For the rectangular population (*), the *p.d.f.* of nth order statistic (the la sample observation), Y_n is : $g(y) = n \cdot \{F(y, \theta)\}^{n-1} \cdot f(y, \theta)$,

where $$F(x, \theta) = P(X \le x) = \int_0^x f(u)du = \int_0^x \frac{1}{\theta} = \frac{x}{\theta}$$

$\therefore$ $$g(y) = n\left(\frac{y}{\theta}\right)^{n-1}\frac{1}{\theta} = \frac{n}{\theta^n} \cdot y^{n-1} ; 0 \le y < \theta$$

$$E(Y_n^r) = \int_0^\theta y^r \cdot g(y)dy = \frac{n}{\theta^n}\int_0^\theta y^{r+n-1}\, dy = \frac{n\theta^r}{n+r}$$

Taking $r = 1$ and 2; $\quad E(Y_n) = \dfrac{n\theta}{n+1}$; $E(Y_n^2) = \dfrac{n\theta^2}{n+2}$

Now $\quad E\left(\dfrac{n+1}{n} \cdot Y_n\right) = \dfrac{n+1}{n} E(Y_n) = \theta$ ...(

$\Rightarrow$ $(n+1)Y_n/n$ is an unbiased estimator of θ. [Using

$$\text{Var}\left(\frac{n+1}{n}Y_n\right) = \left(\frac{n+1}{n}\right)^2 \cdot \text{Var}(Y_n) = \left(\frac{n+1}{n}\right)^2 \{EY_n^2 - (EY_n)^2\}$$

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

$\Rightarrow$ $\text{Var}\left(\dfrac{n+1}{n} \cdot Y_n\right) \le 1 \Big/ \left\{n\,E\left(\dfrac{\partial}{\partial\theta}\log f\right)^2\right\}$. Hence $(n+1)Y_n/n$ is an *MVUE.*

Remark. This example illustrates that if the regularity conditions underlying Cramer-R inequality are violated, then the least attainable variance may be less than the Cramer-R lower bound.

17·6. METHODS OF ESTIMATION

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

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

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

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

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

$$L = f(x_1, \theta)\, f(x_2, \theta) \ldots f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta). \quad \ldots(17·5$$

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

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

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

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

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 :

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

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 :

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

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 $\left(\dfrac{\partial^2}{\partial\theta_i\,\partial\theta_j}\log L\right)_{\theta = \hat{\theta}}$ 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.,* $\dfrac{\partial \log L}{\partial \theta}$ and $\dfrac{\partial^2 \log L}{\partial \theta^2}$ exist and are continuous functions of θ in a range R (including the true value θ_0 of the parameter) for almost all x. For every θ in R, $\left|\dfrac{\partial}{\partial\theta}\log L\right| < F_1(x)$ and $\left|\dfrac{\partial^2}{\partial\theta^2}\log L\right| < F_2(x)$ where $F_1(x)$ and $F_2(x)$ are integrable functions over $(-\infty, \infty)$.

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

(*iii*) For every θ in R,

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

(*iv*) The range of integration is independent of θ. But if the range of integ 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 prop which will be stated in the form of theorems.

Theorem 17·11. (Cramer-Rao Theorem). *"With probability approaching unity a ∞, the likelihood equation $\frac{\partial}{\partial\theta}\log L = 0$, has a solution which converges in probability 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 exam sampling from $N(\mu, \sigma^2)$ population, [c.f. Example 17·31],

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

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

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

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

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

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

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

Proof. If $t = t(x_1, x_2, \ldots, x_n)$ is a sufficient estimator of θ, then Likelihood Func 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 $g(t$ the density function of t and θ and $h(x_1, x_2, \ldots, x_n \mid t)$ is the density function o 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

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 necessa[rily 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 paramete[r α of a] population having density function :* $\frac{2}{\alpha^2}(\alpha - x),\ 0 < x < \alpha$, *for a sample of unit siz[e,] 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 i[s]

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

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 : $\quad \hat{\alpha} = 2x.$

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

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 paramete[r 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 [the] 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 b[y]

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

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

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

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

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

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

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

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

Obtain M.L.E. for θ.

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

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

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

Thus L attains the maximum if

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

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

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

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

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

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

You may use that for large values of λ,

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

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

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

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

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 $\quad -\frac{n\lambda}{\alpha}+\frac{\lambda}{\alpha^2}\cdot n\bar{x} = 0 \quad \Rightarrow \quad -1+\frac{\bar{x}}{\alpha}=0 \quad \Rightarrow \quad \hat{\alpha}=\bar{x}$

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

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

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

$$\Rightarrow \quad 1 + 2\lambda\,(\log G - \log\bar{x}) = 0$$

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

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

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 :

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

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

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

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

Differentiating *w.r. to* θ, we get

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

$$\therefore \quad E(X) = \mu(\theta) = \bar{X},$$

as required.

[From (ii)...

