Unit - II

Point Estimation

Statistics

- Descriptive
- Inferential
  - Theory of estimation
    - Point estimation
    - Interval estimation
  - Testing of hypothesis

Estimate:- An estimate is a statement made to find an unknown parameter i.e μ, σ are the parameters.

Estimator: The method to determine unknown parameters by estimating is called estimators.
i.e: $\theta = \hat{\theta}$

Estimation:- The process of determining estimators in such a way that they are close to parameters value is called as estimation.

* Estimation can be done in two ways
1. Point Estimation
2. Interval Estimation

Point Estimation:- A point estimate is a number obtained from computation on observed value of random variables which serve as an approximation to the parameter.

Defination:- Point estimation is a single valued estimation and it is also called as estimation of parameter.

* Point estimator θ is denoted as $\hat{\theta}$.
* If $E(\hat{\theta}) = \theta$ then it is unbaised estimation
* s^2 is an unbaised estimator of σ^2.

Methods of point estimation:-

There are four methods to find out point estimation. They are:-

1. Methods of Moments
2. Method of Maximum likelihood estimation (MLE)
3. Method of least Squares
4. Bayesian method

Method of Moments: This method was discovered and studied indetail by Karl Pearson.

Let $f(x; \theta_1, \theta_2, ---- \theta_k)$ be the density function of parent population with 'k' parameters $\theta_1, \theta_2, \theta_3, -- \theta_k$ if μ_r' denotes r^{th} moment about origin then $\mu_r' = \int_{-\infty}^{+\infty} x^r f(x : \theta_1, \theta_2, -- \theta_k)\, dx$, $(r = 1, 2, -- k)$. In general $\mu_1', \mu_2', ---- \mu_r'$ are the functions of parameters $\theta_1, \theta_2, --- \theta_k$.

let x_i, where $i=1,2,\dots n$ be a random sample of size n from the given population then the method of moments is given in terms of $\theta_1, \theta_2, \dots \theta_K$ of $\mu_1', \mu_2', \dots \mu_r'$ then replacing these moments μ_r'; $r=1,2,\dots k$.

Ex: $\hat{\theta}_i = \theta_i(\hat{\mu}_1, \hat{\mu}_2, \dots \hat{\mu}_k)$

$= \theta_i(m_1', m_2', \dots m_k')$; $i=1,2,\dots k$

where m_i - is the m^{th} moment about origin, in sample. Then by the method of moments $\hat{\theta}_1, \hat{\theta}_2, \dots \hat{\theta}_K$ are the required estimators of $\theta_1, \theta_2, \dots \theta_K$ respectively.

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

* Generally the method of moments are less efficient estimators than the principle of maximum likelihood estimators.

Method of maximum likelihood estimation:-

In theoretical point of view, the most general method of estimation is the method of maximum likelihood estimation. (MLE) which was introduced by professor RA Fisher. He was developed in a series of methods for find outing estimation.

Likelihood function:- let $x_1, x_2, --- x_n$ be a random sample of size n from a population with density function $f(x, \theta)$ then the likelihood function of small values $x_1, x_2, .. x_n$ this can be denoted $L = L(\theta)$ and joint density function is given by $L = f(x_1, \theta)\, f(x_2, \theta), ..\, f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$. Here L represents the relative likelihood that the random variables of values $x_1, x_2, --- x_n$. for a given sample L becomes a function of variable and θ is parameter.

* The principle of maximum likelihood function consists in finding an estimator for unknown parameters $\theta = (\theta_1, \theta_2, -- \theta_k)$ which maximizes the likelihood function $L(\theta)$ for variations in parameter that can be written as $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, -- \hat{\theta}_k)$ so that $L(\hat{\theta}) > L(\theta)\ \forall\ \theta \in \Theta$. Thus if there exists a function $\hat{\theta} = \hat{\theta}(x_1, x_2, -- x_n)$ of sample values which maximizes L for variations in θ then $\hat{\theta}$ is to be taken as a estimator of θ, where $\hat{\theta}$ is usually called as maximum likelihood estimator.

Properties of maximum likelihood estimator:-

The first and second order derivatives $\left(\frac{dL}{d\theta}\right)\left(\frac{d^2L}{d^2\theta}\right)$ exists and are continuous functions of θ in range 'R'

For every θ in R is $\left|\frac{d}{d\theta}\log L\right| < F_1(x)$ and

$\left|\frac{d^2}{d\theta^2}\log L\right| < F_2(x)$ where

$F_1(x)$ and $F_2(x)$ are integral functions having limits $(-\infty$ to $+\infty)$. Here the third order derivative $\frac{d^2}{d\theta^3}\log L$ exists such that $\left|\frac{d^2}{d\theta^3}\log L\right| \le M_1(x)$ where $M_1(x)$ is K (constant or +ve No)

Theorem:-

If a sufficient estimator exists then it is a function of maximum likelihood estimator.

