Unit III

17·1. INTRODUCTION

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

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

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

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

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

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

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

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

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

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

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

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

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

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

17·2. CHARACTERISTICS OF ESTIMATORS.

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

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

17·2·1. Unbiasedness.

Definition. *An estimator $T_n = T(x_1, x_2, \ldots, x_n)$ is said to be an unbaised estimator of $\gamma(\theta)$ if* $$E(T_n) = \gamma(\theta), \text{ for all } \theta \in \Theta \quad \ldots (17\cdot1)$$

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

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

Remark. If $E(T_n) > \theta$, T_n is said to be positively biased and if $E(T_n) < \theta$, it is said to be negatively biased, the amount of bias $b(\theta)$ being given by $b(\theta) = E(T_n) - \gamma(\theta), \theta \in \Theta$ ...(17·1*a*)

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

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

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

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, $T = \sum_{i=1}^{n} x_i \sim B(n, \theta) \Rightarrow E(T) = n\theta$ 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 that only unbiased estimator of exp* $\{-(k+1)\theta\}$, $k > 0$, *is* $T(X) = (-k)^X$ *so that* $T(x) > 0$ *if x is even and* $T(x) < 0$ *if x is odd.*

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

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

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

17·2·2. Consistency

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

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

where m is some very large value of n.

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

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

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 for indefinitely large values of the sample size n, *i.e.,* as $n \to \infty$. Nothing is regarded of its behaviour for finite n.

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

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

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

Invariance Property of Consistent Estimators.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$P\big[|T_n - E_\theta(T_n)| \le \delta\big] \ge 1 - \frac{\text{Var}_\theta(T_n)}{\delta^2} \quad \ldots(17·4)$$

We have

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

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

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

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

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

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

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

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

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

where η is arbitrarily small positive number.

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

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

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

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

$$T_n \xrightarrow{P} \gamma(\theta) \text{ as } n \to \infty$$ [Using (17·

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

$\therefore$

Example 17·5. *(a) Prove that in sampling from a $N(\mu, \sigma^2)$ population, the sample mean 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 $n \to \infty$, $E(\bar{x}) = \mu$ 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 :

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

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

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

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

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

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

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 Md is an unbiased estimate of the population mean, which of course is same as the population median. Therefore $E(Md) = \mu$. $\ldots(**)$

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

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

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

But f_1 = Median ordinate of (*) = Modal ordinate of (*) [Because of symmetry]

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

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

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

Hence from (**) and (****), using *Theorem 17·2*, we conclude that for Cauchy's 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',

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

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

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

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

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

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

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 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

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

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

We have seen above :

For all n, $V(\bar{x}) = \dfrac{\sigma^2}{n}$ and for large n, $V(Md) = \dfrac{\pi\sigma^2}{2n} = 1.57\dfrac{\sigma^2}{n}$

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.

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

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

$$E = \frac{V_1}{V_2} \quad \ldots(17\cdot11)$$

Obviously, E cannot exceed unity.

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

$$E_i = \frac{\text{Var } T}{\text{Var } T_i}; \; i = 1, 2, \ldots, n \quad \ldots(17\cdot11a)$$

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

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

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

(i) $t_1 = \dfrac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$, (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 :

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

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

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

$\Rightarrow$ 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)

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

Using (*), we get

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

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

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

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

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

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

(i) *Are T_1 and T_2 unbiased estimators ?*

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

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

(iv) *Which is the best estimator ?*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

$$\text{Var}_\theta(T) \leq \text{Var}_\theta(T') \text{ for all } \theta \in \Theta \qquad \ldots(17·14)$$

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

We give below some important Theorems concerning MVU estimators.

Theorem 17·3. *An M.V.U. is unique in the sense that if T_1 and T_2 are M.V.U. estimators for $\gamma(\theta)$, then $T_1 = T_2$, almost surely.*

Proof. We are given that

$$\left.\begin{aligned} E_\theta(T_1) &= E_\theta(T_2) = \gamma(\theta), \text{ for all } \theta \in \Theta \\ \text{and} \quad \text{Var}_\theta(T_1) &= \text{Var}_\theta(T_2), \text{ for all } \theta \in \Theta \end{aligned}\right\} \quad \ldots(17\cdot15)$$

Consider a new estimator, $T = \frac{1}{2}(T_1 + T_2)$ which is also unbiased since,

$$E(T) = \frac{1}{2}\{E(T_1) + E(T_2)\} = \gamma(\theta)$$

$$\text{Var}(T) = \text{Var}\left\{\frac{1}{2}(T_1 + T_2)\right\} = \frac{1}{4}\text{var}(T_1 + T_2) \quad [\because \text{Var}(CX) = C^2\,\text{Var}(X)]$$

$$= \frac{1}{4}\{\text{Var}(T_1) + \text{Var}(T_2) + 2\,\text{Cov}(T_1, T_2)\}$$

$$= \frac{1}{4}\{\text{Var}(T_1) + \text{Var}(T_2) + 2\rho\sqrt{\text{Var}(T_1)\,\text{Var}(T_2)}\}$$

$$= \frac{1}{2}\text{Var}(T_1)(1 + \rho), \quad \ldots [\text{From } (17\cdot15)]$$

where ρ is Karl Pearson's co-efficient of correlation between T_1 and T_2.

Since T_1 is the *MVU* estimator, $\text{Var}(T) \geq \text{Var}(T_1)$

$$\Rightarrow \quad \frac{1}{2}\text{Var}(T_1)(1 + \rho) \geq \text{Var}(T_1) \quad \Rightarrow \quad \frac{1}{2}(1 + \rho) \geq 1 \Rightarrow \rho \geq 1$$

Since $|\rho| \leq 1$, we must have $\rho = 1$, *i.e.,* T_1 and T_2 must have a linear relation of the form :

$$T_1 = \alpha + \beta T_2, \quad \ldots(17\cdot16)$$

where α and β are constants independent of $x_1, x_2, \ldots, x_n$ but may depend on θ, *i.e.,* we may have

$$\alpha = \alpha(\theta) \text{ and } \beta = \beta(\theta).$$

Taking expectation of both sides in (17·16) and using (17·15), we get

$$\theta = \alpha + \beta\theta \quad \ldots(17\cdot17)$$

Also from (17·16), we get $\text{Var}(T_1) = \text{Var}(\alpha + \beta T_2) = \beta^2\,\text{Var}(T_2)$

$$\Rightarrow \quad 1 = \beta^2 \quad \Rightarrow \quad \beta = \pm 1 \quad \ldots [\text{From } (17\cdot15)]$$

But since $\rho(T_1, T_2) = +1$, the coefficient of regression of T_1 on T_2 must be positive.

$$\therefore \quad \beta = 1 \quad \Rightarrow \quad \alpha = 0 \quad [\text{From } 17\cdot17)]$$

Substituting in (17·16), we get $T_1 = T_2$ as desired.