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

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

Obtain the maximum likelihood estimator for θ.

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

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

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

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

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

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

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

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

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

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. Obt 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 fro consideration that the total area under a probability curve is unity.

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

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

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

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

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

The likelihood equations for estimating α and β give

$$\frac{\partial}{\partial \alpha} \log L = 0 = n\beta$$

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

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

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

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

If $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ is ordered sample from this population then $\alpha \le x_{(1)} \le x_{(2)}$ $\le x_{(n)} < \infty$, so that the maximum value of α consistent with the sample is $x_{(1)}$ 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 ran the variable is dependent on the parameter(s) to be estimated, first of all we should dete the constant by taking the total probability as unity and then proceed with the estimation p

2. From the last two examples, it is obvious that whenever the range of the var involves the parameter(s) to be estimated, the likelihood equations fail to give us estimates and in this case M.L.Es are obtained by adopting some other approach of maxim 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. $$L(x,\theta) = \prod_{i=1}^n f(x_i,\theta) = (1+\theta)^n \left(\prod_{i=1}^n x_i\right)^\theta$$

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

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

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

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

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 $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ in the sample. The likelihood equation is : $\frac{\partial}{\partial \theta} \log L = 0$, *i.e.*, $\frac{\partial}{\partial \theta}\left(\sum_x (\log f\right) = 0$

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

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

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 $A = \frac{\partial^2 \psi}{\partial \theta^2}$, 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.

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

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

which is the probability function of the required distribution.

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

$$\frac{\partial}{\partial \theta}\log L = \sum_x \frac{\partial}{\partial \theta}\log f = 0$$

is $\theta = (x_1, x_2, \ldots, x_n)^{1/n} \Rightarrow \log\theta = \frac{1}{n}\sum_x \log x$ or $\sum_x (\log x - \log\theta) = 0$

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

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

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

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

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 :

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

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

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

$$\Rightarrow \log L = k - 2n\log\sigma^2 - \frac{1}{2\sigma^2}\left\{\sum_j(x_{1j}-a-b-c)^2 + \sum_j(x_{2j}-a-b+c)^2 + \sum_j(x_{3j}-a+b-c)^2 + \sum_j(x_{4j}-a+b+c)^2\right\},$$

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 :

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

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

$$\Rightarrow \sum_{j=1}^{n}\left(\sum_{i=1}^{4}x_{ij}\right) + n(-4a) = 0$$

$$\therefore \quad \hat{a} = \frac{1}{4n}\sum_{i=1}^{4}\sum_{j=1}^{n}x_{ij} = \bar{x}$$

(2) gives :

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

$$\Rightarrow \quad \sum_j x_{1j} + \sum_j x_{2j} - \sum_j x_{3j} - \sum_j x_{4j}$$
$$+ n\left[(-a-b-c) + (-a-b+c) - (-a+b-c) - (-a ...\right.$$
$$\sum x_{1j} + \sum x_{2j} - \sum x_{3j} - \sum x_{4j} - 4nb = 0$$
$$\Rightarrow$$
$$\therefore \quad \hat{b} = \frac{1}{4}\left(\frac{1}{n}\sum x_{1j} + \frac{1}{n}\sum x_{2j} - \frac{1}{n}\sum x_{3j} - \frac{1}{n}\sum x_{4j}\right) \Rightarrow \hat{b} = (\bar{x}_1 + \bar{x}_2 - ...$$

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

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

Example 17·42. *The following table gives probabilities and observed frequen[cies in] classes AB Ab, aB and ab in a genetical experiment. Estimate the parameter θ by the 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

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

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

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

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 $\frac{108}{2+\theta} - \frac{(27+30)}{1-\theta} + \frac{8}{\theta} ...$

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

$$\therefore \quad \theta = \frac{-14 \pm \sqrt{196 + 11072}}{346} = -0{\cdot}34 \text{ and } 0{\cdot}26$$

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

M.L.E. of θ is given by $\hat{\theta} = 0{\cdot}26$ ...(**)

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

$$\frac{\partial^2 \log L}{\partial \theta^2} = \frac{-n_1}{(2+\theta)^2} - \frac{(n_2+n_3)}{(1-\theta)^2} - \frac{n_4}{\theta^2}$$
$$-E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right) = \frac{E(n_1)}{(2+\theta)^2} + \frac{E(n_2)+E(n_3)}{(1-\theta)^2} + \frac{E(n_4)}{\theta^2}$$
$$= \frac{np_1}{(2+\theta)^2} + \frac{n(p_2+p_3)}{(1-\theta)^2} + \frac{np_4}{\theta^2} = \frac{n(2+\theta)}{4(2+\theta)^2} + \frac{n(1-\theta)}{2(1-\theta)^2} + \frac{n\theta}{4\theta^2}$$
$$I(\theta) = \frac{n}{4(2+\hat{\theta})} + \frac{n}{2(1-\hat{\theta})} + \frac{n}{4\hat{\theta}}; \qquad n = \sum n_i = 173.$$
$$= 173\left(\frac{1}{4\times 2{\cdot}26} + \frac{1}{2\times 0{\cdot}74} + \frac{1}{4\times 0{\cdot}26}\right) = 301{\cdot}02$$
$$\text{S.E.}(\hat{\theta}) = \sqrt{1/I(\theta)} = \frac{1}{\sqrt{301{\cdot}02}} = 0{\cdot}0576$$ [*c.f.* (17·55), Theorem 17·13)

