Unit - II

POPULATION: The set of all possible observation under study is called population. It is denoted by 'N'.

PARAMETER: Any population constraints are called parameter. For example, A random variable X~N$\left(\mu,\sigma^2\right)$, here μ *and* σ^2 are called the parameters of Normal Distribution.

PARAMETRIC SPACE: The set of all possible values of the parameter is called parametric space. (i.e) X~f$\left(x:\theta\right) \forall\ \theta\, \varepsilon \Theta$. It is denoted as '$\Theta$'. For example: $X \sim N\left(\mu,\sigma^2\right)\ \forall\ \Theta = \left\{\theta = \left\{\mu,\sigma^2\right\}; -\infty < \mu < \infty,\ \sigma > 0\right\}$

SAMPLE: It is the subset of the population. It is denoted by 'n'.

Definition 1: ESTIMATOR

Any function of random samples $x_1, x_2,\ldots,x_n$ that are being observed say $T_n\left(x_1, x_2,\ldots,x_n\right)$ is called an estimator. Clearly a estimator is also a random variable.

If it is used to estimate an unknown parameter, say θ, of the distribution which is also called an estimator.

Definition 2: ESTIMATE

A particular value of the estimator is called estimate of parameter, say θ Eg: $x_1, x_2, \ldots, x_n$ is a random sample then the mean of the random sample is $\bar{x}$ say T(x)= $\bar{x}$ is called estimator.

1.1 SAMPLING DISTRIBUTION

If a number of samples, each of size n (viz., each sample containing n elements) are drawn from the sample population and if for each sample the values of some statistic say, mean is calculated, a set of values of the statistic will be obtained. These values o the statistic will usually vary from one sample to another, as the values of the population members included in different samples, through drawn from the same population may be different and hence may be treated as values of R.V.

The probability distribution of the statistic (a R.V.) that would be obtained, if the number of samples, each of the size n, were infinitely large, is called the sampling distribution of the statistic. If the random sampling technique is adopted, the nature of the sampling distribution of a statistic can be obtained theoretically, using the theory of probability provided the nature of the population distribution is known.

Like any other distribution, a sampling distribution will have its mean, standard deviation and moments of higher order. The standard deviation of the sampling distribution of a statistic isof particular importance in tests of significance for large samples and testing of hypothesis and is known as standard error (S.E). In the case of large samples (viz. n>30), the sampling distribution of many statistics tend to become normal distributions.

If t is a statistic in large samples, then t follows a normal distribution with mean E(t), which is the corresponding population parameter and S.D. equal to S.E.(t). Hence $Z = \frac{t - E(t)}{S.E.(t)}$ is a standard normal variate.Z follows a normal distribution with mean 0 and S.D. 1 and is called the test statistic.

1.2 STANDARD ERRORS

The standard errors of some frequently occurring statistics for large samples of size n are given below, where σ^2 is the population variance, P, the population proportion and Q=1-P and n_1, n_2 represent the sizes of two independent random samples drawn from the given population(s).

S.No.	Statistic	Standard Error
1	Sample Mean $(\bar{X})$	$\frac{\sigma}{\sqrt{n}}$
2	Sample Proportion(p)	$\sqrt{PQ/n}$
3	Sample S.D. (s)	$\sqrt{\sigma^2/2n}$
4	Sample Variance (s^2)	$\sigma^2\sqrt{2/n}$
5	Sample Median	$1.25331\,\sigma/\sqrt{n}$
6	Sample coefficient of Variation (υ)	$\frac{\upsilon}{\sqrt{2n}}\sqrt{1+\frac{2\upsilon^3}{10^4}} \approx \frac{\upsilon}{2n}$
7	Sample Correlation Coefficient (r)	$\frac{(1-\rho^2)}{\sqrt{n}}$ where ρ is the population correlation coefficient.
8	Differences of two sample means $(\bar{X}_1 - \bar{X}_2)$	$\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
9	Difference of two sample S.D's (s_1-s_2)	$\sqrt{\frac{\sigma_1^2}{2n_1}+\frac{\sigma_2^2}{2n_2}}$
10	Difference of two sample proportions (p_1-p_2)	$\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}}$

Definition 3: ONE PARAMETER EXPONENTIAL FAMILY OF DISTRIBUTION

A random variable $X_1, X_2, \ldots, X_n$ said to be distributed according to a member in one parameter exponential family of distributions if its probability density function is expressed as

$$f(x|\theta) = e^{A(\theta)T(x) - B(\theta)} h(x) \forall x \in \aleph, \theta \in \Theta$$

where $A(\theta)$and $B(\theta)$are real valued function of θ, T(x) is a real valued statistic with support x and h(x) is independent of θ.

Definition 4: MULTI PARAMETERS EXPONENTIAL FAMILY OF DISTRIBUTION

A random variables, $X_1, X_2,...,X_n$ which is equal to x_1, x_2,..., x_n is said to be distributed according to a member of multi parameters exponential family of distributions if its probability density function is expressed as

$$f(x_i|\theta)=e^{\sum_{i=1}^{n}A_i(\theta)T_i(x)-B(\theta)}h(x)\forall x\in\aleph,\theta_1,\theta_2..\theta_n\in\Theta$$

where $\sum_{i=1}^{n}A_i(\theta)$and $B(\theta)$are real valued function of θ, $\sum_{i=1}^{n}T_i(x)$ is a real valued statistic with support x and h(x) is independent of θ.

Example 1: Let $X_1, X_2,...,X_n$ ~ Poisson(θ). To check whether, this distribution is a member in one parameter exponential family of distribution.

Solution:

$$P(X=x|\theta)=\begin{cases}\dfrac{e^{-\theta}\theta^x}{x!}, x=0,1,2,\ldots,\theta>0\\ 0, \quad otherwise\end{cases}$$

$$P(X=x|\theta)=e^{-\theta}\, e^{x\log\theta}.\frac{1}{x!}$$

$$P(X=x|\theta)=e^{x\log\theta-\theta}.\frac{1}{x!}$$

Here, $A(\theta)=\log\theta; B(\theta)=\theta; T(x)=x; h(x)=\dfrac{1}{x!}$

Therefore, Poisson distribution is a member in one parameter exponential family of distribution.

Example 2: To check whether, the normal distribution is a member in exponential family of distribution.

Solution:

The Probability density function of normal distribution is

$$f\left(x\middle|\mu,\sigma^2\right)=\begin{cases}\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} & ;-\infty < x,\mu < \infty,\sigma > 0\\ 0, otherwise\end{cases}$$

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

$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}+\frac{\mu x}{\sigma^2}-\frac{\mu^2}{2\sigma^2}-\log\sigma}$$

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

Here, $A(\theta)=\dfrac{-1}{2\sigma^2};B(\theta)=\dfrac{\mu^2}{2\sigma^2}+\log\sigma;T\left(x_1\right)=x^2;T\left(x_1\right)=x;h(x)=\dfrac{1}{\sqrt{2\pi}}$

Therefore, Normal distribution is a member in multi parameter exponential family of distribution.

1.3 IDEAL/PROPERTIES/ CHARACTERISTICS OF AN ESTIMATOR:

Estimation theory is concerned with the properties of estimators (i.e) with defining properties that can be used to compare different estimator for the same quantity based on the same data. Such properties can be used to determine the best rules to use under given circumstance.

The properties of estimators are mainly classified into two, small sample and large sample properties.

Estimators

Small

(i) Unbiasedness

(ii) Efficiency

(iii) Sufficiency

Large

(iv) Consistency

Definition 5: UNBIASEDNESS

The statistic $T_n = T(x_1, x_2, ..., x_n)$ will be called an unbiased estimator of $\gamma(\theta)$ if $E_\theta(T(x)) = \gamma(\theta) \;\; \forall \;\; \theta \; \varepsilon \; \Theta$. (i.e) It has zero bias $\forall \theta$. ($E_\theta(T(x)) - \theta = 0$)

Definition 6: **BIAS**

Let $T_n = T(x_1, x_2, ..., x_n)$ is a biased estimator then $E_\theta(T(x)) - \theta = b(\theta)$. Here $b(\theta)$ is amount of bias.

Remarks:

- $E_\theta(T(x)) > \theta$ then bias is positive
- $E_\theta(T(x)) < \theta$ then bias is negative

Mean Square Error: Let $T_n = T(x_1, x_2, ..., x_n)$ be an estimator of $\gamma(\theta)$. The mean square error of the estimator $\gamma(\theta)$ is defined as $E_\theta[T - \gamma(\theta)]^2$

(i.e) $E_\theta[T - \gamma(\theta)]^2 = E_\theta[T - \gamma(\theta) + E_\theta(T) - E_\theta(T)]^2$

$$= E_\theta\{[E_\theta(T) - \gamma(\theta)] + [T - E_\theta(T)]\}^2$$

$$= E\left\{(E_\theta(T) - \gamma(\theta))^2 + (T - E_\theta(T))^2 + 2(E_\theta(T) - \gamma(\theta))(T - E_\theta(T))\right\}$$

$$= E\;\{E_\theta(T) - \gamma(\theta)\}^2 + E\{T - E_\theta(T)\}^2$$

$$E_\theta[T - \gamma(\theta)]^2 = b^2\gamma(\theta) + Var(T)$$

$$= \left\{\begin{matrix} bias\,is \\ the\ estimator \end{matrix}\right\} + \left\{\begin{matrix} \text{var}\,iability\ of \\ the\ estimator \end{matrix}\right\}$$

$$=\begin{Bmatrix} accuracy\ of \\ the\ estimator \end{Bmatrix}+\begin{Bmatrix} precision\ of \\ the\ estimator \end{Bmatrix}$$

An estimator is preferred over others if it is a MSE is small as compared to that of others which is achieved by the small variance and small biased both together. Controlling over biased does not necessarily result in low mean square error. Sometimes bearing small amount of biased combined with decrease in variance finally that result into a high decreasing mean square error.

Small mean square error of the estimator results in high probability that the estimator too close to true value of parameter θ by chebyshev's inequality.

The Positive square root of mean square error is called standard error. Mean squared error (MSE) combines the notions of bias and standard error. It is defined as

$$\text{MSE}=(\text{Standard Error})^2+(\text{Bias})^2$$

Example 3: Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal population $\mathrm{N}(\mu, 1)$. Show that $t=\frac{1}{n}\sum_{i=1}^{n} x_i^{\,2}$ is an unbiased estimator of μ^2+1.

Solution:

Let, $X_1, X_2, \ldots, X_n \sim \mathrm{N}(\mu, 1)$, $E(x_i)=\mu \text{ and } V(x_i)=1$

w.k.t; $V(x_i)=E(x_i^{\,2})-[E(x_i)]^2$

$$1 = E(x_i^{\,2})-\mu^2$$

$$E(x_i^{\,2})=1+\mu^2$$

$$E\left[\frac{1}{n}\sum_{i=1}^{n} x_i^{\,2}\right]= \frac{1}{n}\sum_{i=1}^{n} E(x_i^{\,2})$$

$$= \frac{1}{n}\sum_{i=1}^{n}(1+\mu^2)$$

$$= \frac{1}{n}.n(1+\mu^2)$$

$$=1+\mu^2$$

Hence $t=\frac{1}{n}\sum_{i=1}^{n} x_i^{\,2}$ is an unbiased estimator of μ^2+1.

Example 4: If T is an unbiased estimator for θ show that T^2 is an unbiased estimator for θ^2.

Solution:

Since T is an unbiased estimator for θ (i.e) $E(T)=\theta$

w.k.t., $V(T)=E(T^2)-[E(T)]^2$

$$V(T)=E(T^2)-\theta^2$$

$$E(T^2)=V(T)+\theta^2$$

$$E(T^2)\neq\theta^2 \qquad \because V(T)>0 \left[T^2 \textit{ is a biased estimator of } \theta^2\right]$$

Example 5: Show that $\frac{\sum x_i(\sum x_i-1)}{n(n-1)}$ is an unbiased estimate of θ^2 for the sample $x_1, x_2, \ldots, x_n$ drawn an X which takes the value 1 or 0 with respective probabilities θ *and* $(1-\theta)$.

Solution:

$$X \sim Bernoulli(\theta)$$

$$T=\sum x_i \sim Binomial(\theta);\ E(T)=n\theta, V(T)=n\theta(1-\theta)$$

$$E\left[\frac{\sum x_i(\sum x_i-1)}{n(n-1)}\right] = E\left[\frac{T(T-1)}{n(n-1)}\right] = \frac{1}{n(n-1)}E(T^2-T)$$

$$= \frac{1}{n(n-1)}\left[E(T^2)-E(T)\right]$$

$$= \frac{1}{n(n-1)}\left[V(T)+[E(T)]^2-E(T)\right]$$

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

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

$$= \frac{1}{n(n-1)}n\theta^2(n-1)$$

$$E\left[\frac{\sum x_i(\sum x_i-1)}{n(n-1)}\right] = \theta^2$$

Definition 7: EFFICIENCY

If T_1 is a most efficient estimator with variance V_1 *and* T_2 is any other estimator with variance V_2 then the efficiency of T_2 is defined as $E = \frac{V_1}{V_2}$ obviously E cannot exceed 1. Similarly if T, $T_1, T_2, \ldots, T_n$ are all estimators of $\gamma(\theta)$ and variance of T is minimum then the efficiency E_i *of* $T_i (i = 1,2,\ldots,n)$ is defined as $E_i = \frac{Var\ T}{Var\ T_i}$ for all i=1,2,…,n obviously $E_i \le 1\ (i = 1,2,\ldots,n)$.