Theorem 17·4. *Let T_1 and T_2 be unbiased estimators of $\gamma(\theta)$ with efficiencies e_1 and e_2 respectively and $\rho = \rho_\theta$ be the correlation coefficient between them. Then*

$$\sqrt{e_1 e_2} - \sqrt{(1 - e_1)(1 - e_2)} \leq \rho \leq \sqrt{e_1 e_2} + \sqrt{(1 - e_1)(1 - e_2)}$$

Proof. Let T be minimum variance unbiased estimator of $\gamma(\theta)$. Then we are given:

$$E_\theta(T_1) = \gamma(\theta) = E_\theta(T_2) \ \forall\, \theta \in \Theta \quad \ldots(17\cdot18)$$

and

$$e_1 = \frac{V_\theta(T)}{V_\theta(T_1)} = \frac{V}{V_1}, \text{ (say)} \quad \Rightarrow \quad V_1 = \frac{V}{e_1} \quad \ldots(17\cdot19)$$

$$e_2 = \frac{V_\theta(T)}{V_\theta(T_2)} = \frac{V}{V_2}, \text{ (say)} \quad \Rightarrow \quad V_2 = \frac{V}{e_2} \quad \ldots(17\cdot20)$$

Let us consider another estimator : $\quad T_3 = \lambda T_1 + \mu T_2, \quad \ldots(17\cdot21)$

which is also unbiased estimator of $\gamma(\theta)$ *i.e.*,

$$E(T_3) = (\lambda + \mu)\,\gamma(\theta) = \gamma(\theta) \text{ [Using (17·18)]} \quad \Rightarrow \quad \lambda + \mu = 1 \qquad \ldots(17·22)$$

$$V_\theta(T_3) = V(\lambda T_1 + \mu T_2) = \lambda^2 V(T_1) + \mu^2 V(T_2) + 2\lambda\mu \,\text{Cov}(T_1, T_2)$$

$$= V\left(\frac{\lambda^2}{e_1} + \frac{\mu^2}{e_2} + 2\,.\frac{\lambda\mu\rho}{\sqrt{e_1 e_2}}\right) \qquad \text{[Using (17·19) and (17·20)]}$$

But $\quad V_\theta(T_3) \geq V$, since V is the minimum variance.

$$\therefore \qquad \frac{\lambda^2}{e_1} + \frac{\mu^2}{e_2} + \frac{2\rho\lambda\mu}{\sqrt{e_1 e_2}} \geq 1 = (\lambda + \mu)^2 \qquad \text{[Using (17·22)]}$$

$$\Rightarrow \qquad \left(\frac{1}{e_1} - 1\right)\lambda^2 + \left(\frac{1}{e_2} - 1\right)\mu^2 + 2\lambda\mu\left(\frac{\rho}{\sqrt{e_1 e_2}} - 1\right) \geq 0$$

$$\Rightarrow \qquad \left(\frac{1}{e_1} - 1\right)\left(\frac{\lambda}{\mu}\right)^2 + 2\left(\frac{\rho}{\sqrt{e_1 e_2}} - 1\right)\left(\frac{\lambda}{\mu}\right) + \left(\frac{1}{e_2} - 1\right) \geq 0, \qquad \ldots(17·23)$$

which is quadratic expression in (λ/μ).

Note that : $\quad e_i < 1 \;\Rightarrow\; \frac{1}{e_i} > 1 \;$ or $\; \left(\frac{1}{e_i} - 1\right) > 0, i = 1, 2$

We know that $AX^2 + BX + C \geq 0 \;\forall x, A > 0, C > 0$; if and only if

$$\text{Discriminant} = B^2 - 4AC \leq 0 \qquad \ldots(17·24)$$

Using (17·24), we get from (17·23) :

$$\left(\frac{\rho}{\sqrt{e_1 e_2}} - 1\right)^2 - \left(\frac{1}{e_1} - 1\right)\left(\frac{1}{e_2} - 1\right) \leq 0 \quad \Rightarrow \quad (\rho - \sqrt{e_1 e_2})^2 - (1 - e_1)(1 - e_2) \leq 0$$

$$\therefore \qquad \rho^2 - 2\sqrt{e_1 e_2}\,\rho + (e_1 + e_2 - 1) \leq 0$$

This implies that ρ lies between the roots of the equation :

$$\rho^2 - 2\sqrt{e_1 e_2}\,\rho + (e_1 + e_2 - 1) = 0,$$

which are given by : $\frac{1}{2}\left\{2\sqrt{e_1 e_2} \pm 2\sqrt{e_1 e_2 - (e_1 + e_2 - 1)}\right\} = \sqrt{e_1 e_2} \pm \sqrt{(e_1 - 1)(e_2 - 1)}$

Hence we have :

$$\sqrt{e_1 e_2} - \sqrt{(e_1 - 1)(e_2 - 1)} \leq \rho \leq \sqrt{e_1 e_2} + \sqrt{(e_1 - 1)(e_2 - 1)}$$

$$\Rightarrow \quad \sqrt{e_1 e_2} - \sqrt{(1 - e_1)(1 - e_2)} \leq \rho \leq \sqrt{e_1 e_2} + \sqrt{(1 - e_1)(1 - e_2)} \qquad \ldots(17·25)$$

Corollary. If we take $e_1 = 1$ and $e_2 = e$ in (17·25), we get $\sqrt{e} \leq \rho \leq \sqrt{e} \;\Rightarrow\; \rho = \sqrt{e}$.

This leads to the following important result, which we state in the form of a theorem.

Theorem 17·5. *If T_1 is an MVU estimator of $\gamma(\theta)$, $\theta \in \Theta$ and T_2 is any other unbiased estimator of $\gamma(\theta)$ with efficiency $e = e_\theta$, then the correlation coefficient between T_1 and T_2 is given by* $\rho = \sqrt{e}$, *i.e.,* $\rho_\theta = \sqrt{e_\theta}$, $\forall\, \theta \in \Theta$.

For an alternate proof, see Examples 17·9 and 17·10.

Theorem 17·6. *If T_1 is an MVUE of $\gamma(\theta)$ and T_2 is any other unbiased estimator of $\gamma(\theta)$ with efficiency $e < 1$, then no unbiased linear combination of T_1 and T_2 can be an MVUE of $\gamma(\theta)$.*

Proof. A linear combination : $\qquad T = l_1 T_1 + l_2 T_2 \qquad \ldots(17·27)$

$i = 1$

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 :*

$$L = g_\theta[t(x)].h(x) \qquad \ldots(17\cdot29)$$

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 $X = (X_1, X_2, \ldots, X_n)$, is always a sufficient statistic.

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

$$L = L(x, \theta) = g(x).h(\theta). \exp\{a(\theta)\psi(x)\} \qquad \ldots(17\cdot30)$$

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

$$L = \prod_{i=1}^{n} f(x_i, \theta) = g_1(t_1, \theta) . k(x_1, x_2, \ldots, x_n) \qquad \ldots(17\cdot31)$$

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

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

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

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

$\Rightarrow$ 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 $\gamma(\theta)$, a function of parameter θ, then*

$$Var(t) \geq \frac{\left\{\dfrac{d}{d\theta}\cdot \gamma(\theta)\right\}^2}{E\left(\dfrac{\partial}{\partial\theta}\log L\right)^2} = \frac{\{\gamma'(\theta)\}^2}{I(\theta)} \qquad \ldots(17{\cdot}32)$$

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 $\gamma(\theta)$.