17·6·2. Method of Minimum Variance. (Minimum Variance Unbiased [E]stimates (M.V.U.E.)}. In this section we shall look for estimates which (*i*) are [u]nbiased 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 [fi]nd a statistic $t = t(x_1, x_2, \ldots, x_n)$, such that

$$E(t) = \int_{-\infty}^{\infty} t.L\,dx = \gamma(\theta) \quad \Rightarrow \quad \int_{-\infty}^{\infty} \{t - \gamma(\theta)\}\,L\,dx = 0 \qquad \ldots(17{\cdot}57)$$

[a]nd

$$V(t) = \int_{-\infty}^{\infty} [t - E(t)]^2\,L\,dx = \int_{-\infty}^{\infty} [t - \gamma(\theta)]^2\,L\,dx \qquad \ldots(17{\cdot}58)$$

[i]s minimum where

$\int_{-\infty}^{\infty} dx$ represents the n-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-[R]ao Inequality (§ 17·7).

17·6·3. Method of Moments. This method was discovered and studied in detail [b]y Karl Pearson.

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

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

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 popula method of moments consists in solving the k-equations (17·59) for $\theta_1, \theta_2, \ldots, \theta_k$ i of $\mu_1', \mu_2', \ldots, \mu_k'$ and then replacing these moments μ_r' ; $r = 1, 2, \ldots, k$ by the sam moments, *e.g.*, $\hat{\theta}_i = \theta_i(\hat{\mu}_1', \hat{\mu}_2', \ldots, \hat{\mu}_k') = \theta_i(m_1', m_2', \ldots, m_k')$; $i = 1, 2, \ldots$ 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 $\theta_2, \ldots, \theta_k$ respectively.

Remarks.1. Let $(x_1, x_2, \ldots, x_n)$ be a random sample of size n from a population w $f(x, \theta)$. 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 then by W.L.L.N., we get

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

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

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

3. Generally the method of moments yields less efficient estimators than those o from the principle of maximum likelihood. The estimators obtained by the method of m are identical with those given by the method of maximum likelihood if the probabili function or probability density function is of the form :

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

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

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

$$\Rightarrow \quad \frac{\partial}{\partial\theta}\log L = a_0 + a_1 \Sigma x_i + a_2 \Sigma x_i^2 + a_3 \Sigma x_i^3 + \ldots$$

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

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

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

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

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

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

Hence $\hat{\alpha} = \frac{m_1'^2}{m_2' - m_1'^2}$ and $\hat{\beta} = \frac{m_1'}{m_2' - m_1'^2}$, 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 $\hat{m}_2 = \mu_1' \pm \sqrt{\mu_2' - \mu_1' - \mu_1'^2}$

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

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

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

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

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

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

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

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

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

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 curve of the form : $\quad y = f(x, a_0, a_1, \ldots, a_n)$

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

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

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 :

$$\frac{\partial E}{\partial a_i} = 0; \; i = 1, 2, \ldots, n$$

Remarks. 1. In chapter 10, we have discussed in detail the method of least squ fitting linear regression, polynomial regression and the exponential family of curves re to linear regression. In chapter 11, we have discussed the method of fitting multiple 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, the x's being known given values, the least square estimators obtained as linear function 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 popu involving a single unknown parameter θ, (say). Let $f(x, \theta)$ be the probability f of the parent distribution from which the sample is drawn and let us suppose th distribution is continuous. Let $t = t(x_1, x_2, \ldots, x_n)$, a function of the sample value estimate of the population parameter θ, with the sampling distribution g $g(t, \theta)$.

Having obtained the value of the statistic t from a given sample, the problem is, "Can we make some reasonable probability statements about the unknown parameter θ in the population, from which the sample has been drawn?" This question is very well answered by the technique of *Confidence interval* due to Neyman and is obtained below :

We choose once for all some small value of α (5% or 1%) and then determine two constants say, c_1 and c_2 such that : $\quad P(c_1 < \theta < c_2 \mid t) = 1-\alpha \quad$ ...(17·65)

The quantities c_1 and c_2, so determined, are known as the *confidence limits* or *fiducial limits* and the interval $[c_1, c_2]$ within which the unknown value of the population parameter is expected to lie, is called the *confidence interval* and $(1-\alpha)$ is called the *confidence coefficient.*

Thus if we take $\alpha = 0{\cdot}05$ (or $0{\cdot}01$), we shall get 95% (or 99%) confidence limits.

