x}$ is the most efficient estimator of μ [*c.f. Remark to Example 17·31*], the effic... 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 n... 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 a... t_1, t_2 and t_3.

Solution. We are given :

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

(i) $E(t_1) = \frac{1}{5}\sum_{i=1}^{5} E(X_i) = \frac{1}{5}\sum_{i=1}^{5}\mu = \frac{1}{5}.5\mu = \mu \Rightarrow t_1$ is an unbiased estima...

(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

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

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

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

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

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

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

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

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

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

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

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

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

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

Since $\text{Var}(T_3)$ 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

$$E_\theta(T) = \gamma(\theta) \text{ for all } \theta \in \Theta \quad \text{...(17·13)}$$

and

$$\text{Var}_\theta(T) \le \text{Var}_\theta(T') \text{ for all } \theta \in \Theta \quad \text{...(17·14)}$$

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

Example 17·11. *(a) Show that if a most efficient estimator A and a less effi[cient] estimator B with efficiency e tend to joint normality for large samp[les], B – A tends to zero correlation with A.*

(b) Show that the error in B may be regarded as composed (for large samples) of two p[arts] which are independent, the error in A and the error in (B –A).

(c) Show further that $V(A-B)=\left(\frac{1}{e}-1\right)V(A)$.

Solution. (a) We have to prove that $r\{A,(B-A)\}=0 \Rightarrow \text{Cov}(A,B-A)=0$

$$\text{Cov}\{A,(B-A)\}=\text{Cov}(A,B)-V(A)=\rho\sigma_A\sigma_B-\sigma_A^2,$$

where ρ is the correlation coefficient between A and B.

If we take $\sigma_A=\sigma$, then $\sigma_B=\frac{\sigma}{\sqrt{e}}$ and $\rho=\sqrt{e}$ (c.f. Theorem 17[…])

$\therefore$ $\text{Cov}(A,B-A)=\sqrt{e}\,.\,\sigma.\frac{\sigma}{\sqrt{e}}-\sigma^2=0.$ Hence $(B-A)$ has zero correlation with A

(b) We have $B=A+(B-A)$

$$\therefore\quad V(B)=V[A+(B-A)]=V(A)+V(B-A)+2\,\text{Cov}(A,B-A)$$
$$=V(A)+V(B-A)$$ [Using part (a)]

$\Rightarrow$ Error in B = Error in A + Error in $(B-A)$

and since A and $(B-A)$ are independent, [c.f. part (a) viz., $r(A,B-A)=0$ and A and tend to joint normality], the result follows.

(c) $$V(A-B)=V(A)+V(B)-2\,\text{Cov}(A,B)=\sigma_A^2+\sigma_B^2-2\rho\sigma_A\sigma_B$$
$$=\sigma^2+\frac{\sigma^2}{e}-2\sqrt{e}.\sigma.\frac{\sigma}{\sqrt{e}}=\frac{\sigma^2}{e}-\sigma^2=\left(\frac{1}{e}-1\right)\sigma^2.$$

Example 17·12. *If T_1 and T_2 are two unbiased estimators of $\gamma(\theta)$, having the sam[e] variance and ρ is the correlation between them, then show that $\rho\ge 2e-1$, where e is th[e] efficiency of each estimator.*

Solution. Let T be MVUE of $\gamma(\theta)$. Then, since $V(T_1)=V(T_2)$, the efficiency e o[f] each estimator is given by: $e=\frac{V(T)}{V(T_1)}=\frac{V(T)}{V(T_2)}$ ...(*)

Consider another unbiased estimator of $\gamma(\theta)$ viz., $T_3=\frac{1}{2}(T_1+T_2)$

$$\Rightarrow\quad V(T_3)=\frac{1}{4}[V(T_1)+V(T_2)+2\,\text{Cov}(T_1,T_2)]$$
$$=\frac{1}{4}\left\{\frac{V(T)}{e}+\frac{V(T)}{e}+2\rho\sqrt{\frac{V(T)}{e}\cdot\frac{V(T)}{e}}\right\}\quad\text{[From (*)]}$$
$$=\frac{V(T)}{4e}(1+1+2\rho)=\frac{(1+\rho)V(T)}{2e}$$

Since $V(T)$ is the minimum variance, $V(T_3)=\frac{(1+\rho).V(T)}{2e}\ge V(T)\Rightarrow\rho\ge(2e-1).$

Aliter. Deduction From (17·25), page 17·11. If T_1 and T_2 have same variances/efficiencies *i.e.*, $e_1=e_2=e$, (say), then (17·25) gives

$$e-(1-e)\le\rho\le e+(1-e)\quad\Rightarrow\quad\rho\ge 2e-1.$$

17·2·4. Sufficiency. An estimator is said to be sufficient for a parameter, if it contains all the information in the sample regarding the parameter.