Example 6: A random sample $(X_1, X_2, X_3, X_4, X_5)$ of size 5 is drawn from a normal population with unknown mean μ. Consider the following estimators of estimate μ:

(a) $t_1 = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$

(b) $t_2 = \frac{X_1 + X_2}{2} + X_3$

(c) $t_3 = \frac{2X_1 + X_2 + \lambda X_3}{3}$

where λ is such that t_3 is an unbiased estimator of μ. Find λ. Are t_1 *and* t_2 unbiased? State giving reasons the estimator which is best among t_1, t_2 *and* t_3.

Solution:

(a) $E\left[\frac{2X_1 + X_2 + \lambda X_3}{3}\right] = \mu$ since t_3 is an unbiased estimator

$\Rightarrow \frac{1}{3} E[2X_1 + X_2 + \lambda X_3] = \mu$

$\Rightarrow \frac{1}{3}[E(2X_1) + E(X_2) + \lambda E(X_3)] = \mu$

$\Rightarrow \frac{1}{3}[2\mu + \mu + \lambda\mu] = \mu$

$\Rightarrow 3\mu + \lambda\mu = 3\mu$

$\Rightarrow \lambda = 0$

(b) $E(t_1) = E\left[\dfrac{X_1 + X_2 + X_3 + X_4 + X_5}{5}\right]$

$$= \frac{1}{5}E(X_1 + X_2 + X_3 + X_4 + X_5)$$

$$= \frac{1}{5}[E(X_1)+E(X_2)+E(X_3)+E(X_4)+E(X_5)]$$

$$= \frac{1}{5}[\mu + \mu + \mu + \mu + \mu]$$

$$= \mu$$

$$E(t_2) = E\left[\frac{X_1 + X_2}{2} + X_3\right]$$

$$= \frac{1}{2}E(X_1 + X_2) + E(X_3)$$

$$= \frac{1}{2}[E(X_1)+E(X_2)] + E(X_3)$$

$$= \frac{1}{2}[\mu + \mu] + \mu$$

$$= 2\mu$$

$\therefore t_1$ *is an unbiased estimator of* μ

t_2 *is a biased estimator of* μ

(c) $V(t_1) = E(t_1^{\,2}) - [E(t_1)]^2$

$$V(t_1) = V\left[\frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}\right]$$

$$= \frac{\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2 + \sigma^2}{25}$$

$$= \frac{\sigma^2}{5}$$

$$V(t_2) = V\left[\frac{X_1 + X_2}{2} + X_3\right]$$

$$= \frac{1}{4}V[X_1 + X_2] + V[X_3]$$

$$= \frac{\sigma^2 + \sigma^2}{4} + \sigma^2$$

$$= \frac{3\sigma^2}{2}$$

$$V(t_3) = V\left[\frac{2X_1 + X_2}{3}\right]$$

$$= \frac{4\sigma^2 + \sigma^2}{9}$$

$$= \frac{5\sigma^2}{9}$$ since $V(t_1)$ *is the least among* $V(t_2), V(t_3)$

$\therefore t_1$ *is the most efficient estimator of* μ.

Example 7: Let X_1, X_2, X_3 is a random sample of size 3 from a population with mean value μ and variance σ^2. T_1, T_2, T_3 are the estimators used to estimate mean value μ, where

$T_1 = X_1 + X_2 - X_3$ $\qquad$ $T_2 = 2X_1 + 3X_3 - 4X_2$ $\qquad$ $T_3 = \frac{1}{3}(\lambda X_1 + X_2 + X_3)$

(i) Are T_1 *and* T_2 unbiased?

(ii) Find the value of λ *such that* T_3 *is ubiased estimator for* μ

(iii) Which is the most efficient estimator?

(i) $E(T_1) = E(X_1 + X_2 - X_3)$

$= E(X_1) + E(X_2) - E(X_3)$

$= \mu + \mu - \mu$

$= \mu$

$E(T_2) = E(2X_1 + 3X_3 - 4X_2)$

$= 2E(X_1) + 3E(X_3) - 4E(X_2)$

$= 2\mu + 3\mu - 4\mu$

$= \mu$

$\therefore T_1$ is an unbiased estimator of μ

T_2 is an unbiased estimator of μ

(ii) $E(T_3) = \mu$ $\qquad$ since T_3 *is an unbiased estimator for* μ

$$E\left(\frac{1}{3}(\lambda X_1 + X_2 + X_3)\right) = \mu$$

$$\frac{1}{3}((\lambda E(X_1) + E(X_2) + E(X_3))) = \mu$$

$$\lambda\mu + \mu + \mu = 3\mu$$

$$\lambda = 1$$

(iii) $V(t_1) = V[X_1 + X_2 - X_3]$

$= \sigma^2 + \sigma^2 + \sigma^2$

$= 3\sigma^2$

$V(t_2) = V[2X_1 + 3X_3 - 4X_2]$

$= 4V(X_1) + 9V(X_3) + 16V(X_2)$

$= 4\sigma^2 + 9\sigma^2 + 16\sigma^2$

$= 29\sigma^2$

$V(t_3) = V\left(\frac{1}{3}(\lambda X_1 + X_2 + X_3)\right)$

$= \frac{1}{9}V[X_1 + X_2 + X_3]$

$= \frac{1}{9}\left[\sigma^2 + \sigma^2 + \sigma^2\right]$

$= \frac{\sigma^2}{3}$

since $V(t_3)$ is the least of all T_1, T_2, T_3

t_3 is the most efficient estimator of μ.

1.4 SUFFICIENCY

Let, the random sample $X_1, X_2, \ldots, X_n$ have the joint distribution function F_θ which is known expect for k parameters $\theta_1, \theta_2, \ldots, \theta_k$. We shall write $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$, a vector with k components, and shall suppose that the parameter space is Θ. Consider k functionally unrelated statistics $T_1, T_2, \ldots, T_k$, the whole set of which may be denoted that by T.

Definition 8:

Let $X_1, X_2, \ldots, X_n$ be a random sample from the cumulative density function $F_\theta(.) = (F_\theta(.) : \theta \in \Theta)$ where θ is unknown and it is a known family of distribution. A statistic for θ if its conditional distribution of $X_1, X_2, \ldots, X_n$ for any given set of values of $T_1, T_2, \ldots, T_k$ is independent of θ.

Theorem 1: NEYMAN-FISHER FACTORIZATION

Statement:

Let X be a discrete random variable with p.m.f $f(x,\theta)$, $\theta\,\varepsilon\,\Theta$. Then T(x) is sufficient iff $f(x,\theta) = g(T(x),\theta)h(x) \quad \forall\ \theta\ \varepsilon\ \Theta$

Proof:

Let $\quad f(x,\theta) = g(T(x),\theta)h(x) \quad \forall\ \theta\ \varepsilon\ \Theta;\ \ x\,\varepsilon\,\Re$

$$= \sum_{x:T(x)=t} g(T(x),\theta)h(x)$$

$$= g(T(x),\theta) \sum_{x:T(x)=t} h(x)$$

Let , $P_\theta(X = x'/T(x) = t) = \begin{cases} 0 & if\ T(x') \neq t \\ \dfrac{P_\theta[X = x',\ T(x) = T(x')]}{P[T(x) = T(x')]} & if\ T(x') = t \end{cases}$

Consider ,

$$\frac{P_\theta[X = x',\ T(x) = T(x')]}{P[T(x) = T(x')]} = \frac{P_\theta(X = x')}{g(T(x),\theta) \sum_{x:T(x)=t} h(x)}$$

$$= \frac{g(T(x),\theta),\ h(x')}{g(T(x),\theta) \sum_{x:T(x)=t} h(x)}$$

$$= \frac{h(x')}{\sum_{x:T(x)=t} h(x)}$$

$P_\theta(X = x'/T(x) = t)$ = $\dfrac{h(x')}{\sum_{x:T(x)=t} h(x)}$ is independent of θ if T(x')=t.

So the conditional distribution of X given T is independent of the parameter. So T is sufficient statistic for θ.

Conversely, Let T is sufficient for θ.

$\Rightarrow P_\theta(X = x'/T(x) = t)$ = C (x', t) $\quad\quad \because$ *independent of* θ

$\Rightarrow \dfrac{P_\theta[X = x',\ T(x) = T(x')]}{P[T(x) = T(x')]}$ = C (x', t) $\quad\quad \because T(x') = t$

$\Rightarrow P_\theta(X = x') =$ C (x', t) $P[T(x) = T(x')]$

$\quad\quad =$ C (x', t) $g(T(x), \theta)$

$\therefore P_\theta(X = x') = g(T(x), \theta)\, h(x)$

Hence proved.

Example 8:

1. Suppose $X_1, X_2, \ldots, X_n$ are Independent Identically Distributed (IID) random variables with common probability density function (p.d.f)

$$f(x) = \begin{cases} \theta^x (1-\theta)^{1-x} & if \;\; x = 0,1 \\ 0 & otherwise \end{cases} \quad \text{where } 0 < \theta < 1$$

Solution:

The p.d.f of Bernoulli distribution is

$P(X = x) = \theta^x (1-\theta)^{1-x}$

The joint probability density function of $x_1, x_2, \ldots, x_n$ is

$$L(\theta) = \prod_{i=1}^{n} P(X_i = x) = \theta^{\sum x_i} (1-\theta)^{n - \sum x_i}$$

$$L(\theta)=\left(\frac{\theta}{1-\theta}\right)^{\sum x_i}(1-\theta)^n$$

$$L(\theta)=g\left(\sum x_i,\theta\right)h(.)$$

Here, $g(t)=\left(\frac{\theta}{1-\theta}\right)^{T}(1-\theta)^n; T=\sum_{i=1}^{n} x_i; h(.)=(1-\theta)^n$

Therefore, $\sum x_i$ is a sufficient statistic for θ.

Example 9: Let X ~ Poisson (θ). Find the sufficient statistic for θ.

Solution:

The p.d.f of Poisson distribution is

$$P(X=x)=\begin{cases}\frac{e^{-\theta}\theta^{x}}{x!}, & x=0,1,2\ldots,\theta>0\\ 0, & otherwise\end{cases}$$

$$L(\theta)=\prod_{i=1}^{n} P(X_i=x)=\frac{e^{-n\theta}\theta^{\sum x_i}}{\prod_{i=1}^{n} x_i!}$$

$$L(\theta)=e^{-n\theta}\theta^{\sum x_i}\left(\frac{1}{\prod_{i=1}^{n} x_i!}\right)$$

Here, $g(t)=e^{-n\theta}\theta^{\sum x_i}; T=\sum_{i=1}^{n} x_i; h(.)=\left(\frac{1}{\prod_{i=1}^{n} x_i!}\right)$

Therefore, $\sum_{i=1}^{n} x_i$ is a sufficient statistic for θ.

Example 10: Let X ~ Exponential (θ). Find the sufficient statistic for θ.

Solution:

The p.d.f of exponential distribution is

$$f(x|\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$$

$$L(\theta|x)=\frac{1}{\theta^n}e^{-\frac{\sum x_i}{\theta}}$$

$$L(\theta|x)=g\left(\sum_{i=1}^{n}x_i\right)h(.)$$

Here, $g(t)=\frac{1}{\theta^n}e^{-\frac{\sum_{i=1}^{n}x_i}{\theta}};T=\sum_{i=1}^{n}x_i;h(.)=1$

Therefore, $\sum_{i=1}^{n}x_i$ is a sufficient statistic for θ.

Example 11: Let X ~ Normal$(0,\sigma^2)$. Find the sufficient statistic for σ^2.

$$f(x|\mu,\sigma^2)=\begin{cases}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2};-\infty<x,\mu<\infty,\sigma>0\\ 0, otherwise\end{cases}$$

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

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

$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}-\log\sigma}$$

$$L(\sigma^2|x)=\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum x^2}{2\sigma^2}-\log\sigma}=g\left(\sum x_i^2\right)h(.)$$

Here, $g(t)=e^{-\frac{\sum_{i=1}^{n}x_i^2}{2\sigma^2}}; T=\sum_{i=1}^{n}x_i^2; h(.)=\left(\frac{1}{\sqrt{2\pi}}\right)^n$

Therefore, $\sum_{i=1}^{n}x_i^2$ is a sufficient statistic for σ^2.

Example 12: Let X ~ Normal$(\mu,1)$. Find the sufficient statistic for μ.

Solution:

$$f(x|\mu,1)=\begin{cases}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2}; -\infty<x,\mu<\infty\\ 0, otherwise\end{cases}$$

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

$$=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}+x\mu-\frac{\mu^2}{2}}$$

$$L(\mu|x)=\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{\mu\sum x_i}\prod_{i=1}^{n}e^{-\frac{1}{2}\left(x_i^2-\mu^2\right)}$$

$$=g\left(\sum x_i,\sum x_i^2\right)\left(\frac{1}{\sqrt{2\pi}}\right)^n$$

Here, $g(t)=e^{\mu\sum x_i}\prod_{i=1}^{n}e^{-\frac{1}{2}\left(x_i^2-\mu^2\right)}; T_1=\sum_{i=1}^{n}x_i; T_2=\sum_{i=1}^{n}x_i^2; h(.)=\left(\frac{1}{\sqrt{2\pi}}\right)^n$

Therefore, $\sum_{i=1}^{n}x_i, \sum_{i=1}^{n}x_i^2$ is a sufficient statistic for μ.

Example 13: Consider X~f(x: θ), X=1,2,3; $\theta=\theta_1,\theta_2,\theta_3$ with the probability function

x	θ_1	θ_2	θ_3
1	0.1	0.2	0.3
2	0.7	0.4	0.1
3	0.2	0.4	0.6

Show that the statistic $T = \begin{cases} 0 & if \ \ x \, is \, odd \\ 1 & if \ \ x \, is \, even \end{cases}$ is sufficient for θ.