How to find c_1 and c_2. ? Let T_1 and T_2 be two statistics such that

$$P(T_1 > \theta) = \alpha_1 \quad \ldots(17{\cdot}66)$$

and

$$P(T_2 < \theta) = \alpha_2 \quad \ldots(17{\cdot}66a)$$

where α_1 and α_2 are constants independent of θ. (17·66) and (17·66a) can be combined to give

$$P(T_1 < \theta < T_2) = 1-\alpha, \quad \ldots(17{\cdot}66b)$$

where $\alpha = \alpha_1 + \alpha_2$. Statistics T_1 and T_2 defined in (17·66) and (17·66a) may be taken as c_1 and c_2 defined in (17·65).

For example, if we take a large sample from a normal population with mean μ and standard deviation σ, then $\quad Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

and $\quad P(-1{\cdot}96 \le Z \le 1{\cdot}96) = 0{\cdot}95 \quad$ (From Normal Probability Tables)

$$\Rightarrow P\left(-1{\cdot}96 \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le 1{\cdot}96\right) = 0{\cdot}95 \quad \Rightarrow \quad P\left(\bar{x}-1{\cdot}96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x}+1{\cdot}96\frac{\sigma}{\sqrt{n}}\right) = 0{\cdot}95$$

Thus $\bar{x} \pm 1{\cdot}96\frac{\sigma}{\sqrt{n}}$ are 95% confidence limits for the unknown parameter μ, the population mean and the interval $\left(\bar{x}-1{\cdot}96\frac{\sigma}{\sqrt{n}}, \; \bar{x}+1{\cdot}96\frac{\sigma}{\sqrt{n}}\right)$ is called the 95% confidence interval.

Also $\quad P(-2{\cdot}58 \le Z \le 2{\cdot}58) = 0{\cdot}99 \quad$ or $\quad P\left(-2{\cdot}58 \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le 2{\cdot}58\right) = 0{\cdot}99$

$$\Rightarrow \quad P\left(\bar{x}-2{\cdot}58\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x}+2{\cdot}58\frac{\sigma}{\sqrt{n}}\right) = 0{\cdot}99$$

Hence 99% confidence limits for μ are : $\bar{x} \pm 2{\cdot}58\frac{\sigma}{\sqrt{n}}$ and

99% confidence interval for μ is $\left(\bar{x}-2{\cdot}58\frac{\sigma}{\sqrt{n}}, \; \bar{x}+2{\cdot}58\frac{\sigma}{\sqrt{n}}\right)$.

Remarks 1. Usually σ^2 is not known and its unbiased estimate S^2 obtained from th samples, is used. However if n is small, $Z = \frac{\bar{x}-\mu}{S/\sqrt{n}}$ is not $N(0, 1)$ and in this case the confidenc limits and confidence intervals for μ are obtained by using Student's 't' distribution.

2. It can be seen that in many cases there exist more than one set of confidence ... with the same confidence coefficient. Then the problem arises as to which particular set ... regarded as better than the others in some useful sense and in such cases we look ... shortest of all the intervals.

Example 17·46. *Obtain 100 $(1-\alpha)$% confidence intervals for the para... (a) θ and (b) σ^2, of the normal distribution :*

$$f(x, \theta; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right\}, -\infty < x < \infty$$

Solution. Let X_i, $(i = 1, 2, ..., n)$ be a random sample of size n from the d... $f(x; \theta, \sigma)$ and let :

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - ...$$

(a) The statistic $t = \dfrac{\bar{X}-\theta}{S/\sqrt{n}}$ follows student's t-distribution with $(n-1)$ degr... freedom. Hence $100(1-\alpha)$% confidence limits for θ are given by :

$$P(|\,t\,| \le t_\alpha) = 1 - \alpha \quad \Rightarrow \quad P\left(|\,\bar{X} - \theta\,| \le \frac{S}{\sqrt{n}}\, t_\alpha\right) = 1 - \alpha$$

$$\therefore \quad P\left(\bar{X} - t_\alpha \cdot \frac{S}{\sqrt{n}} \le \theta \le \bar{X} + t_\alpha \cdot \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

where t_α is the tabulated value of t for $(n-1)$ $d.f.$ at significance level 'α'. Hence th... required confidence interval for θ is :

$$\left(\bar{X} - t_\alpha \frac{S}{\sqrt{n}},\ \bar{X} + t_\alpha \frac{S}{\sqrt{n}}\right)$$

(b) *Case (i) θ is known and equal to μ (say).*

Then

$$\frac{\Sigma(X_i - \mu)^2}{\sigma^2} = \frac{ns^2}{\sigma^2} \sim \chi^2_{(n)}$$