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:

$$P(x_1\cap x_2\cap\ldots\cap x_n\mid T=k)=\frac{P(x_1\cap x_2\cap\ldots\cap x_n\cap T=k)}{P(T=k)}$$
$$=\begin{cases}\dfrac{p^k(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}}=\dfrac{1}{\binom{n}{k}}\\[2ex]0,\text{ if }\sum_{i=1}^n x_i\ne k\end{cases}$$

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

Theorem 15·7. FACTORIZATION THEOREM (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:*

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

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:

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

where $h(\theta)$ and $a(\theta)$ are functions of the parameter θ only and $g(x)$ and $\psi(x)$ 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 para[meter] 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 parame[ter] *if and only if the likelihood function (joint p.d.f. of the sample) can be expressed as :*

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

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 observati[ons] only, independent of θ.

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

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

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

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

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

where $g_\theta[t(x)] = \frac{k\{t(x), \theta\}}{\theta^n}$, $t(x) = \max_{1\le i\le n} x_i$ and $h(x) = k(0, \min_{1\le i\le n} x_i)$

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

Aliter. We have $L = \prod_{i=1}^{n} f(x_i, \theta) = \frac{1}{\theta^n}\,; 0 < x_i < \theta \quad \ldots$

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

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

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

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

Then
$$L = \prod_{i=1}^{n} f_\theta(x_i) = \left\{\frac{1}{\sigma\sqrt{2\pi}}\right\}^n . \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$
$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 - 2\mu\sum x_i + n\mu^2\right)\right\}$$
$$= g_\theta\,[t(x)] . h(x)$$

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

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

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

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

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

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

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

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

17·3. CRAMER-RAO INEQUALITY

Definition. *If t is an unbiased estimator for γ(θ), a function of parameter θ, then*

$$Var(t) \geq \frac{\left\{\frac{d}{d\theta} \cdot \gamma(\theta)\right\}^2}{E\left(\frac{\partial}{\partial\theta}\log L\right)^2} = \frac{\{\gamma'(\theta)\}^2}{I(\theta)} \quad \ldots(17\cdot32)$$

where I(θ) 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 γ(θ).

Proof. In proving this result, we assume that there is only a single parameter θ which is unknown. We also take the case of continuous *r.v.* The case of descrete random variables can be dealt with similarly on replacing the multiple integrals by appropriate multiple sums.

We further make the following assumptions, which are known as the *Regularity conditions* for *Cramer-Rao Inequality.*

(1) The parameter space Θ is a non-degenerate open interval on the real line $R^1(-\infty, \infty)$.

(2) For almost all $x = (x_1, x_2, \ldots, x_n)$, and for all $\theta \in \Theta$, $\frac{\partial}{\partial\theta} L(x, \theta)$ exists, the exceptional set, if any, is independent of θ.

(3) The range of integration is independent of the parameter θ, so that $f(x, \theta)$ is differentiable under integral sign.

If range is not independent of θ and f is zero at the extremes of the range, *i.e.,* $f(a, \theta) = 0 = f(b, \theta)$, then

$$\frac{\partial}{\partial\theta}\int_a^b f\,dx = \int_a^b \frac{\partial f}{\partial\theta}dx - f(a, \theta)\frac{\partial a}{\partial\theta} + f(b, \theta)\frac{\partial b}{\partial\theta}$$

$$\Rightarrow \quad \frac{\partial}{\partial\theta}\int_a^b f\,dx = \int_a^b \frac{\partial f}{\partial\theta}dx, \text{ since } f(a, \theta) = 0 = f(b, \theta)$$

(4) The conditions of uniform convergence of integrals are satisfied so that differentiation under the integral sign is valid.

(5) $I(\theta) = E\left[\left\{\frac{\partial}{\partial\theta}\log L(x, \theta)\right\}^2\right]$, exists and is positive for all $\theta \in \Theta$.

Let X be a *r.v.* following the *p.d.f.* $f(x, \theta)$ and let L be the likelihood function of the random sample $(x_1, x_2, \ldots, x_n)$ from this population. Then

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

Since L is the joint *p.d.f.* of $(x_1, x_2, \ldots, x_n)$, $\int L(x, \theta)\,dx = 1$,

where $\int dx = \iint \ldots \int dx_1\,dx_2 \ldots dx_n$.