Solution:

The distribution of T is given by

T	θ_1	θ_2	θ_3
0	0.3	0.6	0.9
1	0.7	0.4	0.1

The conditional probability function of x|t is given by

$$P(X|T) = \frac{P(X = x, T = t)}{P(T = t)} = \frac{P(X \cap T)}{P(T)}$$

x	θ_1	θ_2	θ_3
1	1/3	1/3	1/3
2	0	0	0
3	2/3	2/3	2/3

The conditional probability function of x|t when t=1 is given by

x	θ_1	θ_2	θ_3
1	0	0	0
2	1	1	1
3	0	0	0

Since the distribution function of x|t, f(x|t) does not depend on the parameter θ. T is sufficient for θ.

Definition 8: MINIMUM VARIANCE UNBIASED ESTIMATOR

Let $U_{\gamma(\theta)}$ be the class of all unbiased estimator of the parametric function $\gamma(\theta)$. If a statistic $T = T(x_1, .. x_n)$ based on sample size n is such that,

(i) T is unbiased for $\gamma(\theta)\ \forall\ \theta\ \varepsilon\ \Theta$ (i.e) $E(T) = \gamma(\theta)\quad \forall\ \theta\ \varepsilon\ \Theta$

(ii) It has the smallest variance among the class of all unbiased estimators of $\gamma(\theta)$ then T is called minimum variance unbiased estimator of $\gamma(\theta)$. (i.e) $Var(T) \leq Var(T^*)$; $T, T^*\ \varepsilon\ U_{\gamma(\theta)}$ *and* $\theta\ \varepsilon\ \Theta$. where T^* is any other unbiased estimator of $\gamma(\theta)$.

Theorem 2:

The minimum variance unbiased estimator is unique in the sense that if T_1 *and* T_2 are minimum variance unbiased estimators for $\gamma(\theta)$ then $T_1 = T_2$ *almost surely*.

Proof:

To prove $T_1 = T_2$

Given T_1 *and* T_2 are unbiased estimator for $\gamma(\theta)$

(i.e) $E(T_1) = E(T_2) = \gamma(\theta)\qquad \forall\ \theta\ \varepsilon\ \Theta$

$Var(T_1) = Var(T_2)\qquad \forall\ \theta\ \varepsilon\ \Theta$

Consider a new estimator $T = \frac{1}{2}(T_1 + T_2)$ which is also unbiased

Since $E(T) = \frac{1}{2}[E(T_1) + E(T_2)]$

$= \frac{1}{2}[\gamma(\theta) + \gamma(\theta)]$

$= \gamma(\theta)$

$$Var(T) = Var\left[\frac{1}{2}(T_1 + T_2)\right]$$

$$= \frac{1}{4}Var(T_1 + T_2)$$

$$= \frac{1}{4}[Var(T_1) + Var(T_2) + 2Cov(T_1, T_2)]$$

$$= \frac{1}{4}\left[Var(T_1) + Var(T_2) + 2\rho\sqrt{Var(T_1)Var(T_2)}\right]$$

$$= \quad \frac{1}{4}\left[Var(T_1)+Var(T_1)+2\rho\sqrt{Var(T_1)Var(T_1)}\right]$$

$$= \quad \frac{1}{4}\left[2Var(T_1)+2\rho\sqrt{Var^2(T_1)}\right]$$

$$= \quad \frac{1}{4}\left[2Var(T_1)+2\rho Var(T_1)\right]$$

$$= \quad \frac{1}{2}\left[Var(T_1)+\rho Var(T_1)\right]$$

$$= \quad \frac{1}{2}Var(T_1)[1+\rho]$$

where ρ is Karl Pearson's coefficient of correlation between T_1 *and* T_2

Since T_1 is MVUE, $Var(T)\geq Var(T_1)$

$$\Rightarrow \frac{1}{2}Var(T_1)[1+\rho]\geq Var(T_1)$$

$$\Rightarrow \frac{1}{2}[1+\rho]\geq 1$$

$$\Rightarrow \rho\geq 1$$

since $|\rho|\leq 1$ we must have $\rho=1$

(i.e) T_1 *and* T_2 must have relation of the form:

$$T_1=\alpha+\beta T_2 \quad \rightarrow 1$$

where α *and* β are constants independent of $x_1,x_2,\ldots,x_n$ but may depend on θ.

(i.e) we have $\alpha=\alpha(\theta)$ *and* $\beta=\beta(\theta)$

Taking expectation on both sides in equation 1

$$\Rightarrow E(T_1)=E(\alpha)+E(\beta T_2)$$

$$\Rightarrow \theta=\alpha+\beta\theta \quad \rightarrow 2$$

$$Var(T_1)=Var(\alpha+\beta T_2)$$

$$Var(T_1)=\beta^2\,Var(T_1)$$

$$\Rightarrow \beta^2=1 \qquad \Rightarrow \beta=\pm 1$$

But since $\rho(T_1,T_2)=\pm 1$ the coefficient of regression of T_1 *on* T_2 must be positive, therefore $\beta=1$

Sub $\beta=1$ in equation 2

$\Rightarrow \alpha = 0$

Sub α *and* β in equation 1

$\Rightarrow T_1 = T_2$

Hence proved.

Theorem 3:

Let T_1, T_2 be unbiased estimators of $\gamma(\theta)$ with efficiencies ρ_1 *and* ρ_2 respectively and $\rho = \rho_\theta$ be the correlation coefficient between them then

$$\sqrt{\rho_1\rho_2} - \sqrt{(1-\rho_1)(1-\rho_2)} \le \rho \le \sqrt{\rho_1\rho_2} + \sqrt{(1-\rho_1)(1-\rho_2)}$$

Proof:

Let T be minimum variance unbiased estimator of $\gamma(\theta)$. Then $E_\theta(T_1) = E_\theta(T_2) = \gamma(\theta) \;\; \forall \; \theta \varepsilon \Theta$ and

$$\rho_1 = \frac{V_\theta(T)}{V_\theta(T_1)} \qquad \Rightarrow V_\theta(T_1) = \frac{V_\theta(T)}{\rho_1}$$

$$\rho_2 = \frac{V_\theta(T)}{V_\theta(T_2)} \qquad \Rightarrow V_\theta(T_2) = \frac{V_\theta(T)}{\rho_2}$$

Let us consider another estimator $T_3 = \lambda T_1 + \mu T_2$ which is also unbiased estimator of $\gamma(\theta)$.

(i.e) $E(T_3) = E(\lambda T_1 + \mu T_2)$

$= \lambda E(T_1) + \mu E(T_2)$

$= (\lambda + \mu)\gamma(\theta)$

$\Rightarrow \lambda + \mu = 1$

$$V_\theta(T_3) = V(\lambda T_1 + \mu T_2)$$

$$= \lambda^2 V(T_1) + \mu^2 V(T_2) + 2\lambda\mu\, Cov(T_1, T_2)$$

$$= \lambda^2 V(T_1) + \mu^2 V(T_2) + 2\rho\lambda\mu \sqrt{Var(T_1)Var(T_2)}$$

$$= \lambda^2 \frac{V(T)}{\rho_1} + \mu^2 \frac{V(T)}{\rho_2} + 2\rho\lambda\mu \sqrt{\frac{V(T)}{\rho_1}\frac{V(T)}{\rho_2}}$$

$$= \lambda^2 \frac{V(T)}{\rho_1} + \mu^2 \frac{V(T)}{\rho_2} + 2\rho\lambda\mu\, V(T)\frac{1}{\sqrt{\rho_1\rho_2}}$$

$$= \quad V(T)\left[\frac{\lambda^2}{\rho_1}+\frac{\mu^2}{\rho_2}+\frac{2\,\rho\lambda\mu}{\sqrt{\rho_1\rho_2}}\right]$$

But $V_\theta(T_3) \geq V_\theta(T)$, since $V_\theta(T)$ has minimum variance

$$\Rightarrow V_\theta(T)\left[\frac{\lambda^2}{\rho_1}+\frac{\mu^2}{\rho_2}+\frac{2\,\rho\lambda\mu}{\sqrt{\rho_1\rho_2}}\right] \geq V_\theta(T)$$

$$\frac{\lambda^2}{\rho_1}+\frac{\mu^2}{\rho_2}+\frac{2\rho\lambda\mu}{\sqrt{\rho_1\rho_2}} \geq 1 = (\lambda+\mu)^2$$

$$\frac{\lambda^2}{\rho_1}+\frac{\mu^2}{\rho_2}+\frac{2\rho\lambda\mu}{\sqrt{\rho_1\rho_2}} \geq \lambda^2+\mu^2+2\lambda\mu$$

$$\left(\frac{\lambda^2}{\rho_1}-\lambda^2\right)+\left(\frac{\mu^2}{\rho_2}-\mu^2\right)+\left(\frac{2\rho\lambda\mu}{\sqrt{\rho_1\rho_2}}-2\lambda\mu\right) \geq 0$$

$$\left(\frac{1}{\rho_1}-1\right)\lambda^2+\left(\frac{1}{\rho_2}-1\right)\mu^2+2\left(\frac{\rho}{\sqrt{\rho_1\rho_2}}-1\right)\mu\lambda \;\geq 0$$

$$\left(\frac{1}{\rho_1}-1\right)\frac{\lambda^2}{\mu^2}+\left(\frac{1}{\rho_2}-1\right)\frac{\mu^2}{\mu^2}+2\left(\frac{\rho}{\sqrt{\rho_1\rho_2}}-1\right)\frac{\mu\lambda}{\mu^2} \;\geq 0$$

which is quadratic equation in $\left(\frac{\lambda}{\mu}\right)$

Note that $\rho_i < 1 \Rightarrow \frac{1}{\rho_i} > 1 \quad (or)\ \left(\frac{1}{\rho_i}-1\right) > 0 \ \ \forall\ i = 1,2,\ldots$

We know that ,

$$AX^2+BX+C \geq 0\ , \quad A>0\ ,\ C>0 \ \text{ iff discriminant is } B^2-4AC \leq 0$$

$$4\left(\frac{\rho}{\sqrt{\rho_1\rho_2}}-1\right)^2 - 4\left(\frac{1}{\rho_1}-1\right)\left(\frac{1}{\rho_2}-1\right) \leq 0$$

$$\Rightarrow \left(\rho-\sqrt{\rho_1\rho_2}\right)^2 - (1-\rho_1)(1-\rho_2) \leq 0$$

$$\Rightarrow \rho^2 - 2\rho\sqrt{\rho_1\rho_2} + \rho_1\rho_2 - 1 + \rho_2 + \rho_1 - \rho_1\rho_2 \leq 0$$

$$\Rightarrow \rho^2 - 2\rho\sqrt{\rho_1\rho_2} + (\rho_1+\rho_2-1) \leq 0$$

$$\rho = \frac{2\sqrt{\rho_1\rho_2} \pm \sqrt{4\rho_1\rho_2 - 4(\rho_1+\rho_2-1)}}{2}$$

$$= \frac{2\left[\sqrt{\rho_1\rho_2} \pm \sqrt{\rho_1\rho_2 - (\rho_1 + \rho_2 - 1)}\right]}{2}$$

$$= \sqrt{\rho_1\rho_2} \pm \sqrt{(\rho_1 - 1)(\rho_2 - 1)}$$

$$\sqrt{\rho_1\rho_2} - \sqrt{(\rho_1 - 1)(\rho_2 - 1)} \le \rho \le \sqrt{\rho_1\rho_2} + \sqrt{(\rho_1 - 1)(\rho_2 - 1)}$$

$$\Rightarrow \sqrt{\rho_1\rho_2} - \sqrt{(1-\rho_1)(1-\rho_2)} \le \rho \le \sqrt{\rho_1\rho_2} + \sqrt{(1-\rho_1)(1-\rho_2)}$$

Theorem 4:

If T_1 is a minimum variance unbiased estimator for $\gamma(\theta)$ $\forall\, \theta\, \varepsilon \Theta$ and T_2 is any other unbiased estimator of $\gamma(\theta)$ with efficiency $\rho = \rho_\theta$ then the correlation coefficient between T_1 *and* T_2 is given by $\rho = \sqrt{\rho}$ (i.e) $\rho_\theta = \sqrt{\rho_\theta}$ $\forall\, \theta\, \varepsilon\, \Theta$

Proof:

Using the previous ***Theorem:3*** statement the correlation coefficient ρ lies between

$$\sqrt{\rho_1\rho_2} - \sqrt{(1-\rho_1)(1-\rho_2)} \le \rho \le \sqrt{\rho_1\rho_2} + \sqrt{(1-\rho_1)(1-\rho_2)}$$

Here T_1 is a minimum variance unbiased estimator of $\gamma(\theta)$ then the efficiency $\rho_1 = 1$ and T_2 is any other unbiased estimator of $\gamma(\theta)$ with efficiency ρ

$\rho_1 = 1$ *and* $\rho_2 = \rho$ sub in 1

$$\sqrt{1.\rho} \le \rho \le \sqrt{\rho}$$

$$\therefore \rho = \sqrt{\rho}$$

Theorem 5:

If T_1 is a minimum variance unbiased estimator for $\gamma(\theta)$ $\forall\, \theta\, \varepsilon \Theta$ and T_2 is any other unbiased estimator of $\gamma(\theta)$ with efficiency $\rho < 1$, then no unbiased linear combination of T_1 *and* T_2 can be an MVUE of $\gamma(\theta)$

Proof:

