11 - 04 - 2022

Tuesday, April 12, 2022 10:00 AM

Unit-4

[redacted] Unit - 4 (Point Estimation)

Unbiasedness
Consistency
Efficiency

* Methods of Estimation -

→ Method of maximum likelihood estimation :-

1st time C.F. Gauss

Prof. R.A. Fisher

§ Likelihood function :- 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 -

$$L = f(x_1,\theta) f(x_2,\theta) f(x_3,\theta) \cdots f(x_n,\theta)$$

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

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

$\theta = \hat{\theta}$ for which L is maximum.

To find the maximum value of likelihood function $L(\theta)$, we follow the steps below :-

(i) $\frac{\partial}{\partial\theta} L(\theta) = 0 \Rightarrow$ points of extrema.

(ii) $\frac{\partial^2}{\partial\theta^2} L(\theta)$. If $\frac{\partial^2}{\partial\theta^2} L(\theta) < 0$, θ is a point of maxima.

Let $\theta = \hat{\theta}$.

Then, the maximum likelihood estimator is-

$L(\hat{\theta})$.

$L(\hat{\theta})$.

§ Properties of Maximum Likelihood Estimators :-

(i) The first & second order derivatives i.e. $\frac{\partial}{\partial \theta} \log L$ and $\frac{\partial^2}{\partial \theta^2} \log L$ exist and are continuous functions of θ in a range R for almost all x.

(ii) The 3rd order derivative $\frac{\partial^3}{\partial \theta^3} \log L$ exists such that

$$\left| \frac{\partial^3}{\partial \theta^3} \log L \right| < M(x), \quad \text{where } E[M(x)] < K, \text{ a +ve quantity.}$$

(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 & 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.

$$\boxed{E(x) = \int_{-\infty}^{\infty} x f(x) \, dx}$$

* 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.

$$\boxed{\begin{array}{l} \text{As } n\to\infty, \quad P\{|E(T_n)-\theta|<\epsilon\}\to 1, \quad \text{as } n\to\infty \\ \qquad\qquad\quad P\{|E(\theta)-\theta_0|<\epsilon\}\to 1. \end{array}}$$

Remark:- MLE's are always consistent estimators but need not be unbiased.

For example- in sampling from $N(\mu,\sigma^2)$ population-

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

but $MLE(\sigma^2)=s^2$ (sample mean), which is consistent but not unbiased.

$$\boxed{E\left(\frac{n}{n-1}s^2\right)=\sigma^2}$$

Theorem (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 $(n)\to\infty$.

$$\frac{\partial}{\partial\theta}\log L=0 \leftarrow \text{maximum of the likelihood as } n\to\infty.$$

Theorem:- If M.L.E. exists, it is the most efficient in the class of such estimators.

Maximum likelihood estimator.

$\theta=\hat{\theta}\longrightarrow$ most efficient

$\theta = \hat{\theta} \longrightarrow$ most efficient $\boxed{\varepsilon = 1}$

§ Invariance property of M.L.E :-

If T is the M.L.E of θ, and $\psi(\theta)$ is one-to-one function of θ, then $\psi(T)$ is the M.L.E of $\psi(\theta)$.

Ex- In random sampling from normal population $N(\mu, \sigma^2)$, find the maximum likelihood estimator for μ when σ^2 is known.

Solution: Here, $X \sim N(\mu, \sigma^2)$, then -

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

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

so, $$f(x_i, \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}.$$

Now, Likelihood function is -

$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_2-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}$$

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

$$e^{x_1} \cdot e^{x_2} \cdots e^{x_n} = e^{\sum_{i=1}^{n} x_i}$$

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

$$2\log a = \log a^2$$

$$L(\theta) = \frac{1}{\ldots} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$$

$\frac{\partial}{\partial\theta}\log L$ $\quad$ $2\log$

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

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

Now, when σ^2 is known, the likelihood equation is –

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

$$\Rightarrow -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(x_i-\mu)(-1) = 0 \quad [\text{from } (1)]$$

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

$$\Rightarrow (x_1-\mu) + (x_2-\mu) + \cdots (x_n-\mu) = 0$$

$$\Rightarrow x_1 + x_2 + \cdots + x_n = n\mu$$

$$\Rightarrow \mu = \frac{x_1+x_2+\cdots+x_n}{n}$$

$$\Rightarrow \mu = \bar{x} \text{ (sample mean)}.$$

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

(b) M.L.E for σ^2, when μ is known.