If we define χ_α^2 as the value of χ^2 such that $P(\chi^2 > \chi_\alpha^2) = \int_{\chi_\alpha^2}^{\infty} p(\chi^2)d\chi^2 = \alpha$

where $p(\chi^2)$ is the $p.d.f.$ of χ^2-distribution with n $d.f.$, then the required confide... interval is given by :

$$P\{\chi^2_{1-(\alpha/2)} \le \chi^2 \le \chi^2_{\alpha/2}\} = 1 - \alpha \quad \Rightarrow \quad P\left\{\chi^2_{1-(\alpha/2)} \le \frac{ns^2}{\sigma^2} \le \chi^2_{\alpha/2}\right\} = 1 - \alpha$$

Now $\dfrac{ns^2}{\sigma^2} \le \chi^2_{\alpha/2} \Rightarrow \dfrac{ns^2}{\chi^2_{\alpha/2}} \le \sigma^2$ and $\chi^2_{1-(\alpha/2)} \le \dfrac{ns^2}{\sigma^2} \Rightarrow \sigma^2 \le \dfrac{ns^2}{\chi^2_{1-(\alpha/2)}}$

Hence (**) gives :

$$P\left\{\frac{ns^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{1-(\alpha/2)}}\right\} = 1 - \alpha,$$

where $\chi^2_{\alpha/2}$ and $\chi^2_{1-(\alpha/2)}$ are obtained from (*) by using n $d.f.$

Thus *e.g.*, 95% confidence interval for σ^2 is :

$$P\left(\frac{ns^2}{\chi^2_{0\cdot025}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{0\cdot975}}\right) = 0\cdot95$$

Case (ii). θ is unknown. In this case the statistic :

$$\frac{\Sigma(X_i - \bar{X})^2}{\sigma^2} = \frac{ns^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

Here also confidence interval for σ^2 is given by (***), where now χ^2_α is ... significant value of χ^2 [as defined in (*)] for $(n-1)$ $d.f.$ at the significance level 'α'.

Example 17·47. *Show that the largest observations L of a sample of n observations from a rectangular distribution with density function :*

$$f(x, \theta) = \begin{cases} \dfrac{1}{\theta}, & 0 \le x \le \theta \\ 0, & \text{otherwise} \end{cases} \quad ...(*)$$

has the distribution :

$$dG(L) = n\left(\frac{L}{\theta}\right)^{n-1} \cdot \frac{dL}{\theta}, 0 \le L \le \theta$$

Show that the distribution of $V = L/\theta$ is given by p.d.f. : $h(v) = nv^{n-1}, 0 \le v \le 1$

Hence deduce that the confidence limits for θ corresponding to confidence coefficient α are

$$L \quad \text{and} \quad \frac{L}{(1-\alpha)^{1/n}}.$$

Solution. Let $X_1, X_2, ..., X_n$ be a random sample of size n from the population (*) and let $L = \max(X_1, X_2, ..., X_n)$. The distribution of L is given by :

$dG(L) = n[F(L)]^{n-1}. f(L)\, dL$, where $F(.)$ is the distribution function of X given by :

$$F(L) = \int_0^L f(x, \theta)dx = \frac{L}{\theta} \qquad \therefore\ dG(L) = n\left(\frac{L}{\theta}\right)^{n-1} \cdot \frac{dL}{\theta}, 0 \le L \le \theta$$

If we take $V = L/\theta$, the Jacobian of transformation is θ. Hence $p.d.f.$ $h(.)$ of V is :

$$h(v) = nv^{n-1} \cdot \frac{1}{\theta} \,|\, J \,| = nv^{n-1}, 0 \le v \le 1,$$

which is independent of θ.

To obtain the confidence limits for θ, with confidence coefficient α, let us define v_α such that

$$P(v_\alpha < V < 1) = \alpha \quad \Rightarrow \quad \int_{v_\alpha}^{1} h(v)\, dv = \alpha \quad ...(**)$$

$$\Rightarrow \quad n\int_{v_\alpha}^{1} v^{n-1}\, dv = \alpha \quad \Rightarrow \quad 1 - v_\alpha^{\ n} = \alpha \quad \Rightarrow \quad v_\alpha = (1-\alpha)^{1/n} \quad ...(***)$$

From (**) and (***), $P\{(1-\alpha)^{1/n} < V < 1\} = \alpha \Rightarrow P\left\{(1-\alpha)^{1/n} < \dfrac{L}{\theta} < 1\right\} = \alpha$

$$\therefore \quad P\left\{L < \theta < \frac{L}{(1-\alpha)^{1/n}}\right\} = \alpha$$

Hence the required confidence limits for θ are L and $L/(1-\alpha)^{1/n}$.

Example 17·48. *Given a random sample from a population with p.d.f. :*

$$f(x, \theta) = \frac{1}{\theta}, \quad 0 \le x \le \theta$$

show that 100 $(1-\alpha)$% confidence interval for θ is given by R and R/ψ, where ψ is given by $\psi^{n-1}[n-(n-1)\psi] = \alpha$, and R is the sample range.