Consider a linear combination:

$$T = l_1T_1 + l_2T_2$$

will be an unbiased estimator of $\gamma(\theta)$ if

$$E(T)=E(l_1T_1+l_2T_2)=l_1\,E(T_1)+l_2\,E(T_2)=\gamma(\theta) \qquad \forall\ \theta\ \varepsilon\,\Theta$$

$$\Rightarrow l_1+l_2=1 \qquad \text{since} \quad E(T_1)=E(T_2)=\gamma(\theta)$$

The efficiency, $\rho=\dfrac{Var_\theta(T_1)}{Var_\theta(T_2)} \qquad \Rightarrow Var_\theta(T_2)=\dfrac{Var(T_1)}{\rho}$

And $\rho=\rho(T_1,T_2)=\sqrt{\rho}$

$$Var_\theta T=Var_\theta[l_1T_1+l_2T_2]$$

$$= l_1^2\,Var_\theta(T_1)+l_2^2\,Var_\theta(T_2)+2l_1l_2\,Cov(T_1,T_2)$$

$$= l_1^2\,Var_\theta(T_1)+l_2^2\,Var_\theta(T_2)+2l_1l_2\,\rho\sqrt{Var(T_1)Var(T_2)}$$

$$= l_1^2\,Var_\theta(T_1)+l_2^2\,\frac{Var_\theta(T_1)}{\rho}+2l_1l_2\,\rho\sqrt{Var(T_1)\frac{Var(T_1)}{\rho}}$$

$$= Var_\theta(T_1)\left[l_1^2+\frac{l_2^2}{\rho}+2l_1l_2\,\frac{\rho}{\sqrt{\rho}}\right]$$

$$= Var_\theta(T_1)\left[l_1^2+\frac{l_2^2}{\rho}+2l_1l_2\right] \qquad \because \rho=\sqrt{\rho}$$

$$Var_\theta(T_1)\left[l_1^2+l_2^2+2l_1l_2\right]<Var_\theta(T) \qquad ;0<\rho<1,\ \frac{1}{\rho}>1$$

$$Var_\theta(T)>Var_\theta(T_1)(l_1+l_2)^2$$

$$Var_\theta(T)>Var_\theta(T_1)$$

$\therefore$ T cannot be MVU estimator.

Information Function (Or) Regularity Conditions

(i) Θ is a non degenerate open interval on the head line $\Re$.

(ii) The support of the random variable is independent of the parameter θ.

(iii) $\dfrac{\partial_i\ f(x/\theta)}{\partial\theta_i}$ exists for all i=1,2,3

(iv) $\dfrac{\partial^i}{\partial\theta_i}\displaystyle\int_x f(x/\theta)dx=\int_x\frac{\partial^i\ f(x/\theta)}{\partial\theta_i}dx$ holds for i=1,2,…n

(v) For some function $T(x)$

$$\frac{\partial^i}{\partial\theta_i}\int_x T(x) f(x/\theta)dx = \int_x T(x)\frac{\partial^i \; f(x/\theta)}{\partial\theta_i}dx$$ holds for i=1,2,…n

(vi) $E_\theta\left[\frac{\partial \log \; f_\theta(x_1, x_2,...,x_n)}{\partial_\theta}\right]$ exists and is positive.

It is also called Fisher information measure.

Theorem 6: CRAMER-RAO INEQUALITY

Under the regularity condition if T is an unbiased estimator for $\gamma(\theta)$ which is assumed to be a differentiable function of θ satisfies the inequality

$$Var_\theta(T) \geq \frac{[\gamma'(\theta)]^2}{E_\theta\left[\frac{\partial \log \; f_\theta(x_1, x_2,...,x_n)}{\partial\theta}\right]^2} \qquad \text{(or)}$$

$$Var_\theta(T) \geq \frac{[\gamma'(\theta)]^2}{I(\theta)}$$

where $I(\theta)$ is information measure.

Proof:

Let X be a random variable from the pdf $f(x/\theta)$ and let L be the likelihood function of the random sample $(x_1, x_2,...,x_n)$ from this population. Thus

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

since L is the joint pdf of $(x_1, x_2,...,x_n)$ then

$$\int_x L(\theta/x)dx = 1 \qquad \text{---(1)}$$

where x represents the domain of $(x_1, x_2,...,x_n)$ and the integral is an n-fold integral.

Differentiating w.r.t. θ and using regularity conditions, we get

$$\int_x \frac{\partial}{\partial\theta} L\,dx = 0$$

$$\Rightarrow \int_x \left(\frac{\partial}{\partial \theta} \log L \right) L dx = 0 \qquad \text{---(2)}$$

$$\Rightarrow \int_x \left(\frac{1}{L} \right) \frac{\partial L}{\partial \theta} L dx = 0$$

$$\Rightarrow E\left(\frac{\partial}{\partial \theta} \log L \right) = 0 \qquad \forall\ \theta\ \varepsilon\ \Theta$$

Let us consider T(x) = T$(x_1, x_2, \ldots, x_n)$ be an unbiased estimator of $\gamma(\theta)$ such that, E(T)= $\gamma(\theta)$.

$$\therefore \int_x T(x) L dx = \gamma(\theta) \qquad \text{---(3)}$$

Differentiating w.r.t. θ we get

$$\int_x T(x) \frac{\partial L}{\partial \theta} dx = \gamma'(\theta)$$

$$\Rightarrow \int_x T(x) \left(\frac{\partial \log L}{\partial \theta} \right) L dx = \gamma'(\theta) \qquad \text{---(4)}$$

Multiplying $\gamma(\theta)$ in Equation 2, we get

$$\int_x \gamma(\theta) \left(\frac{\partial}{\partial \theta} \log L \right) L dx = 0 \qquad \text{---(5)}$$

Subtracting Equation 4 and 5, we get

$$\Rightarrow \int_x [T(x) - \gamma(\theta)] \left(\frac{\partial \log L}{\partial \theta} \right) L dx = \gamma'(\theta) \qquad \text{---(6)}$$

$$\Rightarrow E\left[T(x) . \left(\frac{\partial \log L}{\partial \theta} \right) \right] = \gamma'(\theta) \qquad \text{---(7)}$$

$$Cov\left[T(x) . \left(\frac{\partial \log L}{\partial \theta} \right) \right] = E\left[T(x) . \left(\frac{\partial \log L}{\partial \theta} \right) \right] - E(T(x)) E\left(\frac{\partial \log L}{\partial \theta} \right)$$

$$= \gamma'(\theta)$$

We know that,

$$\{Cov(x,y)\}^2 \le Var(X)Var(Y)$$

$$\Rightarrow [\gamma'(\theta)]^2 \le Var(T).Var\left(\frac{\partial}{\partial\theta}\log L\right)$$

$$\Rightarrow [\gamma'(\theta)]^2 \le Var(T)\left\{E\left(\frac{\partial}{\partial\theta}\log L\right)^2 - \left(E\frac{\partial}{\partial\theta}\log L\right)^2\right\}$$

$$\Rightarrow [\gamma'(\theta)]^2 \le Var(T)\left\{E\left(\frac{\partial}{\partial\theta}\log L\right)^2\right\}$$

$$\Rightarrow \frac{[\gamma'(\theta)]^2}{E\left(\frac{\partial}{\partial\theta}\log L\right)^2} \le Var(T)$$

$$\Rightarrow Var_\theta(T) \ge \frac{[\gamma'(\theta)]^2}{E_\theta\left[\frac{\partial \log f(x_1,x_2,\ldots,x_n)}{\partial\theta}\right]^2} \qquad \text{(or)}$$

$$Var_\theta(T) \ge \frac{[\gamma'(\theta)]^2}{I(\theta)}$$

Hence Proved.

<u>Remarks</u>:

- An unbiased estimator T of $\gamma(\theta)$ for which Cramer-Rao lower bound is attained then it is called minimum variance bound estimator.
- The fisher information measure $I(\theta) = E\left(\frac{\partial}{\partial\theta}\log L\right)^2 = -E\left[\frac{\partial^2}{\partial\theta^2}\log L\right]$

Conditions for the equality sign in Cramer-Rao Inequality

In Cramer-Rao inequality

$$Var_\theta(T) \ge \frac{[\gamma'(\theta)]^2}{E\left(\frac{\partial}{\partial\theta}\log L\right)^2} \qquad \text{---(1)}$$

Rewriting Equation 1, we get

$$Var_\theta(T)E\left(\frac{\partial}{\partial\theta}\log L\right)^2 \geq [\gamma'(\theta)]^2$$

$$\Rightarrow E[T-\gamma(\theta)]^2 \;\; E\left(\frac{\partial}{\partial\theta}\log L\right)^2 \geq \; [\gamma'(\theta)]^2 \qquad \text{---(2)}$$

The sign of equality will hold in CRR inequality if and only if the sign of equality holds in Equation 2. The sign of equality will hold in Equation 2 by Cauchy-Schwartz inequality $Cov(X,Y) = E(X^2)\; E(Y^2)$ iff the variables $(T-\gamma(\theta))$and $\frac{\partial}{\partial\theta}\log L$ are proportional to each other.

$$\therefore \frac{T-\gamma(\theta)}{\frac{\partial}{\partial\theta}\log L} = \lambda\,(say) \; = \lambda(\theta)$$

where λ is constant independent of $(x_1, x_2, \ldots, x_n)$ but depend on θ

$$\frac{\partial}{\partial\theta}\log L = \frac{T-\gamma(\theta)}{\lambda(\theta)}$$

$$\Rightarrow T-\gamma(\theta)[A(\theta)] = \frac{\partial}{\partial\theta}\log L \qquad \text{---(3)}$$

where $A = A(\theta) = \frac{1}{\lambda(\theta)}$

Hence the necessary and sufficient condition for an unbiased estimator T to attain the lower bound of the variance is given by Equation 3

Further the C-R minimum variance bound is given by

$$Var\,(T) \; = \frac{[\gamma'(\theta)]^2}{E\left(\frac{\partial}{\partial\theta}\log L\right)^2} \qquad \text{---(4)}$$

But ,

$$E\left(\frac{\partial}{\partial\theta}\log L\right)^2 \; = \; E[A(\theta).\{T-\gamma(\theta)\}]^2$$

$$= \quad \{A(\theta)\}^2 .\, E\{T-\gamma(\theta)\}^2$$

$$= \quad \{A(\theta)\}^2 .\, Var(T)$$

Substituting in Equation 4,

$$Var(T) = \frac{[\gamma'(\theta)]^2}{\{A(\theta)\}^2 Var(T)}$$

$$\Rightarrow Var(T) = \left|\frac{\gamma'(\theta)}{A(\theta)}\right| = |\gamma'(\theta).\ \lambda(\theta)|$$

Hence , if the likelihood function L is expressible in the form Equation 3 then

1. T is unbiased estimator of $\gamma(\theta)$
2. Minimum variance bound estimator (T) for $\gamma(\theta)$ exists and
3. $Var(T) = \left|\frac{\gamma'(\theta)}{A(\theta)}\right| = |\gamma'(\theta).\ \lambda(\theta)|$

Example 14: Obtain the MVB estimator for μ in normal population $N(\mu,\sigma^2)$ where σ^2 is known.

Solution:

If $x_1, x_2, ..., x_n$ is a random sample of size n from the normal population, then

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

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

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

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

$$= \quad \frac{1}{\sigma^2/n}\left(\sum_{i=1}^{n} xi/n - n\mu/n\right)$$

$$= \quad \frac{n}{\sigma^2}(\bar{x}-\mu)$$

which is of the form

$$\frac{\partial}{\partial\theta}\log L = T - \gamma(\theta)[A(\theta)]$$

then T is a MVB estimator for $\gamma(\theta)$ *and* $A(\theta)$ is a constant.

$\therefore \bar{x}$ is a MVB estimator for μ and

$$\Rightarrow Var(\hat{\mu}) = \left|\frac{\gamma'(\theta)}{A(\theta)}\right|$$

$$= \left|\frac{1}{n/\sigma^2}\right|$$

$$= \frac{\sigma^2}{n}$$

Example 15: A random sample $x_1, x_2, ..., x_n$ is taken from the normal population with mean 0 and variance σ^2. Examine if $\sum_{i=1}^{n} x_i^{\,2}/n$ is a MVB estimator for σ^2.

Solution:

If $x_1, x_2, ..., x_n$ is a random sample of size n from the normal population, then

$$L = \prod_{i=1}^{n} f(x_i, \sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\sum_{i=1}^{n} x_i^{\,2}/2\sigma^2\right\}; \qquad -\infty < x < \infty,\ \sigma > 0$$

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

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

$$\text{Log L} = -n/2\log\sigma^2 - n/2\log 2\pi - \sum_{i=1}^{n} x_i^{\,2}/2\sigma^2$$

$$\frac{\partial}{\partial\sigma^2}\log L = -n/2\sigma^2 - 0 - 1/2\sum_{i=1}^{n} x_i^{\,2}\left(\frac{-1}{\sigma^4}\right)$$

$$= \frac{\sum x_i^{\,2}/n - \sigma^2}{2\sigma^4/n} \text{ which is of the form } T - \gamma(\theta)[A(\theta)]$$

Hence $\hat{\sigma}^2 = \dfrac{\sum x_i^{\,2}}{n}$ is a MVB estimator and

$$Var(\hat{\sigma}^2) = \left|\frac{\gamma'(\theta)}{A(\theta)}\right|$$

$$= \left|\frac{1}{n/2\sigma^4}\right|$$

$$= \frac{2\sigma^4}{n}$$

Example 16: Find if MVB estimator exists for θ in Cauchy's population

$$\partial F(x;\theta) = \frac{1}{\pi}\ \frac{1}{1+(x-\theta)^2} \quad ; \quad -\infty < x < \infty$$

Solution:

Let $x_1, x_2, \ldots, x_n$ be a random sample from Cauchy's population.

$$f(x;\theta) = \frac{1}{\pi}\ \frac{1}{1+(x-\theta)^2}$$

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

$$\log L = -n\log\pi - \sum_{i=1}^{n}\log\left(1+(x_i-\theta)^2\right)$$

$$\frac{\partial}{\partial\theta}\log L = 0 + 2\sum_{i=1}^{n}\log\frac{(x_i-\theta)}{\left(1+(x_i-\theta)^2\right)}$$

Since $\frac{1}{\pi}\ \frac{1}{1+(x-\theta)^2}$ cannot be expressed in the form $T-\gamma(\theta)[A(\theta)]$ MVB estimator does not exist for θ in Cauchy's population and so Cramer-Rao lower bound is attainable by the variance of any
unbiased estimator θ.