Diff^n (1) partially w.r.t σ^2, we have –

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

$$\Rightarrow \left(-\frac{n}{2}\times\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2\right) = 0 \times (-2\sigma^2)$$

$$\Rightarrow n - \frac{1}{\quad}\sum(x_i-\mu)^2 = 0$$

$\frac{\partial}{\partial x}\left(\frac{-1}{x}\right)$

$\frac{1}{x^2}$

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

$\frac{1}{\sigma^4}$

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

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

x_1

(… variance)

$\Rightarrow \sigma^2 = \frac{1}{n} \sum_{i=1}$

$\Rightarrow \sigma^2 = s^2$ (sample variance)

$\therefore$ M.L.E for σ^2 is the sample variance s^2.

(c) Find the M.L.E for the simultaneous estimation of μ & σ^2.

$\rightarrow$ The likelihood equations for the simultaneous estimation of μ & σ^2 are -

✓ $\frac{\partial}{\partial \mu} \log L = 0$ & $\frac{\partial}{\partial \sigma^2} \log L = 0$

$\Rightarrow \underline{\hat{\mu} = \bar{x}}$ $\Rightarrow \underline{\hat{\sigma}^2 = s^2}$ //

[redacted] Unit-4 (Point Estimation)

Some Problems on Maximum Likelihood Estimators

Pb:- Find the maximum likelihood estimate for the parameter λ of a Poisson distribution.

Sol^n: The probability function of the Poisson distribution with parameter λ is -

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

Likelihood function is -

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

$$= \frac{e^{-\lambda n}\lambda^{x_1}\cdot\lambda^{x_2}\cdots\lambda^{x_n}}{x_1!\,x_2!\cdots x_n!}$$

$$\Rightarrow L = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{x_1!\,x_2!\cdots x_n!}$$

Taking log on both sides -

$$\log L = \log e^{-n\lambda} + \log \lambda^{\sum_{i=1}^{n} x_i} - \log(x_1!\,x_2!\cdots x_n!)$$

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

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

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \Rightarrow \sum_{i=1}^{n} x_i = n\bar{x}$$

Likelihood equ^n is -

$$\frac{\partial}{\partial\lambda}\log L = 0$$

$$\frac{\partial}{\partial\lambda}\log L = 0$$

$$\Rightarrow -n + n\bar{x}\cdot\frac{1}{\lambda} - 0 = 0$$

$$\Rightarrow \frac{\bar{x}}{\lambda} = 1$$

$$\Rightarrow \hat{\lambda} = \bar{x}.$$

Thus, the M.L.E. for λ is sample mean $(\bar{x})$.

Pb :- 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.

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

$$= (1+\theta)x_1^{\theta}\,(1+\theta)x_2^{\theta}\cdots(1+\theta)x_n^{\theta}$$

$$= (1+\theta)^n (x_1 x_2 \ldots x_n)^{\theta}$$

$$= (1+\theta)^n \left(\prod_{i=1}^{n} x_i\right)^{\theta}$$

$$\Rightarrow \log L = \log(1+\theta)^n + \log\left(\prod_{i=1}^{n} x_i\right)^{\theta}$$

$$= n\log(1+\theta) + \theta\log\prod_{i=1}^{n} x_i$$

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

Now, the likelihood equation is —

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

$$\Rightarrow n\cdot\frac{1}{1+\theta} + \sum_{i=1}^{n}\log x_i = 0$$

$$\Rightarrow \frac{n}{\ldots} = -\sum_{i=1}^{n}\log x_i$$

$$\Rightarrow \frac{n}{1+\theta} = -\sum_{i=1}^{n} \log x_i \checkmark$$

$$\Rightarrow \frac{1+\theta}{n} = -\frac{1}{\sum_{i=1}^{n} \log x_i} \checkmark$$

$$\Rightarrow 1+\theta = -\frac{n}{\sum_{i=1}^{n} \log x_i}$$

$$\Rightarrow \hat{\theta} = -1 - \frac{n}{\sum_{i=1}^{n} \log x_i} \checkmark$$

$$\Rightarrow \boxed{\hat{\theta} = -1 - \frac{n}{\log\left(\prod_{i=1}^{n} x_i\right)}}$$ // is the M.L.E for θ.

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

$\Rightarrow \theta = \hat{\theta}$ ↓ MLE for θ

Pb :- Let T be the M.L.E for parameter θ of a population. Then–

(a) $\mathcal{E}(T) < \mathcal{E}(T_1)$, T_1 - any other estimator of θ.