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.

$= n \, . \, E\left\{ \frac{\partial}{\partial\theta} \log f(x, \theta) \right\}$ [On using ( )], since x_i's, $i = 1, 2, \ldots, n$ 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

$$[\gamma'(\theta)]^2 \leq E\,[t - \gamma(\theta)]^2 \,.\, E\left(\frac{\partial}{\partial\theta} \log L\right)^2 \qquad \ldots (17{\cdot}39)$$

The sign of equality will hold in C.R. inequality if only if the sign of equality holds in (17·39). The sign of equality will hold in (17·39) by Cauchy Schwartz Inequality, if and only if the variables $[t - \gamma(\theta)]$ and $\left(\frac{\partial}{\partial\theta} \log L\right)$ are proportional to each other, *i.e.*, lower bound.

17·6. 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.* (ii) *Method of Minimum Variance.*

(iii) *Method of Moments.* (iv) *Method of Least Squares.*

(v) *Method of Minimum Chi-square.* (vi) *Method of Inverse Probability.*

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

17·6·1. Method of Maximum Likelihood Estimation. From theoretical point 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 develoed 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 :

$$L = f(x_1, \theta)\, f(x_2, \theta) \ldots f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta). \qquad \ldots(17·53)$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(*iii*) For every θ in R,

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

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

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

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

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

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

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

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

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

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

Remark. Variance of M.L.E. is given by : $V(\hat{\theta}) = \frac{1}{I(\theta)} = \frac{1}{\left\{E\left(-\frac{\partial^2}{\partial\theta^2}\log L\right)\right\}}$ ...(17·5)

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

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

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

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

Differentiating *w.r. to* θ, we get : $\frac{\partial \log L}{\partial\theta} = \frac{\partial}{\partial\theta}\log g(t, \theta) = \psi(t, \theta)$, (say), ...(17·5)

which is a function of t and θ only.

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

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

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

Hence the theorem.

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

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

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

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

MLE for θ is the solution of the likelihood equation :

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

Important Note. It may be pointed out here that though

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The likelihood equation for estimating λ is :

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

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

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

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

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

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

Example 17·34. *Let $x_1, x_2, \ldots, x_n$ denote random sample of size n from a uniform population with p.d.f. :* $\quad f(x, \theta) = 1\,;\, \theta - \frac{1}{2} \le x \le \theta + \frac{1}{2}, -\infty < \theta < \infty$

Obtain M.L.E. for θ.

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

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

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

Thus L attains the maximum if

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

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

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

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

Example 17·35. *Find the M.L.E. of the parameters α and λ, (λ being large), of the distribution :* $\quad f(x; \alpha, \lambda) = \frac{1}{\Gamma(\lambda)}\left(\frac{\lambda}{\alpha}\right)^{\lambda} e^{-\lambda x/\alpha} x^{\lambda-1}\,;\, 0 \le x < \infty, \lambda > 0$

You may use that for large values of λ,

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

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

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

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

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

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

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

where G is independent of λ and α.

The likelihood equations for the simultaneous estimation of α and λ are :

$$\frac{\partial}{\partial \alpha} \log L = 0 \quad \ldots(1) \quad \text{and} \quad \frac{\partial}{\partial \lambda} \log L = 0 \quad \ldots(2)$$

(1) gives $\quad -\dfrac{n\lambda}{\alpha} + \dfrac{\lambda}{\alpha^2} . n\bar{x} = 0 \quad \Rightarrow \quad -1 + \dfrac{\bar{x}}{\alpha} = 0 \quad \Rightarrow \quad \hat{\alpha} = \bar{x}$ ...(3)

(2) gives (for large values of λ), on using (*) :

$$-n\left(\log \lambda - \frac{1}{2\lambda}\right) + n\left\{1.(\log \lambda - \log \alpha) + \lambda . \frac{1}{\lambda}\right\} - \frac{n\bar{x}}{\alpha} + n \log G = 0$$

$$\Rightarrow \quad \frac{1}{2\lambda} + \left(1 - \log \alpha + \log G - \frac{\bar{x}}{\alpha}\right) = 0$$

$$\Rightarrow \quad 1 + 2\lambda (\log G - \log \bar{x}) = 0 \qquad \text{[From (3)]}$$

$$\Rightarrow \quad 1 - 2\lambda \log\left(\frac{\bar{x}}{G}\right) = 0 \quad \Rightarrow \quad \hat{\lambda} = \frac{1}{2 \log (\bar{x}/G)}$$

Hence the M.L.E. for α and λ are given by : $\quad \hat{\alpha} = \bar{x} \quad$ and $\quad \hat{\lambda} = \dfrac{1}{2 \log (\bar{x}/G)}$.

Example 17·36. *In sampling from a power series distribution with p.d.f. :*

$$f(x, \theta) = a_x \theta^x / \psi(\theta) \; ; \; x = 0, 1, 2, \ldots$$

where a_x may be zero for some x, show that MLE of θ is a root of the equation :

$$\bar{X} = \frac{\theta \, \psi'(\theta)}{\psi(\theta)} = \mu(\theta), \text{ where } \mu(\theta) = E(X).$$

Solution. Likelihood function is given by :

$$L = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \left(\frac{a_{x_i} \theta^{x_i}}{\psi(\theta)}\right) = \left(\prod_{i=1}^{n} a_{x_i}\right) \frac{\theta^{\Sigma x_i}}{[\psi(\theta)]^n}$$

$$\Rightarrow \quad \log L = \sum_{i=1}^{n} \log a_{x_i} + \log \theta . \sum_{i=1}^{n} x_i - n \log \psi(\theta)$$

Likelihood equation for estimating θ gives :

$$\frac{\partial}{\partial \theta} \log L = 0 = \frac{\Sigma x_i}{\theta} - \frac{n\psi'(\theta)}{\psi(\theta)} \quad \Rightarrow \quad \bar{X} = \frac{\Sigma x_i}{n} = \frac{\theta \, \psi'(\theta)}{\psi(\theta)} = \mu(\theta), \text{ (say)}. \quad \ldots(i)$$

Hence MLE of θ is a root of equation (*). We have

$$E(X) = \sum_{x=0}^{\infty} x f(x, \theta) = \sum_{x=0}^{\infty} \left[x \left\{\frac{a_x \theta^x}{\psi(\theta)}\right\}\right] \quad \ldots(ii)$$

$$\sum_{x=0}^{\infty} f(x, \theta) = 1 \quad \Rightarrow \quad \sum_{x=0}^{\infty} \frac{a_x \theta^x}{\psi(\theta)} = 1 \quad \Rightarrow \quad \sum_{x=0}^{\infty} a_x \theta^x = \psi(\theta)$$

Differentiating *w.r. to* θ, we get

$$\sum_x [a_x . x\theta^{x-1}] = \psi'(\theta) \quad \Rightarrow \quad \sum_x \left\{a_x . \frac{x \theta^x}{\psi(\theta)}\right\} = \frac{\theta . \psi'(\theta)}{\psi(\theta)}$$

$$\therefore \quad E(X) = \mu(\theta) = \bar{X}, \qquad \text{[From (ii) and (i)]}$$

as required.