Example 17: Show that $\overline{X} = \frac{\Sigma x_i}{n}$ in random sampling from

$$f(x;\theta) = \begin{cases} \frac{1}{\theta}\exp(-x/\theta) & ;\ 0 < x < \infty\ ,\ \theta > 0 \\ 0 & ;\ otherwise \end{cases}$$

is a MVB estimator and has variance $\frac{\sigma^2}{n}$.

Solution:

Let $x_1, x_2, \ldots, x_n$ be a random sample from the population

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}$$

$$L = \prod_{i=1}^{n} f(x;\theta) = \left(\frac{1}{\theta}\right)^n \exp\left[\sum_{i=1}^{n} - x_i / \theta\right]$$

$$\log L = -n \log\theta - \sum_{i=1}^{n} x_i / \theta$$

$$\frac{\partial}{\partial\theta} \log L = \frac{-n}{\theta} + \frac{\Sigma x_i}{\theta^2}$$

$$= \frac{-n\theta + \Sigma x_i}{\theta^2}$$

$$= \frac{n[\Sigma x_i / n \ - \theta]}{\theta^2}$$

$$= \frac{\bar{X} - \theta}{\theta^2 / n} \text{ which is of the form } \quad T - \gamma(\theta)[A(\theta)]$$

Hence $\bar{X}$ is the MVB estimator for θ and

$$Var(\hat{\theta}) = \left|\frac{\gamma'(\theta)}{A(\theta)}\right|$$

$$= \left|\frac{1}{1/\theta^2 / n}\right|$$

$$= \frac{\theta^2}{n}$$

Example 18: Let $x_1, x_2, \ldots, x_n$ be a random sample from a Bernoulli Distribution with parameter p. Then $\theta = p$ and $\Theta = \{\theta \quad 0 < \theta < 1\}$ Find the MVB estimator and its variance.

Solution:

Let $x_1, x_2, \ldots, x_n$ be a random sample from BD.

$$f_\theta(x) = \theta^x (1-\theta)^{1-x}$$

$$L = \prod_{i=1}^{n} f_\theta(x_i) = \theta^{\Sigma x_i} (1-\theta)^{n - \Sigma x_i}$$

$$\log L = \Sigma x_i \log\theta + n - \Sigma x_i \log(1-\theta)$$

$$\frac{\partial}{\partial\theta}\log L = \frac{\Sigma x_i}{\theta} + \left[\frac{(n-\Sigma x_i)}{(1-\theta)}\right]$$

$$= \frac{\Sigma x_i}{\theta} - \frac{(n-\Sigma x_i)}{1-\theta}$$

$$= \frac{(1-\theta)\Sigma x_i - \theta(n-\Sigma x_i)}{\theta(1-\theta)}$$

$$= \frac{\Sigma x_i - n\theta}{\theta(1-\theta)}$$

$$= \frac{n[\Sigma x_i / n \ - \ \theta]}{\theta(1-\theta)}$$

$$= \frac{\overline{X}-\theta}{\theta(1-\theta)/n} \text{ which is of the form } \quad T-\gamma(\theta)[A(\theta)]$$

Hence $\overline{X}$ is the MVB estimator for θ and

$$Var(\hat{\theta}) = \left|\frac{\gamma'(\theta)}{A(\theta)}\right|$$

$$= \left|\frac{1}{1/\theta(1-\theta)/n}\right|$$

$$= \frac{\theta(1-\theta)}{n}$$

Example 19: Let $x_1, x_2, \ldots, x_n$ be a random sample from the Poisson distribution with parameter θ. Find the MVB estimator and its variance.

Solution:

Let $x_1, x_2, \ldots, x_n$ be a random sample from the Poisson population

$$f_\theta(x) = \frac{e^{-\theta}\,\theta^x}{x!}$$

$$\text{L} = \prod_{i=1}^{n} f(x_i;\theta) = \frac{e^{-n\theta}\ \theta^{\Sigma x_i}}{\prod_{i=1}^{n} x!}$$

$$\log L = -n\theta + \Sigma x_i \log\theta - \sum_{i=1}^{n}\log x_i!$$

$$\frac{\partial}{\partial\theta}\log L = -n + \frac{\Sigma x_i}{\theta}$$

$$= \frac{-n\theta + \Sigma x_i}{\theta}$$

$$= \frac{n[\Sigma x_i / n - \theta]}{\theta}$$

$$= \frac{\bar{X} - \theta}{\theta / n} \text{ which is of the form } T - \gamma(\theta)[A(\theta)]$$

Hence $\bar{X} = \frac{\Sigma x_i}{n}$ is MVB estimator for θ and

$$Var(\hat{\theta}) = \left|\frac{\gamma'(\theta)}{A(\theta)}\right|$$

$$= \left|\frac{1}{1/\theta/n}\right|$$

$$= \frac{\theta}{n}$$

Example 20: Let $x_1, x_2, \ldots, x_n$ be a random sample from uniform distribution $U(0,\theta)$. Find the MVB estimator and variance.

Solution:

The support of uniform distribution $[U(0,\theta)]$, $\Theta = \{x : 0 < x < \theta\}$ depends on the parameter θ. This violates the regularity conditions and the C-R lower bound theorem does not produce the result.

Completeness:

We discussed one property, viz., sufficiency, that a statistic T may have in relation to a family of distributions. We shall now consider another property, to be called completeness.

Consider the statistic T based on the random variable $X_1, X_2, \ldots, X_n$ with joint distribution depending on $\theta \in \Theta$. The distribution of T itself will, in general, depend on θ. Hence, related to T, wehave again a family of distributions, say, $\{g(t,\theta), \theta \, \varepsilon \, \Theta\}$

Definition 10:

The statistic T=t(x) or more precisely the family of distributions $\{g(t,\theta), \theta \,\varepsilon\, \Theta\}$ is said to be complete for θ if

E[h(t)] =0 $\quad \forall\, \theta \;\Rightarrow P_\theta[h(t)=0]=1$

(i.e) $\int h(t)\, g(t,\theta) dt = 0 \quad \forall\, \theta\, \varepsilon \Theta$ (or)

$$\sum_t h(t)\, g(t,\theta) = 0 \quad \forall\, \theta\, \varepsilon \Theta$$

$\Rightarrow h(t) = 0 \quad \forall\, \theta\, \varepsilon \Theta$ almost surely(a.s)

Definition 11:

The statistic T, or the family of distributions $\{g(t,\theta), \theta \,\varepsilon\, \Theta\}$ is said to be boundedly complete for θ if, for any (measurable) function $\psi(T)$ is such that

$|\psi(T)| < M, \; for\, some\, M,$

$E_\theta[\psi(T)] = 0 \; for\, all\, \theta \in \Theta$

$\Rightarrow \psi(t) = 0 \; for\, all\, \theta \in \Theta$ *almost everywhere*

Note: If T is complete, then it is necessarily boundedly complete.

Theorem 7: RAO-BLACKWELLIZATION

Statement:

Let X and Y are two random variables such that $E(X) = \theta$, $\theta \,\varepsilon\, \Theta$. If a function $\phi(.)$ is defined as $\phi(y) = E(X \mid Y = y)$. Then

(i) $\quad E[\phi(y)] = \theta$ and

(ii) $\quad Var_\theta[\phi(y)] \le Var_\theta(x)$

Proof:

We will give only the proof for the case where the distribution of (x, y) is absolutely continuous.

Let $f_{XY}(x,y)$ denote the joint density function of X and Y.

$f_X(x)$ is the density function of X and

$f_Y(y)$ is the density function of Y.

To show that $E[\phi(y)] = \theta$

Consider, $\phi(y) = E(X \mid Y = y) = \int_{-\infty}^{\infty} x\, f_{X/Y}(x, y) dx$

Now ,

$$E[\phi(y)] = \int_{-\infty}^{\infty} \phi(y) g_Y(y) dy$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x\, f_{X/Y}(x, y) dx \right] g_Y(y) dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \frac{f(x, y)}{g_Y(y)}\; g_Y(y)\, dy\, dx$$

where $f_{X/Y}(x, y) = \dfrac{f(x, y)}{g_Y(y)}$

$$E[\phi(y)] = \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} f(x, y) dy \right] dx$$

$$E[\phi(y)] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx \quad = E(x) = \theta$$

Next to show that

$$Var_\theta[\phi(y)] \leq Var_\theta(x)$$

Consider, $Var_\theta(x) = E(x - \theta)^2$

$$= E(x - \phi(y) + \phi(y) - \theta)^2$$

$$= E(x - \phi(y))^2 + E(\phi(y) - \theta)^2 + 2E[\,(x - \phi(y))(\phi(y) - \theta)\,] \qquad \text{---(1)}$$

Consider,

$$E[\,(x - \phi(y))(\phi(y) - \theta)\,] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \phi(y))(\phi(y) - \theta)\, f_{X|Y}(x \mid y).\, g_Y(y) dx.dy$$

$$= \int_{-\infty}^{\infty} (\phi(y) - \theta) \left[\int_{-\infty}^{\infty} (x - \phi(y))\, f_{X|Y}(x \mid y) dx \right] g_Y(y) dy$$

$$= \int_{-\infty}^{\infty} (\phi(y) - \theta) \left[\int_{-\infty}^{\infty} x\, f_{X|Y}(x \mid y) dx - \int_{-\infty}^{\infty} \phi(y)\, f_{X|Y}(x \mid y) dx \right] g_Y(y) dy$$

$$= \int_{-\infty}^{\infty} (\phi(y)-\theta)\left[E_\theta(X \mid Y)-\phi(y)\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)dx \right] g_Y(y)dy$$

$$= \int_{-\infty}^{\infty} (\phi(y)-\theta)[E_\theta(X \mid Y)-\phi(y)]\, g_Y(y)dy$$

$$= \int_{-\infty}^{\infty} (\phi(y)-\theta)[\phi(y)-\phi(y)]\, g_Y(y)dy$$

$$= 0 \qquad \text{---(2)}$$

Substitute the Eqn. (2) in Eqn. (1), we get

$$Var_\theta(x) = E(x-\phi(y))^2 + V_\theta(\phi(y)) + 0$$

$$\Rightarrow Var_\theta(\phi(y)) = Var_\theta(x) - E(x-\phi(y))^2$$

$Var_\theta(\phi(y)) \le Var_\theta(x)$

Hence proved.

Theorem 8: LEHMANN-SCHEFFE

Statement:

If T(X) is a complete sufficient statistic and W(X) is an unbiased estimator of $\tau(\theta)$, then $\phi(T) = E(W/T)$ is an UMVUE of $\gamma(\theta)$. Furthermore $\phi(T)$ is the unique UMVUE in the sense that if T* is any other UMVUE, then $P(\phi(T) = T^*) = 1 \quad \forall\ \theta\ \varepsilon\ \Theta$.

Proof:

Let W be any unbiased estimator of $\tau(\theta)$

Then by Rao - blackwell theorem, $\phi(T) = E(W/T)$ is such that $Var_\theta(\phi(T)) \le Var_\theta(W) \quad \forall\ \theta$

Let W* be any other unbiased estimator and

$\phi^*(T) = E[W^*/T]$ then

$E_\theta[\phi(T) - \phi^*(T)] = 0 \quad \forall\ \theta$

And by the definition of completeness of T, it follows that,

$P_\theta[\phi(T) = \phi^*(T)] = 1 \quad \forall\ \theta$

Hence , $\phi(T)$ is the unique UMVUE.

Definition 12: CONSISTENCY

An estimator $T_n = T(x_1, x_2, \ldots, x_n)$ based on a random sample of size n is said to be consistent estimator of $\gamma(\theta)$ $\forall\, \theta \,\varepsilon\, \Theta$ if T_n converges to $\gamma(\theta)$ in probability (i.e) $T_n \xrightarrow{P} \gamma(\theta)$ as $n \to \infty$. In other words T_n is a consistent estimator of $\gamma(\theta)$ if for every $\varepsilon > 0, \eta > 0$ there exist a positive integer n which is $\geq m$ such that

$$P\left[\left|T_n - \gamma(\theta)\right| < \varepsilon\right] \to 1 \quad as\ n \to \infty \quad \forall\, \theta \,\varepsilon\, \Theta$$

$$\Rightarrow P\left[\left|T_n - \gamma(\theta)\right| < \varepsilon\right] > 1 - \eta \qquad \forall\, (n \geq m)$$

where m is some very large value of n.

Remarks:

If $x_1, x_2, \ldots, x_n$ is a random sample from population with finite mean $E(x_i) = \mu < \infty$, then by Khinchin's weak law of large number we have

$$\overline{X}_n = \frac{1}{n}\Sigma x_i \xrightarrow{P} E(X_i) = \mu \quad as\ n \to \infty$$

Hence sample mean $\left(\overline{X}_n\right)$ is always a consistent estimator of population mean (μ).