Differentiating w.r. to θ and using regularity conditions given above, we get

$$\int \frac{\partial}{\partial\theta} L\,dx = 0 \Rightarrow \int \left(\frac{\partial}{\partial\theta}\log L\right) L\,dx = 0 \Rightarrow E\left(\frac{\partial}{\partial\theta}\log L\right) = 0 \quad \ldots(17\cdot33)$$

Let $t = t(x_1, x_2, \ldots, x_n)$ be an unbiased estimator of γ(θ) such that

$$E(t) = \gamma(\theta) \Rightarrow \int t \cdot L\,dx = \gamma(\theta) \quad \ldots(17\cdot34)$$

Differentiating w.r. to θ, we get $\int t \cdot \frac{\partial L}{\partial\theta}dx = \gamma'(\theta) \Rightarrow \int t\left(\frac{\partial}{\partial\theta}\log L\right)L\,dx = \gamma'(\theta)$

$$\Rightarrow \quad E\left(t \cdot \frac{\partial}{\partial\theta}\log L\right) = \gamma'(\theta) \quad \ldots(17\cdot35)$$

$$\text{Cov}\left(t, \frac{\partial}{\partial\theta}\log L\right) = E\left(t \cdot \frac{\partial}{\partial\theta}\log L\right) - E(t) \cdot E\left(\frac{\partial}{\partial\theta}\log L\right)$$

$$= \gamma'(\theta)$$

[From (17·33) and (17·35)] ... (17·35a)

We have: $\{r(X, Y)\}^2 \leq 1 \Rightarrow \{\text{Cov}(X, Y)\}^2 \leq \text{Var}(X) \cdot \text{Var}(Y)$

$$\therefore \quad \left\{\text{Cov}\left(t, \frac{\partial}{\partial\theta}\log L\right)\right\}^2 \leq \text{Var}\,t \cdot \text{Var}\left(\frac{\partial}{\partial\theta}\log L\right)$$

$$\Rightarrow \quad \{\gamma'(\theta)\}^2 \leq \text{Var}\,t\left[E\left(\frac{\partial}{\partial\theta}\log L\right)^2 - \left\{E\left(\frac{\partial}{\partial\theta}\log L\right)\right\}^2\right]$$

$$\Rightarrow \quad \{\gamma'(\theta)\}^2 \leq \text{Var}\,t \cdot E\left\{\left(\frac{\partial}{\partial\theta}\log L\right)^2\right\} \quad \text{[Using (17·33)]}\ldots(17\cdot36)$$

$$\Rightarrow \quad \text{Var}(t) \geq \frac{\{\gamma'(\theta)\}^2}{E\left\{\left(\frac{\partial}{\partial\theta}\log L\right)^2\right\}} \quad \ldots(17\cdot36a)$$

which is Cramer-Rao Inequality.

Corollary. If t is an unbiased estimator of parameter θ, *i.e.,*

$$E(t) = \theta \Rightarrow \gamma(\theta) = \theta \text{ or } \gamma'(\theta) = 1,$$

then from (17·36a), we get

$$\text{Var}(t) \geq \frac{1}{E\left\{\left(\frac{\partial}{\partial\theta}\log L\right)^2\right\}} = \frac{1}{I(\theta)}, \text{ where } I(\theta) = E\left\{\left(\frac{\partial}{\partial\theta}\log L\right)^2\right\} \quad ...(17·$$

is called by R.A. Fisher as the *amount of information* on θ supplied by the sample $(x_1, x_2, \ldots, x_n)$ and its reciprocal $1/I(\theta)$, as the *information limit* to the variance of estimator $t = t(x_1, x_2, \ldots, x_n)$.

Remarks.1. An unbiased estimator t of $\gamma(\theta)$ for which Cramer-Rao lower bound in (17·32) is attained is called a *minimum variance bound (MVB) estimator.*

2. We have :

$$I(\theta) = E\left\{\left(\frac{\partial}{\partial\theta}\log L\right)^2\right\} = -E\left(\frac{\partial^2}{\partial\theta^2}\log L\right) \quad ...(17·3$$

and

$$I(\theta) = n\left\{\frac{\partial}{\partial\theta}\log f(x, \theta)\right\}^2 = -n\left(\frac{\partial^2}{\partial\theta^2}\log f\right) \quad ...(17·38$$

Proof. We have proved in (17·33), $\quad E\left(\frac{\partial}{\partial\theta}\log L\right) = 0 \quad ...($

$$\left(\frac{\partial^2}{\partial\theta^2}\log L\right)L = \frac{\partial}{\partial\theta}\left\{\left(\frac{\partial}{\partial\theta}\log L\right).L\right\} - \left(\frac{\partial}{\partial\theta}\log L\right).\frac{\partial L}{\partial\theta}$$

$$= \frac{\partial}{\partial\theta}\left\{\left(\frac{\partial}{\partial\theta}\log L\right).L\right\} - \left(\frac{\partial}{\partial\theta}\log L\right)^2.L$$

Integrating both sides w.r. to $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, we get

$$E\left(\frac{\partial^2}{\partial\theta^2}\log L\right) = \frac{\partial}{\partial\theta}.E\left(\frac{\partial}{\partial\theta}\log L\right) - E\left(\frac{\partial}{\partial\theta}\log L\right)^2 = -E\left(\frac{\partial}{\partial\theta}\log L\right)^2 \quad \text{[Using (*)]}$$

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

a form which is more convenient to use in practice.