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

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

Obtain the maximum likelihood estimator for θ.

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

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

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

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

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

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

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

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

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

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

The likelihood equations for α and β give

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

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

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

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

$\therefore$ M.L.E. for α and β are given by :

$$\hat{\alpha} = x_{(1)} = \text{The smallest sample observation}$$

and $$\hat{\beta} = x_{(n)} = \text{The largest sample observation.}$$

Example 17·38. *State as precisely as possible the properties of the M.L.E. Obtain the M.L.Es. of α and β for a random sample from the exponential population :*

$$f(x\,;\alpha, \beta) = y_0\, e^{-\beta(x-\alpha)},\ \alpha \le x \le \infty,\ \beta > 0 \text{ and } y_0 \text{ being a constant.}$$

Solution. Here first of all we shall determine the constant y_0 from the consideration that the total area under a probability curve is unity.

$$\therefore\ y_0 \int_\alpha^\infty \exp[-\beta(x-\alpha)]\,dx \Rightarrow y_0 \left| \frac{e^{-\beta(x-\alpha)}}{-\beta} \right|_\alpha^\infty = 1 \Rightarrow -\frac{y_0}{\beta}(0-1) = 1 \Rightarrow y_0 = \beta$$

$$\therefore\ f(x\,;\alpha,\beta) = \beta\, e^{-\beta(x-\alpha)},\ \alpha \le x < \infty$$

If $x_1, x_2, \ldots, x_n$ is a random sample of n observations from this population, then

$$L = \prod_{i=1}^{n} f(x_i\,;\alpha,\beta) = \beta^n \exp\left\{-\beta \sum_{i=1}^{n}(x_i-\alpha)\right\} = \beta^n \exp\left[-n\,\beta(\bar{x}-\alpha)\right] \quad \ldots(*)$$

$$\therefore\ \log L = n \log \beta - n\beta(\bar{x}-\alpha)$$

The likelihood equations for estimating α and β give

$$\frac{\partial}{\partial \alpha} \log L = 0 = n\beta \quad \ldots(**)$$

and $$\frac{\partial}{\partial \beta} \log L = 0 = \frac{n}{\beta} - n(\bar{x}-\alpha) \quad \ldots(***)$$

Equation (**) gives $\beta = 0$, which is obviously inadmissible and this on substitution in (***) gives $\alpha = \infty$, a nugatory result. Thus the likelihood equations fail to give us valid estimates of α and β and we try to locate M.L.Es. for α and β by maximising L directly.

$$L \text{ is maximum} \Rightarrow \log L \text{ is maximum.}$$

From (*), $\log L$ is maximum (for any value of β), if $(\bar{x}-\alpha)$ is minimum, which is so if α is maximum.

If $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ is ordered sample from this population then $\alpha \le x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)} < \infty$, so that the maximum value of α consistent with the sample is $x_{(1)}$, the smallest sample observation, *i.e.*, $\hat{\alpha} = x_{(1)}$.

Consequently, (***) gives $\frac{1}{\beta} = \bar{x} - \hat{\alpha} = \bar{x} - x_{(1)} \Rightarrow \hat{\beta} = \frac{1}{\bar{x} - x_{(1)}}$

Hence M.L.Es. for α and β are given by : $\hat{\alpha} = x_{(1)}$ and $\hat{\beta} = \frac{1}{\bar{x} - x_{(1)}}$

Remarks **1.** Whenever the given probability function involves a constant and the range of the variable is dependent on the parameter(s) to be estimated, first of all we should determine the constant by taking the total probability as unity and then proceed with the estimation part.

2. From the last two examples, it is obvious that whenever the range of the variable involves the parameter(s) to be estimated, the likelihood equations fail to give us valid estimates and in this case M.L.Es are obtained by adopting some other approach of maximising L or $\log L$ directly.

Example 17·39. *Obtain maximum likelihood estimate of θ in $f(x, \theta) = (1 + \theta)\, x^{\theta}$, $0 < x < 1$, based on an independent sample of size n. Examine whether this estimate is sufficient for θ.*

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

$$\Rightarrow \qquad \log L = n \log(1+\theta) + \theta.\sum_{i=1}^{n} \log x_i$$

$$\frac{\partial}{\partial\theta}\log L = \frac{n}{1+\theta} + \sum_{i=1}^{n} \log x_i = 0 \quad \Rightarrow \quad n + \theta \sum_i \log x_i + \sum_i \log x_i = 0$$

$$\therefore \qquad \hat{\theta} = \frac{-n}{\sum_{i=1}^{n} \log x_i} - 1 = \frac{-n}{\log\left(\prod_{i=1}^{n} x_i\right)} - 1 \qquad \ldots(*)$$

Also
$$L(x, \theta) = \left\{(1+\theta)^n . \left(\prod_{i=1}^{n} x_i\right)^{\theta-1}\right\} . \left(\prod_{i=1}^{n} x_i\right)$$

Hence by Factorisation theorem, $T = \left(\prod_{i=1}^{n} x_i\right)$ is a sufficient statistic for θ, and $\hat{\theta}$ being a one to one function of sufficient statistic $\left(\prod_{i=1}^{n} x_i\right)$, is also sufficient for θ.

Example 17·40. (a) *Obtain the most general form of distribution differentiable in θ, for which the sample mean is the M.L.E.*

(b) *Show that the most general continuous distribution for which the M.L.E. of a parameter θ is the geometric mean of the sample is*

$$f(x, \theta) = \left(\frac{x}{\theta}\right)^{\theta \cdot \frac{\partial\psi}{\partial\theta}} \exp\{\psi(\theta) + \xi(x)\},$$

where $\psi(\theta)$ and $\xi(x)$ are arbitrary functions of θ and x respectively.

Solution. (a) We have $L = \prod_{i=1}^{n} f(x_i, \theta) \Rightarrow \log L = \sum_{i=1}^{n} \log f(x_i, \theta) = \sum_{x} \log f$, $[f = f(x, \theta)]$

the summation extending to all the values of $x = (x_1, x_2, \ldots, x_n)$ in the sample. The likelihood equation is : $\frac{\partial}{\partial\theta}\log L = 0$, *i.e.*, $\frac{\partial}{\partial\theta}(\sum_x (\log f) = 0$

$$\Rightarrow \qquad \sum_x \frac{\partial}{\partial\theta}\log f = 0 \quad \Rightarrow \quad \sum_x \frac{1}{f} \cdot \frac{\partial f}{\partial\theta} = 0 \qquad \ldots(*)$$

We are given that the solution of (*) is : $\theta = \frac{1}{n}\Sigma x \Rightarrow n\theta = \Sigma x \Rightarrow \sum_x (x - \theta) = 0 \ldots(**)$

Since this is true for all values of x and θ, we get from (*) and (**),

$\frac{1}{f} \cdot \frac{\partial f}{\partial\theta} = A(x - \theta)$, where A is independent of x but may be function of θ.