Proof:- If $t = t(x_1, x_2, \ldots x_n)$ is a sufficient estimator of θ then likelihood function can be written as

$$L = g(t, \theta)\, h(x_1, x_2, x_3, \ldots x_n / t) \quad \text{(1)}$$

where $g(t,\theta)$ is the density function of t and $h(x_1, x_2, \ldots x_n)/t)$ is the density function of sample and the t is given and is independent of θ.

By applying log on both sides to the eqn ① then we get, $\log L = \log \{ g(t, \theta) + \log h(x_1, x_2, \ldots x_n | t)$.

differentiating w.r.t θ we get,

$$\frac{d \log L}{d\theta} = \frac{d}{d\theta} \log(g)(t, \theta) = \phi(t, \theta) \text{ (say)}$$

which is a function of (t, θ) only.

- Maximum likelihood estimator is given by

$$\frac{d \log L}{d\theta} = 0 \Rightarrow \phi(\hat{t}, \hat{\theta}) = 0.$$

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

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

Hence, the theorem is exists i.e $\hat{\theta}$ is a replaced by θ and it is taken as sufficient estimator. and $\hat{t}$ is replaced by t and it is taken as function of maximum likelihood estimator.

1) Prove that maximum likelihood estimation of parameter α of a population having density function:

$\frac{2}{\alpha^2}(\alpha-x)$, $0<x<\alpha$ for a sample of unit size is $2x$, x is an sample value show that x is biased.

For a random sample of unit size $n=1$ & the likelihood function is $L(\alpha)=f(x,\alpha)=\frac{2}{\alpha^2}(\alpha-x)$ $0<x<\alpha$ by these likelihood equation,

$$\frac{d}{d\alpha}\log L=\frac{d}{dx}\left[\log\frac{2}{\alpha^2}\overbrace{(\alpha-x)}^{b}\right] \quad \left[\because \log\frac{a}{b}=\log a-\log b\right]$$

$$\frac{d}{d\alpha}\left[\log\frac{2}{\alpha^2}+\log(\alpha-x)\right]=0$$

$$\frac{d}{d\alpha}\left[\log 2-\log\alpha^2+\log(\alpha-x)\right]=0$$

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

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

$$\alpha=2(\alpha-x)$$

$$\alpha=2\alpha-2x$$

$$\alpha-2\alpha=-2x$$

$$\cancel{+}\alpha=\cancel{+}2x$$

$$\boxed{\alpha=2x} \quad \therefore \hat{\alpha}=2x$$

$\therefore$ Now $E(\hat{\alpha}) = E(2x)$

$$= 2\int_0^{\alpha} x \cdot f(x,\alpha)\, dx$$

$$= 2\int_0^{\alpha} x \cdot \frac{2}{\alpha^2}(\alpha - x)\, dx$$

$$= \frac{4}{\alpha^2}\int_0^{\alpha} x(\alpha - x)\, dx$$

$$= \frac{4}{\alpha^2}\int_0^{\alpha} (x\alpha - x^2)\, dx$$

$$= \frac{4}{\alpha^2}\left| \frac{\alpha x^2}{2} - \frac{x^3}{3} \right|_0^{\alpha}$$

$$= \frac{4}{\alpha^2}\left| \frac{\alpha^3}{2} - \frac{\alpha^3}{3} \right|$$

$$= \frac{4}{\alpha^2}\left(\frac{3\alpha^3 - 2\alpha^3}{6} \right)$$

$$= \frac{4}{\alpha^2}\left| \frac{\alpha^3}{6} \right|$$

$$\hat{\alpha} = \frac{2\alpha}{3}$$

$\therefore \hat{\alpha} = 2x$ is not baised estimate of x.

2) let $x_1, x_2, \ldots x_n$ denote random sample of size n form a uniform population with probability density function (pdf)

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

obtain maximum likelihood estimation for θ.