Solution. The joint $p.d.f.$ of $x_1, x_2, ..., x_n$ is given by : $\quad L = \dfrac{1}{\theta^n}, 0 \le x_i \le \theta$

If $x_{(1)}, x_{(2)}, ..., x_{(n)}$ is the ordered sample then the joint $p.d.f.$ of $x_{(n)}$ and $x_{(1)}$ is :

$$g[x_{(1)}, x_{(n)}] = \frac{n(n-1)}{\theta^n}\left[x_{(n)} - x_{(1)}\right]^{n-2}, 0 \le x_{(1)} \le x_{(n)} \le \theta$$

To obtain the distribution of the sample range R, let us make the transformation of variables :

$$R = x_{(n)} - x_{(1)} \text{ and } v = x_{(1)} \Rightarrow v = x_{(n)} - R \le \theta - R.$$

The Jacobian of transformation is $|J| = 1$ and the joint $p.d.f.$ of R and v becomes

$$h(R, v) = \frac{n(n-1)}{\theta^n} R^{n-2}, 0 < v < \theta - R$$

The marginal density of R is given by :

$$h_1(R) = \int_0^{\theta - R} \frac{n(n-1)}{\theta^n} . R^{n-2} dv = \frac{n(n-1) R^{n-2} (\theta - R)}{\theta^n}, 0 \le R \le \theta$$

The $p.d.f.$ $h_2(.)$ of $U = R/\theta$ is :

$$h_2(u) = h_1(R) \left| \frac{dR}{du} \right| = \frac{n(n-1) R^{n-2} (\theta - R)}{\theta^n} . \theta = n(n-1) u^{n-2} (1-u), 0 \le u \le 1$$

100 $(1 - \alpha)$% confidence interval for θ is given by : $P(\psi \le U \le 1) = 1 - \alpha$

where ψ is obtained from the equation $\int_0^{\psi} h_2(u) du = \alpha$

$$\Rightarrow \quad n(n-1) \int_0^{\psi} u^{n-2} (1-u) du = \alpha \quad \Rightarrow \quad \left| n u^{n-1} - (n-1) u^n \right|_0^{\psi} = \alpha$$

$$\therefore \quad \psi^{n-1} \{n - (n-1)\psi\} = \alpha$$

From (*), we get

$$P\left(\psi \le \frac{R}{\theta} \le 1\right) = 1 - \alpha \quad \Rightarrow \quad P\left(R \le \theta \le \frac{R}{\psi}\right) = 1 - \alpha$$

Hence the required limits for θ are given by R and R/ψ where ψ is given by

Example 17·49. *Given one observation from a population with p.d.f. :*

$$f(x, \theta) = \frac{2}{\theta^2} (\theta - x), \quad 0 \le x \le \theta,$$

obtain 100 $(1 - \alpha)$% confidence interval for θ.

Solution. The density of $u = x/\theta$ is given by :

$$g(u) = f(x, \theta) . \left| \frac{dx}{du} \right| = \frac{2}{\theta^2} (\theta - x) . \theta = 2(1-u), \quad 0 \le u \le 1$$

To obtain 100 $(1 - \alpha)$% confidence interval for θ, we choose two quantities u_1 and u_2 such that

$$\left.\begin{aligned} P(u_1 \le u \le u_2) &= 1 - \alpha \\ \text{and} \quad P(u < u_1) &= P(u > u_2) = \tfrac{1}{2}\alpha \end{aligned}\right\}$$

$$P(u < u_1) = \frac{\alpha}{2} \quad \Rightarrow \quad \int_0^{u_1} 2(1-u) du = \frac{\alpha}{2} \quad \Rightarrow \quad u_1^2 - 2u_1 + \frac{\alpha}{2} = 0$$

$$\text{and} \quad P(u > u_2) = \frac{1}{2}\alpha \quad \Rightarrow \quad \int_{u_2}^{1} 2(1-u) du = \frac{\alpha}{2} . \quad \Rightarrow \quad u_2^2 - 2u_2 + \left(1 - \frac{\alpha}{2}\right) = 0$$

From (*), we get $\quad P\left(u_1 \le \frac{x}{\theta} \le u_2\right) = 1 - \alpha \quad \Rightarrow \quad P\left(\frac{x}{u_2} \le \theta \le \frac{x}{u_1}\right) = 1 - \alpha$

Hence the required interval for θ is $\left(\frac{x}{u_2}, \frac{x}{u_1}\right)$, where u_1 and u_2 are given by and (***) respectively.