Theorem 9: INVARIANCE PROPERTY OF CONSISTENT ESTIMATOR

Statement:

If T_n is a consistent estimator of $\gamma(\theta)$ and $\psi(\gamma(\theta))$ is continuous function of $\gamma(\theta)$ then $\psi(T_n)$ is a consistent estimator of $\psi(\gamma(\theta))$.

Proof:

Since T_n is a consistent estimator of $\gamma(\theta)$

(i.e) $T_n \xrightarrow{P} \gamma(\theta)$ as $n \to \infty$

Also for every $\varepsilon > 0, \eta > 0$ there exist a positive integer $n \geq m$ such that

$$P[|T_n - \gamma(\theta)| < \varepsilon] > 1 - \eta \qquad \forall\, (n \geq m) \qquad \text{---(1)}$$

Since $\psi(.)$is a continuous for every $\varepsilon > 0$ however small, there exist a positive number ε_1 such that

$$|\psi(T_n) - \psi(\gamma(\theta))| < \varepsilon_1 \quad \text{whenever} \quad |T_n - \gamma(\theta)| < \varepsilon \quad \text{(i.e)}\ |T_n - \gamma(\theta)| < \varepsilon$$

$$\Rightarrow |\{\psi(T_n) - \psi(\gamma(\theta))\}| < \varepsilon_1 \qquad \text{---(2)}$$

For two events A and B if A$\Rightarrow$B then

$$A \subseteq B \ \Rightarrow P(A) \leq P(B) \qquad \text{(or)}$$

$$P(B) \geq P(B) \qquad \text{---(3)}$$

From Equation 2 and 3, we get

$$P\left[|\psi(T_n) - \psi(\gamma(\theta))| < \varepsilon_1\right] \geq P\left[|T_n - \gamma(\theta)| < \varepsilon\right]$$

$$P\left[|\psi(T_n) - \psi(\gamma(\theta))| < \varepsilon_1\right] \geq 1 - \eta \quad \forall\, (n \geq m)$$

$$\Rightarrow \psi(T_n) \xrightarrow{P} \psi(\gamma(\theta)) \quad as \ \ n \to \infty$$

$\therefore \psi(T_n)$ is a consistent estimator of $\psi(\gamma(\theta))$

Theorem 10: SUFFICIENT CONDITION FOR CONSISTENCY

Statement:

Let $\{T_n\}$ be a sequence of estimator such that for all $\theta \ \varepsilon\ \Theta$

1. $E_\theta(T_n) \to \gamma(\theta) \ \ as \ \ n \to \infty$
2. $Var_\theta(T_n) \to 0 \ \ as \ \ n \to \infty$

Then T_n is a consistent estimator of $\gamma(\theta)$.

Proof:

To prove that , T_n is a consistent estimator of $\gamma(\theta)$

(i.e) $T_n \xrightarrow{P} \gamma(\theta)$ as $n \to \infty$

(i.e) $P[|T_n - \gamma(\theta)| < \varepsilon] > 1 - \eta \qquad \forall\,(n \geq m)$

where ε *and* η are arbitrarily small positive numbers and m is some large value of n.

Applying Chebyshev's inequality to the statistic T_n we get,

$$P\left[\,|T_n - E_\theta(T_n)| \leq \delta\,\right] \geq 1 - \frac{Var(T_n)}{\delta^2}$$

We have ,

$$|T_n - \gamma(\theta)| = \left[\,|T_n - E_\theta(T_n) + E_\theta(T_n) - \gamma(\theta)|\,\right]$$

$$\leq |T_n - E_\theta(T_n)| + |E_\theta(T_n) - \gamma(\theta)|$$

Now,

$$|T_n - E_\theta(T_n)| \leq \delta \quad \Rightarrow |T_n - \gamma(\theta)| \leq \delta + |E(T_n) - \gamma(\theta)|$$

Since for two events A and B if $A \Rightarrow B$ then

$$A \subseteq B \quad \Rightarrow P(A) \leq P(B) \quad or \quad P(B) \geq P(A)$$

$$P\left\{|T_n - \gamma(\theta)| \leq \delta + |E_\theta(T_n) - \gamma(\theta)|\right\} \geq P\left\{T_n - E\theta(T_n) \leq \delta\right\}$$

$$\Rightarrow P\left\{|T_n - \gamma(\theta)| \leq \delta + |E_\theta(T_n) - \gamma(\theta)|\right\} \geq 1 - \frac{Var_\theta(T_n)}{\delta^2} \qquad \text{---(1)}$$

Given , $E_\theta(T_n) \to \gamma(\theta) \quad \forall\ \theta\ \varepsilon\ \Theta \quad as\ n \to \infty$

Hence for every $\delta_1 > 0$ there exists a particular positive integer $n \geq n_0(\delta_1)$ such that

$$|E_\theta(T_n) - \gamma(\theta)| \leq \delta_1 \qquad n \geq n_0(\delta_1) \qquad \text{---(2)}$$

Also given $Var_\theta(T_n) \to 0 \quad as\ n \to \infty$

$$\frac{Var_\theta(T_n)}{\delta_2} \leq \eta \quad \forall \; n \geq n'_0(\eta) \qquad \text{---(3)}$$

where η is arbitrarily small positive number.

Substituting from eqn 2 and 3 in eqn 1 we get,

$$P\left[\left|T_n - \gamma(\theta)\right| \leq \delta + \delta_1\right] \geq 1 - \eta \quad , \quad n \geq m(\delta_1, \eta)$$

$$P\left[\left|T_n - \gamma(\theta)\right| \leq \varepsilon\right] \geq 1 - \eta \quad , \quad n \geq m$$

where m=max $(n_0 \; , n'_0)$ and $\varepsilon = \delta + \delta_1 > 0$

$$\Rightarrow T_n \xrightarrow{P} \gamma(\theta) \quad as \quad n \to \infty$$

$$\therefore T_n \; is \; a \; consistent \; estimator \; of \; \gamma(\theta)$$

UNIT-II

In the previous chapter, we have discussed different optimum properties of good point estimators, viz. Unbiasedness, minimum variance, sufficiency, efficiency and consistency. In this chapter, we shall discuss different methods of point estimation which are expected to yield estimators enjoying some of these important properties. Also we shall discuss the confidence interval for proportions, mean(s), variance(s) based on chi-square, Student's t, F and normal distributions.

2.1 METHODS OF ESTIMATION:

There are several methods in estimation theory such as

1. Method of maximum likelihood estimation
2. Method of moments
3. Method of least square
4. Method of minimum variance
5. Method of minimum chi-square
6. Method of inverse probability

METHOD OF MAXIMUM LIKELIHOOD ESTIMATION:

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

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

$$L = \prod_{i=1}^{n} f(x_i;\theta) \qquad \text{---(1)}$$

L gives the relative likelihood that the random variables assume a particular set of values $x_1, x_2, \ldots, x_n$ L becomes a function of a variable θ. The principle of maximum likelihood consist in finding an estimator for the unknown parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ which maximize the likelihood function $L(\theta)$ for variations in parameters (i.e) we want to find $\hat{\theta} = \left(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k\right)$ so that

$$L\left(\hat{\theta}\right) > L(\theta) \quad \forall\ \theta\ \varepsilon\, \Theta$$

(i.e) $L\left(\hat{\theta}\right) = \sup L(\theta) \quad \forall\ \theta\ \varepsilon\, \Theta$

Thus if there exist a function $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$ of the sample values which maximise L for variations in θ. Then $\hat{\theta}$ is to be taken as an estimator of θ. Therefore $\hat{\theta}$ is called maximum likelihood estimator. Thus $\hat{\theta}$ is the solution if any of

$$\frac{\partial L}{\partial \theta} = 0 \quad and \quad \frac{\partial^2 L}{\partial \theta^2} < 0 \qquad \text{---(3)}$$

Since L>0 and logL is a non decreasing function of L. L and logL attain their extreme value (maxima or minima) at the same value of $\hat{\theta}$. Therefore the Equation 2, can be rewritten as

$$\frac{1}{L}\frac{\partial L}{\partial \theta} = 0 \quad \Rightarrow \frac{\partial \log L}{\partial \theta} = 0 \quad and \quad \frac{\partial^2 \log L}{\partial \theta^2} < 0$$

PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATOR:

1. Maximum likelihood estimators are consistent.
2. Any consistent solution of the likelihood equation provides a maximum of the likelihood with probability tends to unity as the sample size tends to unity.
3. Asymptotic normality of MLE: A consistent solution of the likelihood equation is asymptotically normally distributed about the true value of θ_0 (i.e) $\hat{\theta}$ is asymptotically $N\left(\theta_0, \frac{1}{I(\theta_0)}\right)$ $as\ \ n \to \infty$ where $Var\left(\hat{\theta}\right) = \frac{1}{I(\theta)} = \frac{1}{-E\left(\frac{\partial^2}{\partial \theta^2}\log L\right)}$
4. If MLE exist if it is the most efficient in the class of such estimators.
5. If a sufficient estimator exist it is a function of MLE.

6. If for a given population with pdf $f(x:\theta)$ and MVBE T exist for θ then the likelihood equation will have a solution equal to the estimator T.
7. Invariance property of MLE: If T is a MLE of θ and $\psi(\theta)$ is a one-to-one function of θ then $\psi(T)$ is a MLE of $\psi(\theta)$.

Example 1: In a random sample from normal population $N(\mu,\sigma^2)$ to find the maximum likelihood estimator for the first case (i) μ when σ^2 is known (ii) σ^2 when μ is known.

Solution:

The density function of normal distribution is

$$f(x:\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\sigma^2 (x-\mu)^2} \qquad ; -\infty < x,\ \mu < \infty; \sigma > 0$$

Likelihood function is

$$L = \prod_{i=1}^{n} f(x_i : \mu, \sigma^2)$$

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

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

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

Case(i): when σ^2 is known to estimate μ

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

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

$$\frac{\partial^2 \log L}{\partial\mu^2} = \frac{-1}{\sigma^2} < 0$$

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

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

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

$\therefore$ Maximum likelihood estimator for μ when σ^2 is known is a sample mean $\bar{x}$.

Case(ii) : when μ is known to estimate σ^2

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

$$\frac{\partial^2 \log L}{\partial\sigma^4} = \frac{-n}{2}\left(\frac{-1}{\sigma^4}\right)-\frac{2}{2\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2$$

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

$$= \frac{n}{2\sigma^4}-\frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2 < 0$$

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

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

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

$\therefore$ Maximum likelihood function for σ^2 when μ is known is $\dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$

Example 2: In a random sample from Poisson distribution with parameter λ. To find maximum likelihood estimator for λ.

Solution:

The Probability density function is

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

The likelihood function is

$$L = \frac{e^{-n\lambda}\ \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

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

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

$$\frac{\partial \log L}{\partial \lambda} = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}$$

$$\frac{\partial^2 \log L}{\partial \lambda^2} = \frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0$$

$$\frac{\partial \log L}{\partial \lambda} = 0 \qquad \Rightarrow \ -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Maximum likelihood estimator for $\lambda\ is\ \dfrac{\Sigma x_i}{n} = \bar{x}$

Example 3: In a random sample from exponential distribution with parameter θ, find maximum likelihood estimator for θ.

Solution:

The Probability density function is,

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta} \quad ; \quad 0 < x < \infty,\ \theta > 0$$

The likelihood function is

$$L = \left(\frac{1}{\theta}\right)^n e^{-\sum_{i=1}^{n} x_i/\theta}$$

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

$$\frac{\partial \log L}{\partial \theta} = \frac{-n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i$$

$$\frac{\partial^2 \log L}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{1}{\theta^4}\sum_{i=1}^{n} x_i < 0$$

$$\frac{\partial \log L}{\partial \theta} = \frac{-n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i = 0$$

$$\frac{-n\theta + \Sigma x_i}{\theta^2} = 0$$

$$\hat{\theta} = \frac{\Sigma x_i}{n} = \bar{x}$$

Maximum likelihood estimator for θ is the sample mean $\bar{x}$

2.2 METHOD OF MOMENTS:

This method was discovered by Karl Pearson. Let $f(x:\theta_1,\theta_2,\ldots,\theta_k)$ be the density function of the parent population with k parameters $\theta_1,\theta_2,\ldots,\theta_k$. If μ'_r denotes the rth moment about origin then

$$\mu_r' = \int x^r \, f(x:\theta_1,\theta_2,\ldots,\theta_k)dx \quad \forall\, r = 1,2,\ldots,k \;\rightarrow 1$$

In general $\mu_1',\mu_2',\ldots,\mu_k'$ will be a function of the parameters $\theta_1,\theta_2,\ldots,\theta_k$. Let $x_i \quad \forall i = 1,2,\ldots,n$ be a random sample of size n from the given population . The method of moments consist in solving the k equations 1 for $\theta_1,\theta_2,\ldots,\theta_k$ in terms of $\mu_1',\mu_2',\ldots,\mu_k'$ and replacing these moments μ'_r for all r = 1,2,..,k by the sample moments.

For example, $\hat{\theta}_i = \theta_i(\hat{\mu}_1',\hat{\mu}_2',\ldots,\hat{\mu}_k')$

$$= \theta_i(m_1',m_2',\ldots,m_k') \qquad \forall\, i = 1,2,\ldots,k$$

where m_i' is the ith moment about the origin in the sample. Then by the method of moments $\hat{\theta}_1,\hat{\theta}_2,\ldots,\hat{\theta}_k$ are the required estimators of $\theta_1,\theta_2,\ldots,\theta_k$ respectively.