The maximum likelihood function ~~is~~

$$L = L(\theta; x_1, x_2, \ldots x_n) = 1$$

$$\theta - \frac{1}{2} \leq x \leq \theta + \frac{1}{2}$$

If $x_1, x_2, \ldots x_n$ is an ordered sample then

$$\theta - \frac{1}{2} \leq x_1 \leq x_2 \leq \ldots x_n \leq \theta + \frac{1}{2}.$$

Here L attends maximum

$$\theta - \frac{1}{2} \leq x_1 ; \quad x_n \leq \theta + \frac{1}{2}$$

$$\theta \leq x_1 + \frac{1}{2} ; \quad x_n - \frac{1}{2} \leq \theta$$

Hence every statistics $t = t(x_1, x_2, \ldots x_n)$ such that $x_n - \frac{1}{2} \leq t(x_1, x_2, \ldots x_n) \leq x_n + \frac{1}{2}$

It provides maximum likelihood estimate θ as - $x_n - \frac{1}{2}$ & $x_1 + \frac{1}{2}$

nce:-

Method of Minimum Variance:-

It can be written as MBVE (Minimum Variance Unbaised Estimator). These can be study it by using maximum likelihood estimate of mean and variance that is

i) unbaised ii) Minimum variance.

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

$$t = t(x_1, x_2, \ldots x_n)$$

Such that $E(t) = \int_{-\infty}^{\infty} t \cdot L \, dx$

$= \gamma(\theta)$ (say)

$\int_{-\infty}^{\infty} [t - \gamma(\theta)] L \cdot dx$

variance of $v(t) = \int_{-\infty}^{\infty} [t - \gamma(\theta)]^2 L \, dx$

Here integral $-\infty$ and ∞ represents n-fold integration. For these a detailed information is given in Cramer Roo inequality.

Cramer Rao Inequality:- With probability approaching unity as n tends to infinity the likelihood equation $(n \to \infty)$

$\frac{\partial}{\partial \theta} \log L = 0$ has solution which converge. In probability to the true value θ_0. In otherwords maximum likelyhood estimation are consistent.

For Ex:- A sample is drawn from a $N(\mu, \sigma^2)$ population Then $M.L.E(\mu) = \bar{x}$ (sample mean)

$M.L.E(\sigma^2) = s^2$ (sample variance)

Here sample mean is $\bar{x}$ is both unbaised and consistent and s^2 is only consistence but not unbaised.

Method of least Squares:-

The principle of least squares is used to fit a curve of the form $y = (x, a_0, a_1, -- a_n)$ where a_i's are unknown parameters to set n sample observations $n(x_i, y_i)$ where $i = 1, 2, -- n$ from a bio variate population it consists of minimising the sum of squares of Residuals.

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

subject to the variations in $a_0, a_1, -- a_n$.

The normal eqns for substituting $a_0, a_1, a_2 -- a_n$ are given by $\frac{dE}{da_i} = 0$ $i = 1, 2, -- n$

This is also called as estimating of normal eqns.

Method of minimum chi-square:-

In statistic the minimum chi-square estimation is a method of estimation of unbaised quantities based on observed data.

In certain chi-square test rejects a null hypothesis about a population distribution if a specified test statistic is too large. when that statistics would have approximately a chi-square distribution if the null hypothesis is true. In this estimation we have to find the values of parameters of statistic that makes the test statistic as small as possible.

The chi-square test can be solved by using formula as $\chi^2 = \sum_{i=1}^{n} (O_i - E_i)$

Here O represents obsened data of sample
E represents expected values of sample

This test is not valid for small tests of sample.

Modified Minimum Chi-Square :-

let $x_1, x_2, --- x_n$ be n sample observations with obsened frequencies $o_1, o_2, -- o_k$ respectively. Assume that these observations are grouped into k - classes.

let $p_1, p_2, -- p_k$ be k - unknown probabilities for k classes which are functions of unknown parameters $\theta = (\theta_1, \theta_2, -- \theta_k)$ then $p_i = P_i(\theta)$ for $i = 1, 2, -- k$. By the defination of expected frequencies for k classes are given as $e_1, e_2, -- e_k$ where $e_i = np_i$ and $n = \sum_{i=1}^{n} o_i$. A measure of distance between obsened and expected frequencies is supplied by the statistic chi-square is given by $\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \cdots + \frac{(o_i - e_i)}{e_i}$

$= \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i}$ where $e_i = np_i$ and $n = \sum^{n} o_i$

In linear model, the modified chi-square statistic is given by $\chi^2 = \sum_{i=1}^{k} \frac{(e_i - o_i)^2}{o_i}$

Method:-

The minimum chi-square method provides some computational difficulties for estimating parameters since o_i's are occuring in the denominators of minimum chi-square equation. In such cases one can use the method of modified minimum chi-square. The modified chi-square is given as ~~chi-s~~ $\chi^2 = \sum_{i=1}^{k} \frac{(np_i - o_i)^2}{o_i}$

Consider the likelihood function then

$$L = \frac{n!}{\prod_{i=1}^{n} o_i!} \prod_{i=1}^{k} p_i^{o_i}$$

$$= \frac{n!}{\prod_{i=1}^{n} o_i!} \prod_{i=1}^{k} \left(\frac{np_i}{o_i}\right)^{o_i} \prod_{i=1}^{k} \left(\frac{o_i}{n}\right)^{o_i}$$

$$\left[\because p_i^{o_i} = \frac{np_i}{o_i} \cdot \frac{o_i}{n}\right]$$

Taking log on both sides

$\log L = c + \sum_{i=1}^{n} o_i \cdot \log\left(\frac{np_i}{o_i}\right)$ where c - is independent of p_i's. For large samples assume that $np_i = o_i + c_i\sqrt{n}$ where c_i's are small compared to o_i's, and $\sum_{i=1}^{k} c_i = 0$.

