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

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

MLE

$r(\theta)$

θ. $X \sim N(\mu, \sigma^2)$

$f(x, \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$\theta \leftrightarrow (\mu, \sigma^2)$

$(\theta_1, \theta_2) \leftrightarrow (\mu, \sigma^2)$

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

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

L gives the relative likelihood that the random variables assume a particular set of values $x_1, x_2, \ldots, x_n$. For a given sample $x_1, x_2, \ldots, x_n$, L becomes a function of the variable θ, the parameter.

$L(\theta)$

$L(\theta)$

$f(x)$

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

$\theta \leftrightarrow (\theta_1, \theta_2)$

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

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

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

$\frac{d}{dx} f(x) = 0$ $\qquad \frac{\partial L}{\partial \theta} = 0$

$x_1, x_2, \ldots, x_n$

$$\frac{\partial L}{\partial \theta} = 0 \quad \text{and} \quad \frac{\partial^2 L}{\partial \theta^2} < 0 \qquad \ldots(15\cdot54)$$

$\frac{\partial^2 L}{\partial \theta^2} < 0$

$\frac{d^2 f}{dx^2} < 0$

Since $L > 0$, and $\log L$ is a non-decreasing function of L ; L and $\log L$ attain their extreme values (maxima or minima) at the same value of $\hat{\theta}$. The first of the two equations in (15·54) can be rewritten as

$$\frac{1}{L} \cdot \frac{\partial L}{\partial \theta} = 0 \quad \Rightarrow \quad \frac{\partial \log L}{\partial \theta} = 0, \qquad \ldots(15\cdot54a)$$

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

$\frac{\partial L}{\partial \theta} = 0$ $\qquad \frac{\partial^2}{\partial \theta^2}(\log L) < 0$

$\frac{\partial}{\partial \theta}(\log L) = 0$

$\frac{1}{L}\frac{\partial}{\partial \theta} L = 0$

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

$$\frac{\partial}{\partial \theta_i} \log L = \frac{\partial}{\partial \theta_i} \log L(\theta_1, \theta_2, \ldots, \theta_k) = 0\ ;\ i = 1, 2, \ldots, k \qquad \ldots(15\cdot54b)$$

$\theta \leftrightarrow (\theta_1, \theta_2, \ldots, \theta_k)$

Equations (15·54*a*) and (15·54*b*) are usually referred to as the *Likelihood Equations* for estimating the parameters.

Example 15·31. *In random sampling from normal population $N(\mu; \sigma^2)$, find the maximum likelihood estimators for*

(i) μ when σ^2 is known,
(ii) σ^2 when μ is known, and
(iii) the simultaneous estimation of μ and σ^2.

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

$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2} \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_2-\mu}{\sigma}\right)^2} \times \ldots \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2}$

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

$e^{m_1} \times e^{m_2} \times \ldots \times e^{m_n} = e^{m_1+m_2+\ldots+m_n} = e^{\sum m_i}$

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

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

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

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

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

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

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

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

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

$\hat{u} = \bar{x}$

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

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

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

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

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

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

$$= n\left[\log(1) - \left\{\log\sigma + \log(2\pi)^{1/2}\right\}\right]$$

$$-\frac{1}{2}\sum\left(\frac{x_i-\mu}{\sigma}\right)^2 \times \log_e e$$

$$= n\left[0 - \log\sigma - \frac{1}{2}\log(2\pi)\right] - \frac{1}{2}\sum\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$= -\frac{n}{2}\left[2\log\sigma + \log 2\pi\right] - \frac{1}{2}\left(\sum\right)$$

$$\log L = -\frac{n}{2}\left[\log\sigma^2 + \log 2\pi\right] - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$

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

$$\frac{\partial}{\partial\mu}(\log L) = -0 - 0 - \frac{1}{\cancel{2}\sigma^2}\sum \cancel{2}(x_i-\mu)^{2-1} \times (0-1) = 0$$

$$\frac{d}{d\mu}\left[f(\mu)\right]^n = n\left[f(\mu)\right]^{n-1} \times \frac{d}{d\mu}f(\mu)$$

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

$$\Rightarrow \sum_{i=1}^{n} x_i - n\mu = 0$$

$$\hat{\mu} = \frac{\sum x_i}{n} = \bar{x}$$

$$\frac{\partial}{\partial\mu}(\log L) =$$

$$-0-0-\frac{1}{\cancel{2}\sigma^2}\sum \cancel{2}(x_i-\mu)^1(-1)$$

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

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

$$\log L = -\frac{n}{2}\overset{\log y}{\log\sigma^2} - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$

$$\frac{d}{dy}\left(\frac{1}{y}\right) = -\frac{1}{y^2}$$

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

$$\frac{d}{d\sigma^2}\left(\frac{1}{\sigma^2}\right) = \frac{-1}{(\sigma^2)^2}$$

$$\sigma^2 = y$$

$$y = \sigma^2$$

$$\frac{1}{\cancel{2}\sigma^{\cancel{4}2}}\sum(x_i-\mu)^2 = \frac{n}{\cancel{2}} \times \frac{1}{\cancel{\sigma^2}}$$

$$\sum(x_i-\mu)^2 = n\sigma^2 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial}{\partial\mu}(\log L) = 0$$

$$\hat{\mu} = \bar{x}$$

$$\frac{\partial}{\partial\sigma^2}(\log L) = 0$$

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

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = s^2$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = s^2$$

Poll Que. (i) Which of the following is the maximum likelihood estimate of the parameter 'α' of a population having density function: $\frac{2}{\alpha^2}(\alpha - x), 0 < x < \alpha$ when a sample of unit size is drawn from the population? (Here 'x' is the sample value)

(a) $2x$ (b) x (c) $\frac{x}{2}$ (d) *None of these.*

Poll Que. (ii) Is the above MLE is biased?

(a) Yes (b) No.

$\theta \to \alpha$

$\theta \leftrightarrow (\theta_1, \theta_2)$

$\mu \quad \sigma^2$

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

$i = 1, 2, \ldots, n \qquad n = 1$

$$\prod_{i=1}^{n} f(x_i; \alpha) = f(x_1, \alpha) = f(x, \alpha) = \frac{2}{\alpha^2}(\alpha - x)$$

$$\log L = \log 2 + \log(\alpha - x) - 2\log\alpha$$

$$\frac{\partial}{\partial\alpha}(\log L) = 0 + \frac{1}{\alpha - x}(1 - 0) - \frac{2}{\alpha} = 0$$

$$\frac{2}{\alpha} = \frac{1}{\alpha - x} \Rightarrow 2\alpha - 2x = \alpha \Rightarrow \hat{\alpha} = 2x$$

$$\boxed{\frac{-1}{(\alpha - x)^2} + \frac{2}{\alpha^2} < 0} \qquad \boxed{\frac{2}{\alpha^2} < \frac{1}{(\alpha - x)^2} \qquad 0 < x < \alpha}$$

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

$$= \frac{4}{\alpha^2}\left[\frac{\alpha^3}{2} - \frac{\alpha^3}{3}\right] = \frac{4}{\alpha^2}\left[\frac{\alpha^3}{6}\right] = \frac{2\alpha}{3} \neq \alpha$$