Example 4: Let X has the following distribution function

X=x	0	1	2
P(X=x)	1-$\theta-\theta^2$	θ	θ^2

Obtain the moment estimate of θ, if in a sample of 25 observations there were 10 one's an 4 two's.

Solutions:

From the given information,

X=x	$P_\theta(X = x)$	Frequency(f)
0	1-$\theta-\theta^2$	11
1	θ	10
2	θ^2	4
Total		25

μ'_1 = E(X) = 0(1-$\theta-\theta^2$) + 1(θ) + 2(θ^2)

$$= \theta + 2\theta^2$$

$$m'_1 = \frac{\Sigma fx}{N} = \frac{18}{25}$$

$$\mu'_1 = m'_1 \Rightarrow \theta + 2\theta^2 = \frac{18}{25}$$

$$\Rightarrow 25\theta + 50\theta^2 - 18 = 0$$

$$\Rightarrow 50\theta^2 + 25\theta - 18 = 0$$

$$\Rightarrow (10\theta + 9)(5\theta - 2) = 0$$

$$\Rightarrow \theta = -0.9 \text{ and } \theta = 0.41$$

Therefore, the moment estimate of θ=0.41.

Example 5: A random variable X takes the values 0,1,2 with respective probabilities $\frac{6}{4N}+\frac{1}{2}\left(1-\frac{\theta}{N}\right)$, $\frac{\theta}{2N}+\frac{\alpha}{2}\left(1-\frac{\theta}{N}\right)$, $\frac{\theta}{4N}+\frac{1-\alpha}{2}\left(1-\frac{\theta}{N}\right)$ where N is a known number and α *and* θ are unknown parameters. If 75 independent observations on X give the values 0,1,2 with frequencies 27,38,10 respectively. To estimate, α *and* θ by using method of moments.

Solution:

From the given information,

X=x	$P_\theta(X = x)$	Frequency(f)
0	$\frac{6}{4N}+\frac{1}{2}\left(1-\frac{\theta}{N}\right)$	27
1	$\frac{\theta}{2N}+\frac{\alpha}{2}\left(1-\frac{\theta}{N}\right)$	38
2	$\frac{\theta}{4N}+\frac{1-\alpha}{2}\left(1-\frac{\theta}{N}\right)$	10
Total		75

$$\mu'_1 = \mathrm{E(X)} = 0\left(\frac{6}{4N}+\frac{1}{2}\left(1-\frac{\theta}{N}\right)\right)+1\left(\frac{\theta}{2N}+\frac{\alpha}{2}\left(1-\frac{\theta}{N}\right)\right)+2\left(\frac{\theta}{4N}+\frac{1-\alpha}{2}\left(1-\frac{\theta}{N}\right)\right)$$

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

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

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

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

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

$$= 1-\frac{\alpha}{2}\left(1-\frac{\theta}{2}\right)$$

$$\mu'_2 = 1\left(\frac{\theta}{2N}+\frac{\alpha}{2}\left(1-\frac{\theta}{N}\right)\right)+4\left(\frac{\theta}{4N}+\frac{1-\alpha}{2}\left(1-\frac{\theta}{N}\right)\right)$$

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

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

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

$$= \frac{3\theta}{2N} + \left(1 - \frac{\theta}{N}\right)\left[\frac{\alpha + 4 - 4\alpha}{2}\right]$$

$$= \frac{3\theta}{2N} + \left(1 - \frac{\theta}{N}\right)\left[\frac{4 - 3\alpha}{2}\right]$$

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

$$= \frac{3\theta}{2N} + 2 - \frac{3\alpha}{2} - \frac{2\theta}{N} + \frac{3\theta\alpha}{2N}$$

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

$$m'_1 = \frac{\Sigma fx}{N} = \frac{1}{75}\left[0(27) + 1(38) + 2(10)\right]$$

$$= \frac{58}{75}$$

$$m'_2 = \frac{1}{75}\left[0^2(27) + 1^2(38) + 4(10)\right] = \frac{78}{75}$$

$$\mu'_1 = m'_1 = 1 - \frac{\alpha}{2}\left(1 - \frac{\theta}{N}\right) = \frac{58}{75}$$

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

$$\frac{\alpha}{2}\left(1 - \frac{\theta}{N}\right) = \frac{17}{75} \qquad \text{---(1)}$$

$$\mu'_2 = m'_2 = 2 - \frac{\theta}{2N} - \frac{3\alpha}{2}\left[1 - \frac{\theta}{N}\right] = \frac{78}{75}$$

$$\frac{\theta}{2N}+3\left(\frac{17}{75}\right) = 2-\frac{78}{75}$$

$$\frac{\theta}{2(75)}+\frac{17}{25} = \frac{150-78}{75}$$

$$\frac{\theta+17(6)}{150} = \frac{72}{75}$$

$$\frac{\theta+102}{150} = \frac{72}{75}$$

$$\theta+102=144$$

$$\hat{\theta}=42$$

Substituting $\theta=42$ in Equation 1,we get

$$\frac{\alpha}{2}\left(1-\frac{42}{75}\right) = \frac{17}{75}$$

$$\frac{\alpha}{2}\left(\frac{75-42}{75}\right) = \frac{17}{75}$$

$$33\alpha=34$$

$$\hat{\alpha}=\frac{34}{33}$$

Example 6: To find the moment estimator of Bernoulli population with parameter p.

Solution:

The density function of Bernoulli distribution is

$$P(X=x)=\begin{cases} pq^{1-x} & ; \; x=0\,(or)\,1\,, 0\le p\le 1\;,\; p+q=1 \\ 0 & ; \; otherwise \end{cases}$$

Raw moment of Bernoulli distribution

$$\mu'_1 = \text{p}$$

Same moment , $m'_1 = \dfrac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$

The moment estimator is $\mu'_1 = m'_1 \;\Rightarrow\; \hat{p} = \bar{x}$

Example 7: To find the moment estimator of Poisson population with parameter λ .

Solution:

$$P(X=x)=p(x)=\begin{cases} \dfrac{e^{-\lambda}\,\lambda^{x}}{x!} & ; \; x=1,2,\ldots \quad \lambda>0 \\ 0 & ; \; otherwise \end{cases}$$

Since, $\mu'_1 = \lambda$

$$m'_1 = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

$$\mu'_1 = m'_1 \;\Rightarrow \bar{x} = \hat{\lambda}$$

Example 8: To find the moment estimator of Exponential distribution with parameter θ .

Solution:

f(x) = $\theta e^{-\theta x}$; $\theta > 0$, $x = 0,1,\ldots$

Since, $\mu'_1 = \theta$

$$m'_1 = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

$$\mu'_1 = m'_1 \Rightarrow \bar{x} = \hat{\theta}$$

Example 9: To find the moment estimator of Normal distribution with parameter μ *and* σ^2.

Solution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; \quad -\infty < x , \mu < \infty , \sigma > 0$$

Since, $\mu'_1 = \mu$

$$m'_1 = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

$$\mu_2' = \mu^2 + \sigma^2$$

$$m'_2 = \frac{\sum_{i=1}^{n} x_i^{\,2}}{n}$$

$$\mu'_1 = m'_1 \Rightarrow \hat{\mu} = \bar{x}$$

$$\mu'_2 = m'_2 \Rightarrow \mu^2 + \sigma^2 = \frac{\Sigma x_i^{\,2}}{n}$$

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

$$\Rightarrow \hat{\sigma}^2 = \frac{\Sigma x_i^{\,2}}{n} - \bar{x}^2$$

2.3 METHOD OF LEAST SQUARES

For fitting a curve of the form

y=f(x; b_0,b_1,...) ---(1)

where b_0,b_1,... are unknown parameters, to the observed sample observations (x_1,y_1), (x_2,y_2),..., (x_n,y_n)by the principle of least squares, we have to minimise

$$\sum_i \{y_i - f(x_i;b_0,b_1,b_2,\ldots)\}^2 \text{ ---(2)}$$

With respect to the parameters $b_0,b_1,\ldots$.

This is the same as to minimise the sum of squares of the distances of the observed points from the curve measured in the direction of the y-axis.

In case Equation 1 is the regression equation of Y on X, $x_1,x_2,\ldots,x_n$ may be taken as observed values of the independent variable X, and Y is dependent variable and $e_i = y_i - f(x_i;b_0,b_1,b_2,\ldots)$ are the residuals or errors. If we assume that the errors are independently normally distributed with zero means and constant variance σ_e^2, then the joint probability density of the errors, or the likelihood function, is given by

$$L = Const.\exp\left[-\frac{1}{2\sigma_e^2}\sum_i \{y_i - f(x_i;b_0,b_1,b_2,\ldots)\}^2\right]$$

Hence maximising L amounts to minimizing

$$\sum_i \{y_i - f(x_i;b_0,b_1,b_2,\ldots)\}^2$$

In case e_i's are independently normally distributed with zero means and variances $\sigma_{e_i}^2$, maximizing L will amount to minimizing

$$\sum_i \frac{1}{\sigma_{e_i}^2}\{y_i - f(x_i;b_0,b_1,b_2,\ldots)\}^2$$

Which is the sum of squares of residuals each weighted by the inverse of its variance. This may be called the weighted least-squares method. In general, we may consider the regression of Y on $X_1,X_2,\ldots,X_p$ and the method of least squares appropriate for this case may be similarly deduced.

The least-squares estimators do not have any optimum properties even asymptotically. However, in linear estimation this method provides good estimators in small samples.When we are estimating $f(x_i;b_0,b_1,b_2,\ldots)$ as a linear function of the parameters $b_0,b_1,b_2,\ldots$, the x_i's being known given values, the least squares estimators obtained as linear functions of the Y's will be minimum-variance unbiased estimators.

Example 10: 1. consider $f(x) = b_0 + b_1x + b_2x^2 + \ldots + b_kx^k, where\, n > k+1.$

Here we have to minimise

$$\sum_i \left(y_i - b_0 - b_1x_i - b_2x_i^2 - \ldots - b_kx_i^k\right)^2,$$

with respect to $b_0, b_1, b_2, \ldots, b_k$. Differentiating this with respect to $b_0, b_1, b_2, \ldots, b_k$, we have k+1 equations, called the normal equations, given by

$$\sum_i x_i^j e_i = 0 \ \ (j = 0,1,2,\ldots,k) \text{ or}$$

$$\sum_i x_i^j y_i = b_0 \sum_i x_i^j + b_1 \sum_i x_i^{j+1} + \ldots + b_k \sum_i x_i^{j+k} \ (j = 0,1,2,\ldots,k)$$

Hence $b_0, b_1, b_2, \ldots, b_k$ would be obtained as linear functions of the y's.

Example 11: Consider the multiple linear regression $Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_p X_p$

Here we have to minimize $\sum \left[y_i - b_0 - b_1 x_{1i} - b_2 x_{2i} - \ldots - b_p x_{pi} \right]^2$ with respect to $b_0, b_1, b_2, \ldots, b_p$.

The Normal equations are

$$\left.\begin{aligned} &\sum_i e_i = 0 \\ &and \ \sum_i x_{ij} e_i = 0 \ (for \ \ j = 0,1,2,\ldots,p) \end{aligned}\right\}$$

or

$$\left.\begin{aligned} &\sum_i y_i = n b_0 + b_1 \sum_i x_{1i} + b_2 \sum_i x_{2i} + \ldots + b_p \sum_i x_{pi} \\ &and \ \sum_i x_{ij} y_i = b_0 \sum_i x_{ij} + b_1 \sum_i x_{ji} \, x_{1i} + b_2 \sum_i x_{ji} \, x_{2i} + \ldots + b_p \sum_i x_{ji} \, x_{pi} \ (j = 0,1,2,\ldots,k) \end{aligned}\right\}$$

and hence $b_0, b_1, b_2, \ldots, b_p$ may be obtained as linear functions of y_i's and of the given known values x's.

Definition 1: CONFIDENCE INTERVAL AND LIMITS

Let $x_1, x_2, \ldots, x_n$ be a random sample from the density $f(. , \theta)$. Let $T_1 = t_1(x_1, x_2, \ldots, x_n)$ and $T_2 = t_2(x_1, x_2, \ldots, x_n)$ be a two statistic satisfying the condition of $T_1 \le T_2$ for which $P_\theta[T_1 < \tau(\theta) < T_2] \equiv \gamma$ where γ does not depend on θ, then the random interval part $\tau(\theta)$, γ is called confidence coefficient and T_1 *and* T_2 are called lower and upper confidence limits respectively for $\tau(\theta)$. A value t_1, t_2 of the random interval T_1 *and* T_2 is also called a 100 γ % confidence interval for $\tau(\theta)$.

Definition 2: ONE SIDED CONFIDENCE INTERVAL

Let $x_1, x_2, \ldots, x_n$ be a random sample from the density $f(. , \theta)$. Let $T_1 = t_1(x_1, x_2, \ldots, x_n)$ be a statistic for which $P_\theta[T_1 < \tau(\theta)] \equiv \gamma$ then T_1 is called a one sided lower confidence for $\tau(\theta)$. Similarly,

$T_2 = t_2\left(x_1, x_2, \ldots, x_n\right)$ be a statistic for which $P_\theta\left[\tau(\theta) < T_2\right] \equiv \gamma$ then T_2 is called a one sided upper confidence for $\tau(\theta)$.

