Maximum Likelihood Estimation

Session 11

05/04/2021

Methods of Estimation

- So far we have been discussing the requisites of a good estimator. Now we shall briefly outline some(will discuss only one in detail here) 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*

Method of Maximum Likelihood Estimation

- From theoretical point of view, the most general method of estimation known is the method of Maximum Likelihood Estimation which was initially formulated by C.F.Gauss but as a general method of estimation was first introduced by Prof.R.A.Fisher.

Likelihood function(Definition)

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

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

L gives the relative likelihood that the random variables assume a particular set of values $x_1, x_2, \dots x_n$

- The principle of maximum likelihood consist in finding an estimator for the unknown parameter $\theta = (\theta_1, \theta_2, \dots \theta_k)$, say which maximizes the likelihood function $L(\theta)$ for variations in parameter.

- i.e., we wish to find $\hat{\theta} = \left(\widehat{\theta_1}, \widehat{\theta_2}, \dots, \widehat{\theta_n}\right)$ so that

  i.e., $L\left(\hat{\theta}\right) = \mathrm{Sup} L(\theta)$ for all $\theta \in \Theta$

- Thus if there exist a function $\hat{\theta} = \left(\widehat{\theta_1}, \widehat{\theta_2}, \dots, \widehat{\theta_n}\right)$ of sample values which maximizes L for variations in θ, then $\hat{\theta}$ is to be taken as an estimator of θ and $\hat{\theta}$ is usually called M.L.E.

- *Thus* $\hat{\theta}$ is the solution, if any, of

$$\frac{\partial L}{\partial \theta} = 0 \quad \text{and} \quad \frac{\partial^2 L}{\partial \theta^2} < 0$$

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 $\hat{\theta}$.

- **Thus,**

$$\frac{1}{L} \cdot \frac{\partial L}{\partial \theta} = 0 \quad \Rightarrow \quad \frac{\partial \log L}{\partial \theta} = 0,$$

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

- Thus from the practical point of view, we find the following likelihood equation for estimating the parameters.

$$\frac{\partial logL}{\partial\theta} = 0$$

- If θ is a vector valued parameter, then $\hat{\theta} = \left(\widehat{\theta_1}, \widehat{\theta_2}, \ldots, \widehat{\theta_n}\right)$, is given by the solution of simultaneous equations

$$\frac{\partial}{\partial\theta_i} logL = \frac{\partial}{\partial\theta_i} logL(\theta_1, \theta_2, \ldots \theta_n) = 0; \mathrm{i} = 1,2, \ldots \mathrm{n}$$

Example 1

- In random sampling from a 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
- (iii) the simultaneous estimation for 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

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

or

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

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

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

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

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

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

$$\frac{\partial}{\partial \mu} \log L = 0 \quad \text{and} \quad \frac{\partial}{\partial \sigma^2} \log L = 0, \text{ thus giving}$$

$$\hat{\mu} = \bar{x} \qquad \text{[From (*)]}$$

and

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2 \qquad \text{[From (**)]}$$

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

- IMPORTANT : It may be noted that MLE need not be necessarily unbiased

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 2

- Find the M.L.E. for the parameter λ of a Poisson distribution on the basis of a random sample of size n.

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

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

$$\therefore \qquad \log L = -n\lambda + \left(\sum_{i=1}^{n} x_i\right) \log \lambda - \sum_{i=1}^{n} \log (x_i\,!)$$

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

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 $\bar{x}$.