(b) $\mathcal{E}(T) = \frac{1}{2}$.

(c) T is an unbiased estimator of θ.

(d) T is a consistent estimator of θ.

$\mathcal{E}(T) = 1$

(a) $Var(T) = Var(T')$ ✗

(b) $Eff(T') = 1$ ✗

(c) $Eff(T) > 1$ ✗

(d) $Var(T) < Var(T')$ ✓

Pb :- If T is the most effecient estimator and T' is any other estimator of parameter θ, then which is true?

(i) $\mathcal{E}(T) < \mathcal{E}(T')$ ✗

(ii) $\mathcal{E}(T) + \mathcal{E}(T') < 1$ ✗

(iii) $\mathcal{E}(T) > 0$ $\quad = 1$

(iv) $\mathcal{E}(T') < 0$

$\mathcal{E} \in (0, 1]$

$0 < \mathcal{E} \leq 1$.

$\mathcal{E}(T') = \frac{V(T)}{V(T')}$

$\to 0$

$\mathcal{E}(T) > 0$ $(=1)$

$x^{(iv)}$ $\mathcal{E}(T') < 0$

$\mathcal{E}(T') = \frac{V(\cdot)}{V(T')}$

> 0

$\mathcal{E}(T) = 1$

$\mathcal{E}(T') > 0$

$\mathcal{E}(T) + \mathcal{E}(T') > 1$.

Pb If T is an unbiased estimator of parameter θ, then—

(i) $\mathcal{E}(T) > \theta$ $\underline{\underline{+ve}}$

(ii) $\mathcal{E}(T) < \theta$ $\underline{\underline{-ve}}$

(iii) $\mathcal{E}(T) - \theta = 0$

(iv) None of these

$\underline{\underline{\mathcal{E}(T)}} = \theta$

$\mathcal{E}(T) - \theta = 0$

$\Rightarrow \mathcal{E}(T) = \theta$

Lecture 32

[redacted] Unit-4 (Point Estimation)

4) Sufficiency (sufficient Estimators) :-

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

→ Let $x_1, x_2, \ldots, x_n$ be a random sample of size n drawn from a population with p.m.f or p.d.f $f(x, \theta)$.

If the conditional distribution of sample values $x_1, x_2, \ldots, x_n$ given T i.e., $f(x_1, x_2, \ldots, x_n \backslash T)$ is independent of θ, then T is sufficient estimator for θ.

- If the conditional distribution of sample values $x_1, x_2, \ldots, x_n$ given (T_1, T_2) i.e., $f(x_1, x_2, \ldots, x_n \backslash (T_1, T_2))$ is independent of θ, then (T_1, T_2) is called 'joint sufficient estimator of θ.

§ A single sufficient statistic (estimator) may or may not exist but a joint sufficient estimator will always exist for a parameter θ.

$\dfrac{f(x_1, x_2 \ldots x_n / T)}{\neq g(\theta)}$

Defn: If $T = t(x_1, x_2, \ldots, x_n)$ is an estimator of a parameter θ, based on a sample $x_1, x_2, \ldots, x_n$ of size n from a population with density function $f(x, \theta)$ such that the conditional distribution of $x_1, x_2, \ldots, x_n$ given T, is independent of θ, then T is sufficient estimator of θ. ✓

i.e., $\boxed{f(x_1, x_2, \ldots, x_n \backslash T = k) \text{ is independent of } \theta.}$

§ Illustration: Let $x_1, x_2, \dots, x_n$ be a random sample from Bernoulli population with parameter 'p', $0<p<1$ i.e.,

$$x_i = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with prob. } q = 1-p \end{cases}$$

Then $T = t(x_1, x_2, \dots, x_n) = x_1 + x_2 + \dots + x_n = \sum_{i=1}^{n} x_i \sim B(n,p)$

$$\therefore P(T=k) = \binom{n}{k} p^k (1-p)^{n-k} \; ; \; k = 0,1,2,\dots,n.$$

So, the conditional distribution of sample values given $T=k$ is —

$$P(x_1 \cap x_2 \cap \dots \cap x_n \mid T=k) = \frac{P(x_1 \cap x_2 \cap \dots \cap x_n \cap T=k)}{P(T=k)}$$

$$= \begin{cases} \dfrac{\cancel{p^k}\,\cancel{(1-p)^{n-k}}}{\binom{n}{k}\cancel{p^k}\,\cancel{(1-p)^{n-k}}}, & \sum_{i=1}^{n} x_i = k \\ 0, & \sum_{i=1}^{n} x_i \neq k \end{cases}$$