17·7·1. Confidence Intervals for Large Samples. It has been proved that under certain regularity conditions, the first derivative of the logarithm of the likelihood function w.r.to parameter θ *viz.*, $\frac{\partial}{\partial \theta} \log L$, is asymptotically normal with mean zero and variance given by :

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

Hence for large n,

$$Z = \frac{\frac{\partial}{\partial \theta} \log L}{\sqrt{\text{Var}\left(\frac{\partial}{\partial \theta} \log L\right)}} \sim N(0, 1) \quad \text{...(17·68)}$$

The result enables us to obtain confidence interval for the parameter θ in large samples. Thus for large samples, the confidence interval for θ with confidence coefficient $(1 - \alpha)$ is obtained by converting the inequalities in

$$P(|Z| \le \lambda_\alpha) = 1 - \alpha \quad \text{...(17·69)}$$

where λ_α is given by $\frac{1}{\sqrt{2\pi}} \int_{-\lambda_\alpha}^{\lambda_\alpha} \exp(-u^2/2) \, du = 1 - \alpha$ ...[17·69(*a*)]

Example 17·50. *Obtain 100 $(1 - \alpha)$% confidence limits (for large samples) for the parameter λ of the Poisson distribution :*

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

Solution. We have

$$\frac{\partial}{\partial \lambda} \log L = \frac{\partial}{\partial \lambda} \left\{ -n\lambda + \left(\sum_{i=1}^{n} x_i\right) \log \lambda - \sum_{i=1}^{n} \log (x_i) \right\} = -n + \frac{\Sigma x_i}{\lambda} = n\left(\frac{\bar{x}}{\lambda} - 1\right)$$

$$\text{Var}\left(\frac{\partial}{\partial \lambda} \log L\right) = E\left(- \frac{\partial^2}{\partial \lambda^2} \log L\right) = E\left(\frac{n\bar{x}}{\lambda^2}\right) = \frac{n}{\lambda^2} E(\bar{x}) = \frac{n}{\lambda} \quad [\because E(\bar{x}) = \lambda]$$

$$\therefore \quad Z = \frac{n\left(\frac{\bar{x}}{\lambda} - 1\right)}{\sqrt{n/\lambda}} = \sqrt{(n/\lambda)} (\bar{x} - \lambda) \sim N(0, 1) \quad \text{[Using (17·68)]}$$

Hence 100 $(1 - \alpha)$% confidence interval for λ is given by (for large samples)

$$P\{ | \sqrt{(n/\lambda)} (\bar{x} - \lambda) | \le \lambda_\alpha \} = 1 - \alpha$$

Hence the required limits for λ are the roots of the equation :

$$| \sqrt{n/\lambda} (\bar{x} - \lambda) | = \lambda_\alpha \quad \Rightarrow \quad n(\bar{x} - \lambda)^2 - \lambda . \lambda_\alpha^2 = 0$$

$$\Rightarrow \lambda^2 - \lambda\left(2\bar{x} + \frac{\lambda_\alpha^2}{n}\right) + \bar{x}^2 = 0 \Rightarrow \lambda = \frac{\left(2\bar{x} + \frac{\lambda_\alpha^2}{n}\right) \pm \left\{\left(2\bar{x} + \frac{\lambda_\alpha^2}{n}\right)^2 - 4\bar{x}^2\right\}^{1/2}}{2} \quad \text{...(*)}$$

For example, 95% confidence interval for λ is given by taking $\lambda_\alpha = 1·96$ in (*), thus giving :

$$\lambda = \frac{1}{2}\left(2\bar{x} + \frac{3·84}{n}\right) \pm \left(\frac{3·84\bar{x}}{n} + \frac{3·69}{n^2}\right)^{1/2} = \bar{x} \pm 1·96 \sqrt{\bar{x}/n}, \text{ to the order } n^{-1/2}.$$

Example 17·51. *Show that for the distribution :* $\quad dF(x) = \theta e^{-x\theta}; 0 < x < \infty,$ *central confidence limits for large samples with 95% confidence coefficient are given by*

$$\theta = \left(1 \pm \frac{1.96}{\sqrt{n}}\right) \Big/ \bar{x}.$$

Solution. Here $$L = \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right)$$

$$\frac{\partial}{\partial\theta}\log L = \frac{\partial}{\partial\theta}(n \log\theta - \theta \Sigma x_i) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i = n\left(\frac{1}{\theta} - \bar{x}\right)$$

and $$\frac{\partial^2}{\partial\theta^2}\log L = -\frac{n}{\theta^2} \quad \Rightarrow \quad \text{Var}\left(\frac{\partial}{\partial\theta}\log L\right) = E\left(-\frac{\partial^2}{\partial\theta^2}\log L\right) = \frac{n}{\theta^2}$$

Hence, for large samples, using (17·68), we have

$$Z = \frac{n\left(\frac{1}{\theta} - \bar{x}\right)}{\sqrt{n/\theta^2}} \sim N(0, 1) \quad \Rightarrow \quad \sqrt{n}\,(1 - \theta\bar{x}) \sim N(0, 1)$$