Hence $\log L = C + \sum_{i=1}^{k} O_i \log\left(1 + \frac{c_i n^{1/2}}{O_i}\right)$

$$= C + \sum_{i=1}^{k} O_i\left[\frac{c_i n^{1/2}}{O_i} - \frac{c_i^2 n}{2O_i^2} + \frac{c_i^3 n^{3/2}}{3O_i^3} \cdots\right.$$

$$= C - \frac{1}{2}\sum_{i=1}^{k} \frac{(np - O_i)^2}{O_i} + O(n^{-1/2})$$

$$= C - \frac{1}{2}(\chi'^2) + O(n^{-1/2})$$

Therefore, if we neglect the maximization of $\log L$ amongst the minimization of χ'^2

Asymptotic Normality of MLE:-

A consistent solution of likelihood eqn is called asymptotically normally of MLE It is normally distributed about the true value θ_0. Thus $\hat{\theta} = N\left[\theta_0, \frac{1}{I(\theta_0)}\right)$ as $n \to \infty$ variance of MLE is given by

$$V(\hat{\theta}) = \frac{1}{I(\theta_0)} = \frac{1}{\left[E\left(-\frac{d^2}{d\theta^2}\log L\right)\right]}$$

Theorem:-

1) If a sufficient estimator exists, it is a function of maximum likelihood estimator.

Proof:- If $t = ~~(t_1, t_2.~~~ t(x_1, x_2 \cdots x_n)$ is a sufficient estimator of θ then likelihood function can be written as

$L = g(t,\theta)\, h(x_1, x_2, \dots x_n | t)$ where $g(t,\theta)$ is a density function of t and $h(x_1, x_2, \dots x_n | t)$ is a density function of sample i.e given 't' and is independent of θ.

Apply log on b.s

~~$\log L = \log g(\theta$~~

$$\log L = \log g(t,\theta) + \log h(x_1, x_2 \dots x_n | t)$$

differentiating w.r.t "θ"

$$\frac{d \log L}{d\theta} = \frac{d}{d\theta} \log g(t,\theta) = \phi(t,\theta) \text{ (say)}$$

which is a function of t and θ only.

* MLE is given by $\frac{d \log L}{d\theta} = 0$ $[\because \phi(t,\theta) = 0]$

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

$\hat{t} = \phi(\theta) =$ some function of MLE

Hence the theorem is proved.

Invariance property of MLE :-

If T_r is MLE of θ and $\phi(\theta)$ is one-one function of θ and $\phi(T)$ is MLE of $\phi(\theta)$.

Ex:- 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.

sol:-

Given $X \sim N(\mu, \sigma^2)$ then maximum likelihood estimation $L = \prod_{i=1}^{n} \left[\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ \frac{1}{2\sigma^2}(x_i - \mu)^2 \right\} \right]$

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

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

Case i:- when σ^2 is known

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

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

$$\therefore \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Hence, MLE for μ is sample mean $(\bar{x})$.

case ii :- when μ is known the likelihood equation for estimator σ^2 is

diff w.r.t σ^2

$$\frac{d\log L}{d\sigma^2} = \cancel{-n}$$

$$\frac{d\log L}{d\mu^2} = 0$$

$$= -\frac{n}{2} \times \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \Sigma(x_i-\mu)^2 = 0$$

$$\Rightarrow \frac{1}{2\sigma^{4}{}_{2}} \Sigma(x_i-\mu)^2$$

$$\frac{n}{2\sigma^2} \Sigma(x_i-\mu)^2$$

$$\frac{\Sigma(x_i-\mu)^2}{\sigma^2} = n$$

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

case iii :- The likelihood eqn's for simultaneous estimation of μ and σ^2 are

$$\frac{d\log L}{d\mu} = 0 \Rightarrow \frac{d\log L}{d\sigma^2} = 0$$

$$\Rightarrow \hat{\mu} = \bar{x} \text{ and } \hat{\sigma}^2 = \frac{\Sigma(x_i-\mu)^2}{n}$$