$\binom{n}{k} = {}^nC_k$

$$= \begin{cases} \dfrac{1}{\binom{n}{k}}, & \sum_{i=1}^{n} x_i = k \\ 0, & \sum_{i=1}^{n} x_i \neq k \end{cases}$$

p

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

Next class

§ Neymann Factorization Theorem:- $T = t(x)$ is sufficient for θ if and only if the joint density function L of sample values and only if the joint density function L of sample values can be expressed in the form –

$$L = \underbrace{g_\theta[t(x)]}_{x,\theta} \cdot \underbrace{h(x)}_{x}, \text{ where } g_\theta[t(x)] \text{ depends on}$$

θ & x only through the value of $t(x)$ and $h(x)$ is independent of θ.

$g_\theta[t(x)]$

$g(x,\theta)h(x)$

Remarks: (i) $h(x)$ is a function independent of θ. That is, it should not include θ and its domain also do not contain θ.

$g_\theta(x) \cdot h(x)$

i.e., the fun^c $h(x) = \frac{1}{2a}$, $a-\theta < x < a+\theta$, $-\infty < \theta < \infty$, is not independent of θ.

(ii) The original sample $X = (X_1, X_2, \ldots, X_n)$ is always statistic.

(iii) Invariance property of sufficient estimators:

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 estimator of $\psi(\theta)$.

* Fisher-Neymann 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 sample values) can be expressed as: –

$$L = \prod^{n} f(x_i;\theta) = g_1(t_1,\theta)\, h(x_1, x_2, \ldots, x_n),$$

$$L = \prod_{i=1}^{n} f(x_i;\theta) = \underline{g_1(t_1,\theta)}\, h(\underline{x_1, x_2}, \dots, x_n),$$

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

Ex- Let $X_1, X_2, \dots, X_n$ be a random sample from a population with p.d.f $f(x,\theta) = \theta x^{\theta-1}$; $0<x<1$, $\theta>0$.

Check whether $t_1 = \prod_{i=1}^{n} X_i$, is sufficient for θ or not.

Solⁿ: Here, likelihood function is given by—

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

$$= \prod_{i=1}^{n} \theta x_i^{\theta-1}$$

$$= (\theta x_1^{\theta-1}) \cdot (\theta x_2^{\theta-1}) \dots (\theta x_n^{\theta-1})$$

$$= \theta^n x_1^{\theta-1} \cdot x_2^{\theta-1} \dots x_n^{\theta-1} = \theta^n (x_1 x_2 \dots x_n)^{\theta} (x_1 \dots x_n)^{-1}$$

$$= \theta^n \prod_{i=1}^{n} x_i^{\theta-1}$$

$$= \theta^n \prod_{i=1}^{n} x_i^{\theta} \cdot \prod_{i=1}^{n} x_i^{-1}$$

$$= \left(\theta^n \prod_{i=1}^{n} x_i^{\theta}\right) \cdot \frac{1}{\prod_{i=1}^{n} x_i}$$

$$\therefore L(x,\theta) = g(t_1,\theta) \cdot h(x_1, x_2, \dots, x_n)$$

Hence, by Factorization Theorem, $t_1 = \prod_{i=1}^{n} x_i$ is sufficient estimator for θ.

Pb:- Let $X_1, X_2, \ldots, X_n$ be a random sample from Cauchy population $f(x,\theta) = \frac{1}{\pi} \cdot \frac{1}{1+(x-\theta)^2}$; $-\infty < x < \infty$, $-\infty < \theta < \infty$.

Examine if there exists a sufficient statistic for θ.

YES NO

Solⁿ:

$$L(x,\theta) = \prod_{i=1}^{n} f(x_i,\theta) = \prod_{i=1}^{n} \frac{1}{\pi} \cdot \frac{1}{1+(x_i-\theta)^2}$$

$$= \frac{1}{\pi^n} \prod_{i=1}^{n} \frac{1}{1+(x_i-\theta)^2}$$

$$= \frac{1}{\pi^n} \left\{ \frac{1}{1+(x_1-\theta)^2} \cdot \frac{1}{1+(x_2-\theta)^2} \cdots \frac{1}{1+(x_n-\theta)^2} \right\}$$

$$\Rightarrow L(x,\theta) \neq g(t,\theta) \cdot h(x_1, x_2, \ldots, x_n)$$

Thus, there does not exists a single sufficient estimator for θ.