Let us take $A = \dfrac{\partial^2 \psi}{\partial \theta^2}$, where $\psi = \psi(\theta)$ is any arbitrary function of θ. Thus

$$\frac{\partial}{\partial \theta} \log f = \frac{\partial^2 \psi}{\partial \theta^2} (x - \theta).$$

Integrating *w.r. to.* θ (partially), we get

$$\log f = (x - \theta) . \frac{\partial \psi}{\partial \theta} - \int \frac{\partial \psi}{\partial \theta} (-1) \, d\theta + \xi(x) + k,$$

where $\xi(x)$ is an arbitrary function of x and k is arbitrary constant.

$$\therefore \qquad \log f = (x - \theta) . \frac{\partial \psi}{\partial \theta} + \psi(\theta) + \xi(x) + k$$

Hence
$$f = \text{const. exp} \left\{ (x - \theta) \frac{\partial \psi}{\partial \theta} + \psi(\theta) + \xi(x) \right\},$$

which is the probability function of the required distribution.

Remark. In particular, if we take. $\psi(\theta) = \dfrac{\theta^2}{2}$ and $\xi(x) = -\dfrac{x^2}{2}$, then

$$f = \text{Const. exp} \left\{ (x - \theta) . \theta + \frac{\theta^2}{2} - \frac{x^2}{2} \right\}$$

$$= \text{Const. exp} \left\{ -\frac{1}{2} (x^2 + \theta^2 - 2\theta x) \right\} = \text{Const. exp} \left\{ -\frac{1}{2} (x - \theta)^2 \right\}$$

which is the probability function of the normal distribution with mean θ and unit variance.

(*b*) Here the solution of the likelihood equation

$$\frac{\partial}{\partial \theta} \log L = \sum_x \frac{\partial}{\partial \theta} \log f = 0 \qquad \ldots(*)$$

is $\theta = (x_1, x_2, \ldots, x_n)^{1/n}$ $\Rightarrow$ $\log \theta = \dfrac{1}{n} \sum\limits_x \log x$ or $\sum\limits_x (\log x - \log \theta) = 0$ ...(**)

Since this is true for all x and all θ, we get from (*) and (**),

$$\frac{\partial}{\partial \theta} \log f = (\log x - \log \theta) \, A(\theta),$$

where $A(\theta)$ is an arbitrary function of θ and is independent of x.

Integrating *w.r. to* θ (partially),

$$\log f = \log x \int A(\theta) \, d\theta - \int A(\theta) \log \theta \, d\theta + \xi(x),$$

where $\xi(x)$ is an arbitrary function of x alone.

If we take $\int A(\theta) \, d\theta = A_1(\theta)$, then

$$\log f = \log x . A_1(\theta) - \left\{ A_1(\theta) \log \theta - \int A_1(\theta) . \frac{1}{\theta} d\theta \right\} + \xi(x)$$

$$= A_1(\theta) \log (x/\theta) + \int \frac{A_1(\theta)}{\theta} d\theta + \xi(x)$$

Let us take $A_1(\theta) = \theta \dfrac{\partial \psi}{\partial \theta}$. (suggested by the answer), where $\psi = \psi(\theta)$ is an arbitrary function of θ alone.

$$\therefore \quad \log f = \theta \frac{\partial \psi}{\partial \theta} \log (x/\theta) + \int \frac{\partial \psi}{\partial \theta} d\theta + \xi(x)$$

$$= \theta \frac{\partial \psi}{\partial \theta} \cdot \log (x/\theta) + \psi(\theta) + \xi(x) = \log \left[\left(\frac{x}{\theta} \right)^{\theta \frac{\partial \psi}{\partial \theta}} \right] + \psi(\theta) + \xi(x)$$

Hence $$f = f(x, \theta) = \left(\frac{x}{\theta} \right)^{\theta \frac{\partial \psi}{\partial \theta}} \cdot \exp \{\psi(\theta) + \xi(x)\}.$$

Example 17·41. *A sample of size n is drawn from each of the four normal populations which has the same variance σ^2. The means of the four populations are $a + b + c$, $a + b - c$, $a - b + c$ and $a - b - c$. What are the M.L.Es. for a, b, c, and σ^2 ?*

Solution. Let the sample observations be denoted by x_{ij}, $i = 1, 2, 3, 4; j = 1, 2, \ldots, n$. Since the four samples, from the four normal populations are independent, the likelihood function L of all the sample observations x_{ij}, $(i = 1, 2, 3, 4, ; j = 1, 2, \ldots, n)$,

is given by : $$L = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^{4n} \cdot \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{4} \sum_{j=1}^{n} (x_{ij} - \mu_i)^2 \right\},$$

where μ_i, $(i = 1, 2, 3, 4)$ is mean of the ith population. Therefore

$$L = \left(\frac{1}{\sqrt{2\pi}\,\sigma} \right)^{4n} \cdot \exp \left[-\frac{1}{2\sigma^2} \left\{ \sum_j (x_{1j} - \mu_1)^2) + \sum_j (x_{2j} - \mu_2)^2 + \sum_j (x_{3j} - \mu_3)^2 + \sum_j (x_{4j} - \mu_4)^2 \right\} \right]$$

$$\Rightarrow \log L = k - 2n \log \sigma^2 - \frac{1}{2\sigma^2} \left\{ \sum_j (x_{1j} - a - b - c)^2 + \sum_j (x_{2j} - a - b + c)^2 \right.$$

$$\left. + \sum_j (x_{3j} - a + b - c)^2 + \sum_j (x_{4j} - a + b + c)^2 \right\},$$

where k is a constant w.r. to a, b, c and σ^2. The M.L.Es. for a, b, c and σ^2 are the solutions of the simultaneous equations (maximum likelihood equations for estimating a, b, c and σ^2) :

$$\frac{\partial}{\partial a} \log L = 0 \quad \ldots(1) \qquad \frac{\partial}{\partial b} \log L = 0 \quad \ldots(2)$$

$$\frac{\partial}{\partial c} \log L = 0 \quad \ldots(3) \qquad \frac{\partial}{\partial \sigma^2} \log L = 0 \quad \ldots(4)$$

(1) gives : $$-\frac{1}{2\sigma^2} \left\{ \sum_j (x_{1j} - a - b - c)(-2) + \sum_j (x_{2j} - a - b + c)(-2) \right.$$

$$\left. + \sum_j (x_{3j} - a + b - c)(-2) + \Sigma (x_{4j} - a + b + c)(-2) \right\} = 0$$

$$\Rightarrow \sum_j (x_{1j} + x_{2j} + x_{3j} + x_{4j}) + n \left[(-a - b - c) + (-a - b + c) + (-a + b - c) + (-a + b + c) \right] = 0$$

$$\Rightarrow \sum_{j=1}^{n} \left(\sum_{i=1}^{4} x_{ij} \right) + n(-4a) = 0 \qquad \therefore \quad \hat{a} = \frac{1}{4n} \sum_{i=1}^{4} \sum_{j=1}^{n} x_{ij} = \bar{x}$$