Hence 95% confidence limits for θ are given by :

$$P[-1.96 \le \sqrt{n}\,(1 - \theta\bar{x}) \le 1.96] = 0.95$$

$$\sqrt{n}\,(1 - \theta\bar{x}) \le 1.96 \quad \Rightarrow \quad \left(1 - \frac{1.96}{\sqrt{n}}\right)\frac{1}{\bar{x}} \le \theta$$

and $$-1.96 \le \sqrt{n}\,(1 - \theta\bar{x}) \quad \Rightarrow \quad \theta \le \left(1 + \frac{1.96}{\sqrt{n}}\right)\frac{1}{\bar{x}}$$

Hence, from (*), (**) and (***), the central 95% confidence limits for θ are gi

$$\theta = \left(1 \pm \frac{1.96}{\sqrt{n}}\right) \cdot \frac{1}{\bar{x}}.$$

CHAPTER CONCEPTS QUIZ

1. Comment on the following statements :
   (i) In the case of Poisson distribution with parameter λ, $\bar{x}$ is sufficient for λ.
   (ii) If $(X_1, X_2, \ldots X_n)$ be a sample of independent observation from the u distribution on $(\theta, \theta + 1)$, then the maximum likelihood estimator of θ is u
   (iii) A maximum likelihood estimator is always unbiased.
   (iv) Unbiased estimator is necessarily consistent
   (v) A consistent estimator is also unbiased.
   (vi) An unbiased estimator whose variance tends to zero as the sample size inc is consistent.
   (vii) If t is a sufficient statistic for θ then $f(t)$ is a sufficient statistic for $f(\theta)$.
   (viii) If t_1 and t_2 are two independent estimators of θ, then $t_1 + t_2$ is less efficie both t_1 and t_2.
   (ix) If T is consistent estimator of a parameter θ, then $aT + b$ is a consistent es of $a\theta + b$, where a and b are constants.
   (x) If x is the number of successes in n independent trials with a co probability p of success in each trial, then x/n is a consistent estimator of p.

2. Fill in the blanks :
   (i) In a random sample of size n from a population with mean μ, the sample mean $(\bar{x})$ is … estimate of …
   (ii) The sample median is … estimate for the mean of normal population.
   (ii) An estimator $\hat{\theta}$ of a parameter θ is said to be unbiased if…
   (iii) The variance s^2 of a sample of size n is a … estimator of population variance σ^2.
   (iv) If a sufficient estimator exists, it is a function of the … estimator.
   (v) …estimate may not be unique.
3. (a) Give example of a statistic t which is unbiased for a parameter θ but t^2 is not unbiased for θ^2.
   (b) Give example of an M.L. estimator which is not unbiased.
4. (i) If $\bar{x}$ is an unbiased estimator for the population mean μ, state which of the following are unbiased estimators for μ^2 :
   (a) $\bar{x}^2$, (b) $\bar{x}^2 - \frac{\sigma^2}{n}$ (σ^2 is known/unknown)
   (ii) If t is the maximum likelihood estimator for θ, state the condition under which $f(t)$ will be the maximum likelihood estimator for $f(\theta)$.
   (iii) Write down the condition for the Cramer-Rao lower bound for the variance of the estimator to be attained.
   (iv) Write down the general form of the distribution admitting sufficient statistic.
5. A random variable X takes the values 1, 2, 3 and 4, each with probability $\frac{1}{4}$. A random sample of three values of x is taken, $\bar{x}$ is the mean and m is the median of this sample. Show that both $\bar{x}$ and m are unbiased estimators of the mean of the population, but $\bar{x}$ is more efficient than m. Compare their efficiencies.
6. Give an example of estimates which are (i) Unbiased and efficient, (ii) Unbiased and inefficient.
7. Mark the correct alternative :
   (i) Let T_n be an estimator, based on a sample $x_1, x_2, \ldots, x_n$, of the parameter θ. Then T is a consistent estimator of θ if
   (a) $P(T_n - \theta > \varepsilon) = 0 \ \forall \ \varepsilon > 0$, (b) $P(|T_n - \theta| < \varepsilon) = 0$,
   (c) $\lim_{n\to\infty} P(|T_n - \theta| > \varepsilon) = 0 \ \forall \ \varepsilon > 0$, (d) $\lim_{n\to\infty} P(T_n - \theta > \varepsilon) = 0 \ \forall \ \varepsilon > 0$
   (ii) Let $E(T_1) = \theta = E(T_2)$, where T_1 and T_2 are the linear functions of the sample observations. If $V(T_1) \le V(T_2)$ then:
   (a) T_1 is an unbiased linear estimator.
   (b) T_1 is the best linear unbiased estimator.
   (c) T_1 is a consistent linear unbiased estimator.
   (d) T_1 is a consistent best linear unbiased estimator.
   (iii) Let X be a random variable with $E(X) = \mu$ and $V(X) = \sigma^2$. Let $\overline{X}$ be the sample mean based on a random sample of size n, then $\overline{X}$ is :
   (a) the best linear unbiased estimator of μ.
   (b) an unbiased and consistent estimator of μ.