Also

$$I(\theta) = E\left\{\left(\frac{\partial}{\partial\theta}\log L\right)^2\right\} = E\left\{\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i, \theta)\right\}^2$$

$$= E\left[\sum_{i=1}^{n}\left\{\frac{\partial}{\partial\theta}\log f(x_i, \theta)\right\}^2 + \sum_{i \neq j = 1}^{n}\left\{\left(\frac{\partial}{\partial\theta}\log f(x_i, \theta)\right)\left(\frac{\partial}{\partial\theta}\log f(x_j, \theta)\right)\right\}\right]$$

$$= n.E\left\{\frac{\partial}{\partial\theta}\log f(x, \theta)\right\}^2 \quad \text{[On using (*)], since } x_i\text{'s}; i = 1, 2, \ldots, n \text{ are i.i.d. r.v.'s]}$$

17·3·1. Conditions for the Equality Sign in Cramer-Rao Inequality.

In proving (17·32) we used [*c.f.* (17·36)] that

$$[\gamma'(\theta)]^2 \leq E[t - \gamma(\theta)]^2 . E\left(\frac{\partial}{\partial\theta}\log L\right)^2 \quad ...(17·3$$

The sign of equali...

$$\frac{t - \gamma(\theta)}{\frac{\partial}{\partial\theta}\log L} = \lambda = \lambda(\theta),$$

where λ is a constant independent of $(x_1, x_2, \ldots, x_n)$ but may depend on θ.

$$\frac{\partial}{\partial\theta}\log L = \frac{t - \gamma(\theta)}{\lambda(\theta)} = [t - \gamma(\theta)]\, A(\theta), \quad ...(17·40)$$

where

$$A = A(\theta) = 1/[\lambda(\theta)], \text{ say.}$$

Hence, *a necessary and sufficient condition for an unbiased estimator t to attain the lower bound of its variance is given by (17·40).*

Further, the C-R minimum variance bound is given by :

$$\text{Var}(t) = [\gamma'(\theta)]^2 / E\left(\frac{\partial}{\partial\theta}\log L\right)^2 \quad ...(17·41)$$

But

$$E\left(\frac{\partial}{\partial\theta}\log L\right)^2 = E[A(\theta).\{t - \gamma(\theta)\}]^2 \quad \text{[From (17·40)]}$$

$$= \{A(\theta)\}^2 . E\{t - \gamma(\theta)\}^2 = \{A(\theta)\}^2 . \text{Var}(t)$$

Substituting in (17·41), we get $\quad \text{Var}(t) = \dfrac{\{\gamma'(\theta)\}^2}{\{A(\theta)\}^2 . \text{Var}(t)}$

$$\Rightarrow \quad \text{Var}(t) = \left|\frac{\gamma'(\theta)}{A(\theta)}\right| = |\gamma'(\theta).\lambda(\theta)| \quad ...(17·42)$$

Hence if the likelihood function L is expressible in the form (17·40) then

(*i*) t is an unbiased estimator of $\gamma(\theta)$,

(*ii*) Minimum Variance Bound (MVB) estimator (t) for $\gamma(\theta)$ exists, and

(*iii*) $\text{Var}(t) = \left|\frac{\gamma'(\theta)}{A(\theta)}\right| = |\gamma'(\theta)\,\lambda(\theta)|$

The importance of this result lies in the fact that C.R. inequality, in addition to find if MVBU estimator for $\gamma(\theta)$ exists, also gives us the variance of such an estimator, which is given by (17·42).

Remarks 1. If $\gamma(\theta) = \theta$, *i.e.*, if t is an unbiased estimator of θ, then (15·40) can be written as :

$$\frac{\partial}{\partial\theta}\log L = \frac{t - \theta}{\lambda} \quad ...(17·43)$$

Hence if (17·43) holds, then t is an MVB estimator for θ with

$$\text{Var}(t) = |\lambda(\theta)| = 1/|A(\theta)| \quad ...(17·43a)$$

2. We have seen in (17·40) that MVB estimator exists for $\gamma(\theta)$ if

$$\frac{\partial}{\partial\theta}\log L = \frac{t - \gamma(\theta)}{\lambda} = \{t - \gamma(\theta)\}.\frac{1}{\lambda}, \text{ where } \lambda = \lambda(\theta), \text{ say.} \quad ...(*)$$