(2) gives : $$-\frac{1}{2\sigma^2} \left\{ \sum_j (x_{1j} - a - b - c)(-2) + \sum_j (x_{2j} - a - b + c)(-2) \right.$$

$$\left. + \sum_j (x_{3j} - a + b - c)(2) + \sum_j (x_{4j} - a + b + c)(2) \right\} = 0$$

$$\Rightarrow \quad \sum_j x_{1j} + \sum_j x_{2j} - \sum_j x_{3j} - \sum_j x_{4j} + n\left[(-a-b-c)+(-a-b+c)-(-a+b-c)-(-a+b+c)\right]=0$$

$$\Rightarrow \quad \sum x_{1j} + \sum x_{2j} - \sum x_{3j} - \sum x_{4j} - 4nb = 0$$

$$\therefore \quad \hat{b} = \frac{1}{4}\left(\frac{1}{n}\sum x_{1j} + \frac{1}{n}\sum x_{2j} - \frac{1}{n}\sum x_{3j} - \frac{1}{n}\sum x_{4j}\right) \Rightarrow \hat{b} = (\bar{x}_1 + \bar{x}_2 - \bar{x}_3 - \bar{x}_4)/4$$

where $\bar{x}_i$ is the mean of the ith sample.

Similarly (3) will give : $\hat{c} = (\bar{x}_1 - \bar{x}_2 + \bar{x}_3 - \bar{x}_4)/4$

(4) gives :
$$-\frac{2n}{\sigma^2} + \frac{1}{2\sigma^4}\left\{\sum_j (x_{1j}-a-b-c)^2 + \sum_j (x_{2j}-a-b+c)^2 + \sum_j (x_{3j}-a+b-c)^2 + \sum_j (x_{4j}-a+b+c)^2\right\} = 0$$

$$\therefore \quad \hat{\sigma}^2 = \frac{1}{4n}\left\{\sum_j (x_{1j}-\hat{a}-\hat{b}-\hat{c})^2 + \sum_j (x_{2j}-\hat{a}-\hat{b}+\hat{c})^2 + \sum_j (x_{3j}-\hat{a}+\hat{b}-\hat{c})^2 + \sum_j (x_{4j}-\hat{a}+\hat{b}+\hat{c})^2\right\}$$

Example 17·42. *The following table gives probabilities and observed frequencies in four classes AB Ab, aB and ab in a genetical experiment. Estimate the parameter θ by the method of maximum likelihood and find its standard error.*

Class	*Probability*	*Observed frequency*
AB	$\frac{1}{4}(2+\theta)$	108
Ab	$\frac{1}{4}(1-\theta)$	27
aB	$\frac{1}{4}1-\theta)$	30
ab	$\frac{1}{4}\theta$	8

Solution. Using multinomial probability law, we have

$$L = L(x, \theta) = \frac{n!}{n_1!\,n_2!\,n_3!\,n_4!}\, p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4}, \quad \sum p_i = 1, \quad \sum n_i = n$$

$$\Rightarrow \quad \log L = C + n_1 \log p_1 + n_2 \log p_2 + n_3 \log p_3 + n_4 \log p_4,$$

where $C = \log\left[\dfrac{n!}{n_1!\,n_2!\,n_3!\,n_4!}\right]$, is a constant.

$$\therefore \quad \log L = C + n_1 \log(2+\theta/4) + n_2 \log(1-\theta/4) + n_3 \log(1-\theta/4) + n_4 \log(\theta/4) \quad \ldots(*)$$

Likelihood equation gives :

$$\frac{\partial \log L}{\partial \theta} = \frac{n_1}{2+\theta} - \frac{n_2}{1-\theta} - \frac{n_3}{1-\theta} + \frac{n_4}{\theta} = 0$$

$$\Rightarrow \quad \frac{n_1}{2+\theta} - \frac{(n_2+n_3)}{1-\theta} + \frac{n_4}{\theta} = 0$$

Taking $n_1 = 108$, $n_2 = 27$, $n_3 = 30$ and $n_4 = 8$, we get $\dfrac{108}{2+\theta} - \dfrac{(27+30)}{1-\theta} + \dfrac{8}{\theta} = 0$

$$\Rightarrow \quad 108\theta(1-\theta) - 57\theta(2+\theta) + 8(1-\theta)(2+\theta) = 0 \quad \Rightarrow \quad 173\,\theta^2 + 14\theta - 16 = 0$$

$$\therefore \qquad \theta = \frac{-14 \pm \sqrt{196 + 11072}}{346} = -0.34 \text{ and } 0.26$$

But θ, being the probability cannot be negative. Hence,

$$\text{M.L.E. of } \theta \text{ is given by } \hat{\theta} = 0.26 \qquad \ldots(**)$$

Differentiating (*) again partially w.r.to. θ, we get

$$\frac{\partial^2 \log L}{\partial \theta^2} = \frac{-n_1}{(2+\theta)^2} - \frac{(n_2+n_3)}{(1-\theta)^2} - \frac{n_4}{\theta^2}$$

$$-E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right) = \frac{E(n_1)}{(2+\theta)^2} + \frac{E(n_2)+E(n_3)}{(1-\theta)^2} + \frac{E(n_4)}{\theta^2}$$

$$= \frac{np_1}{(2+\theta)^2} + \frac{n(p_2+p_3)}{(1-\theta)^2} + \frac{np_4}{\theta^2} = \frac{n(2+\theta)}{4(2+\theta)^2} + \frac{n(1-\theta)}{2(1-\theta)^2} + \frac{n\theta}{4\theta^2}$$

$$\therefore \qquad I(\theta) = \frac{n}{4(2+\hat{\theta})} + \frac{n}{2(1-\hat{\theta})} + \frac{n}{4\hat{\theta}}; \qquad n = \Sigma n_i = 173.$$

$$= 173\left(\frac{1}{4 \times 2.26} + \frac{1}{2 \times 0.74} + \frac{1}{4 \times 0.26}\right) = 301.02$$

$$\text{S.E.}(\hat{\theta}) = \sqrt{1/I(\theta)} = \frac{1}{\sqrt{301.02}} = 0.0576 \qquad [\textit{c.f.}\ (17.55), \text{Theorem } 17.13)$$

17·6·2. Method of Minimum Variance. (Minimum Variance Unbiased Estimates (M.V.U.E.)}. In this section we shall look for estimates which (*i*) are unbiased and (*ii*) have minimum variance.

If $L = \prod_{i=1}^{n} f(x_i, \theta)$, is the likelihood function of a random sample of n observations $x_1, x_2, \ldots, x_n$ from a population with probability function $f(x, \theta)$, then the problem is to find a statistic $t = t(x_1, x_2, \ldots, x_n)$, such that

$$E(t) = \int_{-\infty}^{\infty} t.L\,dx = \gamma(\theta) \quad \Rightarrow \quad \int_{-\infty}^{\infty} \{t - \gamma(\theta)\}\, L\,dx = 0 \qquad \ldots(17.57)$$

and
$$V(t) = \int_{-\infty}^{\infty} [t - E(t)]^2\, L\,dx = \int_{-\infty}^{\infty} [t - \gamma(\theta)]^2\, L\,dx \qquad \ldots(17.58)$$

is minimum where

$$\int_{-\infty}^{\infty} dx \text{ represents the } n\text{-fold integration } \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dx_1\, dx_2 \ldots dx_n$$

In other words, we have to minimise (17·58) subject to the condition (17·57).

For detailed discussion of this method see *MVU* Estimators (§ 17·5·2) and Cramer-Rao Inequality (§ 17·7).

17·6·3. Method of Moments. This method was discovered and studied in detail by Karl Pearson.

Let $f(x\,;\, \theta_1, \theta_2, \ldots, \theta_k)$ be the density function of the parent population with k parameters $\theta_1, \theta_2, \ldots, \theta_k$. If μ'_r denotes the rth moment about origin, then

$$\mu_r' = \int_{-\infty}^{\infty} x^r f(x ; \theta_1, \theta_2, \ldots, \theta_k)\, dx, \ (r = 1, 2, \ldots, k) \quad \ldots(17{\cdot}59)$$

In general $\mu_1', \mu_2, \ldots, \mu_k'$ will be function of the parameters $\theta_1, \theta_2, \ldots, \theta_k$.

Let x_i, $i = 1, 2, \ldots, n$ be a random sample of size n from the given population. The method of moments consists in solving the k-equations (17·59) for $\theta_1, \theta_2, \ldots, \theta_k$ in terms of $\mu_1', \mu_2', \ldots, \mu_k'$ and then replacing these moments μ_r' ; $r = 1, 2, \ldots, k$ by the sample moments, *e.g.*, $\hat{\theta}_i = \theta_i (\hat{\mu}_1', \hat{\mu}_2', \ldots, \hat{\mu}_k') = \theta_i (m_1', m_2', \ldots, m_k')$; $i = 1, 2, \ldots, k$, where m_i' is the ith moment about origin in the sample.

Then by the method of moments $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ are the required estimators of $\theta_1, \theta_2, \ldots, \theta_k$ respectively.

Remarks.1. Let $(x_1, x_2, \ldots, x_n)$ be a random sample of size n from a population with p.d.f. $f(x, \theta)$. Then X_i, $(i = 1, 2, \ldots, n)$ are *i.i.d.* $\Rightarrow X_i^r$, $(i = 1, 2, \ldots, n)$ are *i.i.d.* Hence if $E(X_i^r)$ exists, then by W.L.L.N., we get

$$\frac{1}{n}\sum_{i=1}^{n} x_i^r \xrightarrow{p} E(X_1^r) \quad \Rightarrow \quad m_r' \xrightarrow{p} \mu_r' \quad \ldots(17{\cdot}60)$$

Hence the sample moments are consistent estimators of the corresponding population moments.

2. It has been shown that under quite general conditions, the estimates obtained by the method of moments are asymptotically normal but not, in general, efficient.

3. Generally the method of moments yields less efficient estimators than those obtained from the principle of maximum likelihood. The estimators obtained by the method of moments are identical with those given by the method of maximum likelihood if the probability mass function or probability density function is of the form : $\quad \ldots(17{\cdot}61)$

$$f(x, \theta) = \exp\,(b_0 + b_1 x + b_2 x^2 + \ldots)$$

where b's are independent of x but may depend on $\theta = (\theta_1, \theta_2, \ldots)$.

(17·61) implies that : $L(x_1, x_2, \ldots, x_n ; \theta) = \exp\,(nb_0 + b_1 \Sigma x_i + b_2 \Sigma x_i^2 + \ldots)$

$$\Rightarrow \quad \frac{\partial}{\partial\theta}\log L = a_0 + a_1 \Sigma x_i + a_2 \Sigma x_i^2 + a_3 \Sigma x_i^3 + \ldots \quad \ldots(17{\cdot}61a)$$

where $\quad a_i = \dfrac{\partial}{\partial\theta}(b_i), \ (i = 1, 2, \ldots\ldots) \quad$ and $\quad a_0 = n\dfrac{\partial b_0}{\partial\theta}$

Thus both the methods yield identical estimators if *MLE*'s are obtained as linear functions of the moments.

Example 17·43. *Estimate α and β in the case of Pearson's Type III distribution by the method of moments :*

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, 0 \le x < \infty$$

Solution. We have

$$\mu_r' = \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty x^r x^{\alpha-1} e^{-\beta x}\, dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+r)}{\beta^{\alpha+r}} = \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)\,\beta^r}$$

$$\mu_1' = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha).\beta} = \frac{\alpha}{\beta}, \qquad \mu_2' = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\,\beta^2} = \frac{(\alpha+1)\,\alpha}{\beta^2}$$

$$\therefore \quad \frac{\mu_2'}{\mu_1'^2} = \frac{\alpha+1}{\alpha} = \frac{1}{\alpha} + 1 \quad \Rightarrow \quad \alpha = \frac{\mu_1'^2}{\mu_2' - \mu_1'^2}, \qquad \beta = \frac{\alpha}{\mu_1'} = \frac{\mu_1'}{\mu_2' - \mu_1'^2}$$

Hence $\hat{\alpha} = \dfrac{m_1'^2}{m_2' - m_1'^2}$ and $\hat{\beta} = \dfrac{m_1'}{m_2' - m_1'^2}$, where m_1' and m_2' are sample moments.

Example 15·44. *For the double Poisson distribution :*

$$p(x) = P(X = x) = \frac{1}{2} \cdot \frac{e^{-m_1} . m_1^x}{x\,!} + \frac{1}{2} \cdot \frac{e^{-m_2} . m_2^x}{x\,!}\ ; x = 0, 1, 2, \ldots$$

show that the estimates for m_1 and m_2 by the method of moments are : $\mu_1' \pm \sqrt{\mu_2' - \mu_1' - \mu_1'^2}$.

Solution. We have

$$\mu_1' = \sum_{x=0}^{\infty} x . p(x) = \frac{1}{2} \sum_{x=0}^{\infty} x . \frac{e^{-m_1} m_1^x}{x\,!} + \frac{1}{2} \sum_{x=0}^{\infty} x . \frac{e^{-m_2} m_2^x}{x\,!} = \frac{1}{2} m_1 + \frac{1}{2} m_2 \qquad \ldots(*)$$

(since the first and second summations are the means of Poisson distributions with parameters m_1 and m_2 respectively).

$$\mu_2' = \sum_{x=0}^{\infty} x^2 . p(x) = \frac{1}{2} \left\{ \sum_{x=0}^{\infty} x^2 . \left(\frac{e^{-m_1} m_1^x}{x\,!} \right) + \sum_{x=0}^{\infty} x^2 . \left(\frac{e^{-m_2} m_2^x}{x\,!} \right) \right\}$$

$$= \frac{1}{2} \{ (m_1^2 + m_1) + (m_2^2 + m_2) \}$$

$$\Rightarrow \quad \mu_2' = \frac{1}{2} \{ (m_1 + m_2) + (m_1^2 + m_2^2) \}. \qquad \ldots(**)$$

$$= \frac{1}{2} \{ 2\mu_1' + m_1^2 + (2\mu_1' - m_1)^2 \} \qquad \text{[Using (*)]}$$

$$= \frac{1}{2} (2\mu_1' + m_1^2 + 4\mu_1'^2 + m_1^2 - 4m_1 \mu_1')$$

$$\Rightarrow \quad \mu_2' = \mu_1' + m_1^2 + 2\mu_1'^2 - 2\mu_1' m_1 \quad \Rightarrow \quad m_1^2 - 2m_1\mu_1' + (2\mu_1'^2 + \mu_1' - \mu_2') = 0$$

$$\therefore \quad \hat{m}_1 = \frac{2\mu_1' \pm \sqrt{4\mu_1'^2 - 4(2\mu_1'^2 + \mu_1' - \mu_2')}}{2} = \mu_1' \pm \sqrt{\mu_2' - \mu_1' - \mu_1'^2}$$

Similarly on substituting for m_1 in terms of m_2 from (*) in (**), we get

$$m_2^2 - 2m_2\mu_1' + (2\mu_1'^2 + \mu_1' - \mu_2') = 0$$

Solving for m_2, we get $\quad \hat{m}_2 = \mu_1' \pm \sqrt{\mu_2' - \mu_1' - \mu_1'^2}$

Example 17·45. *A random variable X takes the values, 0, 1, 2, with respective probabilities* $\dfrac{\theta}{4N} + \dfrac{1}{2}\left(1 - \dfrac{\theta}{N}\right), \dfrac{\theta}{2N} + \dfrac{\alpha}{2}\left(1 - \dfrac{\theta}{N}\right)$ *and* $\dfrac{\theta}{4N} + \dfrac{1-\alpha}{2}\left(1 - \dfrac{\theta}{N}\right)$,

where N is a known number and α, θ are unknown parameters. If 75 independent observations on X yielded the values 0, 1, 2 with frequencies 27, 38, 10 respectively, estimate θ and α by the method of moments.

Solution.

$$E(X) = 0 . \left\{ \frac{\theta}{4N} + \frac{1}{2}\left(1 - \frac{\theta}{N}\right) \right\} + 1 . \left\{ \frac{\theta}{2N} + \frac{\alpha}{2}\left(1 - \frac{\theta}{N}\right) \right\} + 2 \left\{ \frac{\theta}{4N} + \frac{1-\alpha}{2}\left(1 - \frac{\theta}{N}\right) \right\}$$

$$= \frac{\theta}{N} + \left(1 - \frac{\theta}{N}\right) \left[\frac{\alpha}{2} + (1 - \alpha) \right]$$

$$\Rightarrow \quad \mu_1' = \frac{\theta}{N} + \left(1 - \frac{\theta}{N}\right)\left(1 - \frac{\alpha}{2}\right) = 1 - \frac{\alpha}{2}\left(1 - \frac{\theta}{N}\right) \qquad \ldots(*)$$

$$E(X^2) = 1^2 \cdot \left\{\frac{\theta}{2N} + \frac{\alpha}{2}\left(1 - \frac{\theta}{N}\right)\right\} + 2^2 \cdot \left\{\frac{\theta}{4N} + \frac{1-\alpha}{2}\left(1 - \frac{\theta}{N}\right)\right\}$$

$$= \frac{3\theta}{2N} + \left(1 - \frac{\theta}{N}\right)\left[\frac{\alpha}{2} + 2(1-\alpha)\right] = \frac{3\theta}{2N} + \left(1 - \frac{\theta}{N}\right)\left(2 - \frac{3\alpha}{2}\right)$$

$$\Rightarrow \quad \mu_2' = 2 - \frac{\theta}{2N} - \frac{3}{2}\alpha\left(1 - \frac{\theta}{N}\right) \qquad \ldots(**)$$

The sample frequency distribution is :

x	0	1	2
f	27	38	10

$$\mu_1' = \frac{1}{N}\Sigma fx = \frac{1}{75}(38 + 20) = \frac{58}{75}, \qquad \mu_2' = \frac{1}{N}\Sigma fx^2 = \frac{1}{75}(38 + 40) = \frac{78}{75}$$

Equating the sample moments to theoretical moments, we get

$$1 - \frac{\alpha}{2}\left(1 - \frac{\theta}{N}\right) = \frac{58}{75} \quad \Rightarrow \quad \frac{\alpha}{2}\left(1 - \frac{\theta}{N}\right) = 1 - \frac{58}{75} = \frac{17}{75} \qquad \ldots(***)$$

Substituting in (**), we get $\quad 2 - \frac{\theta}{2N} - 3 \times \frac{17}{75} = \frac{78}{75} \quad \Rightarrow \quad \hat{\theta} = \frac{42}{75}N$

Substituting in (***), we get $\quad \frac{\alpha}{2}\left(1 - \frac{42}{75}\right) = \frac{17}{75} \quad \Rightarrow \quad \hat{\alpha} = \frac{34}{33}$

17·6·4. Method of Least Squares. The principle of least squares is used to fit a curve of the form :

$$y = f(x, a_0, a_1, \ldots, a_n) \qquad \ldots(17{\cdot}62)$$

where a_i's are unknown parameters, to a set of n sample observations (x_i, y_i); $i = 1, 2, \ldots, n$ from a bivariate population. It consists in minimising the sum of squares of residuals, *viz.*,

$$E = \sum_{i=1}^{n} \{y_i - f(x_i, a_0, a_1, \ldots, a_n)\}^2 \qquad \ldots(17{\cdot}63)$$

subject to variations in $a_0, a_1, \ldots, a_n$.

The normal equations for estimating $a_0, a_1, \ldots, a_n$ are given by :

$$\frac{\partial E}{\partial a_i} = 0; \; i = 1, 2, \ldots, n \qquad \ldots(17{\cdot}64)$$

Remarks. **1.** In chapter 10, we have discussed in detail the method of least squares for fitting linear regression, polynomial regression and the exponential family of curves reducible to linear regression. In chapter 11, we have discussed the method of fitting multiple linear regression (§ 11·12·1).

2. If we are estimating $f(x, a_0, a_1, \ldots, a_n)$ as a linear function of the parameters $a_0, a_1, \ldots, a_n$, the x's being known given values, the least square estimators obtained as linear functions of the y's will be MVU estimators.

17·7. CONFIDENCE INTERVAL AND CONFIDENCE LIMITS

Let x_i, $(i = 1, 2, \ldots, n)$ be a random sample of n observations from a population involving a single unknown parameter θ, (say). Let $f(x, \theta)$ be the probability function