CONSTRUCTION OF CONFIDENCE INTERVAL FOR POPULATION MEAN (when the variance is known)

Let $x_1, x_2, \ldots, x_n$ be a random sample from the normal population with mean μ and variance σ^2. We take a large sample from a normal population with mean μ and SD σ . Then

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

To claim, $100(1-\alpha)\%$ confidence interval for the level of significance at 5% from the normal probability table

$$P[-1.96 \le Z \le 1.96] = 0.95$$

$$\Rightarrow P\left[-1.96 \le \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \le 1.96\right] = 0.95$$

$$\Rightarrow P\left[\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le 1.96\frac{\sigma}{\sqrt{n}} + \bar{x}\right] = 0.95$$

$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$ are 95% confidence limit for the unknown parameter μ and the interval $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \ , \ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$ is called the 95% confidence interval for μ. Also to construct $100(1-\alpha)\%$ confidence interval for the level of significance at 1% from the normal probability table

$$P[-2.58 \le Z \le 2.58] = 0.99$$

$$\Rightarrow P\left[-2.58 \le \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \le 2.58\right] = 0.99$$

$$\Rightarrow P\left[\bar{x} - 2.58\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}\right] = 0.99$$

$\bar{x} \pm 2.58\frac{\sigma}{\sqrt{n}}$ are 99% confidence limit for the unknown parameter μ and the interval $\left(\bar{x} - 2.58\frac{\sigma}{\sqrt{n}} \ , \ \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}\right)$ is called the 95% confidence interval for μ.

In general, $P(-z_\alpha \le z \le z_\alpha) = 1 - \alpha$

$$\Rightarrow P\left[-z_\alpha \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le z_\alpha\right] = 1-\alpha$$

$$\Rightarrow P\left[\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

Hence the confidence interval for μ is $\left(\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$ where z_α is the standard normal value for given level of α .

CONFIDENCE INTERVAL FOR POPULATION MEAN (when variance is unknown)

Let $x_1, x_2, \ldots, x_n$ be a random sample from the normal population with mean μ and variance σ^2. We know that population variance $s^2 = \frac{1}{n-1}\Sigma(x_i - \bar{x})^2$. A statistic $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{(n-1)}$. Hence $100(1-\alpha)\%$ confidence limit for μ is given by

$$P(|t| \le t_\alpha) = 1-\alpha$$

$$\Rightarrow P\left[\left|\frac{\bar{x}-\mu}{s/\sqrt{n}}\right|\right] = 1-\alpha$$

$$\Rightarrow P\left[\bar{x} - t_\alpha\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x} + t_\alpha\left(\frac{s}{\sqrt{n}}\right)\right] = 1-\alpha$$

where t_α is a tabulated value of student t for (n-1) degrees of freedom at significance level α . Hence required confidence interval for population mean μ is $\left(\bar{x} - t_\alpha\left(\frac{s}{\sqrt{n}}\right)\ ,\ \bar{x} + t_\alpha\left(\frac{s}{\sqrt{n}}\right)\right)$.

CONSTRUCTION OF CONFIDENCE INTERVAL FOR POPULATION VARIANCE (when mean is known)

Let $x_1, x_2, \ldots, x_n$ be a random sample from the normal population with mean μ and variance σ^2. The statistic

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

where $s^2 = \frac{1}{n}\Sigma(x_i - \mu)^2$

Let $\chi^2{}_\alpha$ at the value of χ^2 such that

$$P\left[\chi^2 > \chi^2{}_\alpha\right] = \int_{\chi^2{}_\alpha}^{\infty} P\left(\chi^2\right) d\chi^2$$

where $P\left(\chi^2\right)$ is the probability density function of χ^2 distribution with n degrees of freedom and significance level α. Thus the required confidence interval is given by

$$P\left[\chi^2{}_{1-\alpha/2} \le \chi^2 \le \chi^2{}_{\alpha/2}\right] = 1-\alpha$$

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

Now , $\dfrac{ns^2}{\sigma^2} \le \chi^2{}_{\alpha/2} \qquad \Rightarrow \dfrac{ns^2}{\chi^2{}_{\alpha/2}} \le \sigma^2$

$$\chi^2{}_{1-\alpha/2} \ge \frac{ns^2}{\sigma^2} \qquad \Rightarrow \frac{ns^2}{\chi^2{}_{1-\alpha/2}} \ge \sigma^2$$

Then, $P\left[\dfrac{ns^2}{\chi^2{}_{\alpha/2}} \le \sigma^2 \le \dfrac{ns^2}{\chi^2{}_{1-\alpha/2}}\right] = 1-\alpha$

where $\chi^2{}_{\alpha/2}$ and $\chi^2{}_{1-\alpha/2}$ are obtained from χ^2 table with n degrees of freedom and significant level α.

CONSTRUCTION OF CONFIDENCE INTERVAL FOR POPULATION VARIANCE (When Mean is Unknown)

Let $x_1, x_2, \ldots, x_n$ be a random sample from the normal population with mean μ and variance σ^2.

Here the statistic $\dfrac{\Sigma\left(x_i - \bar{x}\right)}{\sigma^2} = \dfrac{ns^2}{\sigma^2} \sim \chi^2{}_{(n-1)}$

where $s^2 = \dfrac{1}{n}\Sigma\left(x_i - \bar{x}\right)^2$

Let $\chi^2{}_\alpha$ as the value of χ^2 such that

$$P\left[\chi^2 > \chi^2{}_\alpha\right] = \int_{\chi^2{}_\alpha}^{\infty} P\left(\chi^2\right) d\chi^2$$

where $P\left(\chi^2\right)$ is the probability density function with (n-1) degrees of freedom and significance level α .Thus the required confidence interval is given by

$$P\left[\chi^2{}_{1-\alpha/2} \le \chi^2 \le \chi^2{}_{\alpha/2}\right] = 1-\alpha$$

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

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

$$\Rightarrow P\left[\frac{ns^2}{\chi^2_{1-\alpha/2}} \ge \sigma^2 \ge \frac{ns^2}{\chi^2_{\alpha/2}}\right] = 1-\alpha$$

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

where $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ are obtained from χ^2 table with (n-1) degrees of freedom and significant level α.

CONSTRUCTION OF CONFIDENCE INTERVAL FOR DIFFERENCE OF MEANS OF TWO INDEPENDENT NORMAL POPULATION WHEN VARIANCE IS KNOWN

Let $x_1, x_2, ..., x_n \sim N(\mu_x, \sigma^2_x)$ and $y_1, y_2, ..., y_n \sim N(\mu_y, \sigma^2_y)$. The statistic

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}}$$

The required confidence interval for given level of significance

$$P[-z_{\alpha/2} \le z \le z_{\alpha/2}] = 1-\alpha$$

$$\Rightarrow P\left[-z_{\alpha/2} \le \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}} \le z_{\alpha/2}\right] = 1-\alpha$$

$$\Rightarrow P\left[-z_{\alpha/2}\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}} \le \bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2) \le z_{\alpha/2}\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}\right] = 1-\alpha$$

$$\Rightarrow P\left[(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}} \le (\mu_1 - \mu_2) \le (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}\right] = 1-\alpha$$

Hence the difference of population mean confidence interval for the given level of significance α is given by $\left((\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}\ ,\ (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}\right)$ and the confidence limit is

$$(\bar{x}_1-\bar{x}_2)\pm z_{\alpha/2}\sqrt{\frac{\sigma^2{}_1}{n_1}+\frac{\sigma^2{}_2}{n_2}}$$

CONSTRUCTION OF CONFIDENCE INTERVAL FOR DIFFERENCE OF MEANS OF TWO INDEPENDENT NORMAL POPULATION WHEN VARIANCE IS UNKNOWN

$$x_1,x_2,\ldots,x_n \sim N\left(\mu_x\ ,s^2{}_x\right) \qquad y_1,y_2,\ldots,y_n \sim N\left(\mu_y\ ,s^2{}_y\right)$$

The statistic $\quad z=\dfrac{\bar{x}_1-\bar{x}_2-(\mu_1-\mu_2)}{\sqrt{\dfrac{s^2{}_1}{n_1}+\dfrac{s^2{}_2}{n_2}}}$

The required confidence interval for given level of significance

$$P\left[-z_{\alpha/2}\le z\le z_{\alpha/2}\right]=1-\alpha$$

$$\Rightarrow P\left[-z_{\alpha/2}\le \frac{\bar{x}_1-\bar{x}_2-(\mu_1-\mu_2)}{\sqrt{\dfrac{s^2{}_1}{n_1}+\dfrac{s^2{}_2}{n_2}}}\le z_{\alpha/2}\right]=1-\alpha$$

$$\Rightarrow P\left[-z_{\alpha/2}\sqrt{\frac{s^2{}_1}{n_1}+\frac{s^2{}_2}{n_2}}\le \bar{x}_1-\bar{x}_2-(\mu_1-\mu_2)\le z_{\alpha/2}\sqrt{\frac{s^2{}_1}{n_1}+\frac{s^2{}_2}{n_2}}\right]=1-\alpha$$

$$\Rightarrow P\left[(\bar{x}_1-\bar{x}_2)-z_{\alpha/2}\sqrt{\frac{s^2{}_1}{n_1}+\frac{s^2{}_2}{n_2}}\le (\mu_1-\mu_2)\le (\bar{x}_1-\bar{x}_2)+z_{\alpha/2}\sqrt{\frac{s^2{}_1}{n_1}+\frac{s^2{}_2}{n_2}}\right]=1-\alpha$$

Hence the difference of population mean confidence interval for the given level of significance α is given by $\left((\bar{x}_1-\bar{x}_2)-z_{\alpha/2}\sqrt{\dfrac{s^2{}_1}{n_1}+\dfrac{s^2{}_2}{n_2}}\ ,\ (\bar{x}_1-\bar{x}_2)+z_{\alpha/2}\sqrt{\dfrac{s^2{}_1}{n_1}+\dfrac{s^2{}_2}{n_2}}\right)$ and the confidence limit is

$$(\bar{x}_1-\bar{x}_2)\pm z_{\alpha/2}\sqrt{\frac{s^2{}_1}{n_1}+\frac{s^2{}_2}{n_2}}$$

CONSTRUCTION OF CONFIDENCE INTERVAL FOR DIFFERENCE OF MEANS OF TWO SAMPLES OF NORMAL POPULATION WITH COMMON VARIANCE (COMMON VARIANCE IS UNKNOWN)

Let $x_1,x_2,\ldots,x_m$ be a random sample from the normal population with mean μ_1 and variance $\sigma^2{}_1$. Let $y_1,y_2,\ldots,y_n$ be a random sample from the normal population with mean μ_2 and variance $\sigma^2{}_2$. Assume that the two samples are independent to each other. Let $\bar{y}-\bar{x}$ is normally distributed with mean $\mu_2-\mu_1$ and variance $\dfrac{\sigma_2}{m}+\dfrac{\sigma^2}{n}$ (i.e) $(\bar{y}-\bar{x})\sim N\left(\mu_2-\mu_1\ ,\dfrac{\sigma_2}{m}+\dfrac{\sigma^2}{n}\right)$. $\dfrac{\Sigma(x_i-\bar{x})^2}{\sigma^2}$ is chi-square distributed with (m-1) degrees of freedom

$$\frac{\Sigma(x_i-\bar{x})^2}{\sigma^2}\sim\ \chi^2{}_{(m-1)} \quad \text{and}$$

$$\frac{\Sigma(y_i-\bar{y})^2}{\sigma^2}\sim\ \chi^2{}_{(n-1)}$$

$$\therefore \frac{\Sigma(x_i-\bar{x})^2}{\sigma^2}+\frac{\Sigma(y_i-\bar{y})^2}{\sigma^2}\sim\chi^2{}_{(m+n-2)}$$

The statistic

$$Q=\frac{(\bar{y}-\bar{x})-(\mu_2-\mu_1)}{\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right).\,s_{p^2}}}\sim F_{(m+n-2)}$$

Thus the confidence interval for difference of means for two samples of normal population with the given level of significance α

$$P[-t_{\alpha/2}\le Q\le t_{\alpha/2}]\ =1-\alpha$$

$$\Rightarrow P\left[-t_{\alpha/2}<\frac{(\bar{y}-\bar{x})-(\mu_2-\mu_1)}{\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right).\,s_{p^2}}}<t_{\alpha/2}\right]=1-\alpha$$

$$\Rightarrow P\left[(\bar{y}-\bar{x})-t_{\alpha/2}\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right).\,s_{p^2}}<(\mu_2-\mu_1)<(\bar{y}-\bar{x})+t_{\alpha/2}\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right).\,s_{p^2}}\ \right]=1-\alpha$$

Hence $100(1-\alpha)\%$ confidence interval is $\left((\bar{y}-\bar{x})-t_{\alpha/2}\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right).\,s_{p^2}}\ ,\ (\bar{y}-\bar{x})+t_{\alpha/2}\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right).\,s_{p^2}}\right)$ and the confidence limits are

$$(\bar{y}-\bar{x})\pm t_{\alpha/2}\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right).\,s_{p^2}}$$

Suppose the samples are dependent on each other with common variance. Let $D_i=y_i-x_i\quad\forall\ i=1,2,\ldots,n$ *then* $D_1,D_2,\ldots,D_n$ are independently identically distributed random variables with common normal distribution having mean $\mu_D=\mu_2-\mu_1$ and variance $\sigma^2{}_D=\sigma_1{}^2+\sigma_2{}^2-2\rho\sigma_1\sigma_2$

$\therefore 100(1-\alpha)\%$ confidence interval for $\mu_D = \mu_2 - \mu_1$ is

$$\left(\overline{D} - t_{\alpha/2}\sqrt{\frac{\Sigma\left(D_i - \overline{D}\right)^2}{n(n-1)}} \;,\; \overline{D} + t_{\alpha/2}\sqrt{\frac{\Sigma\left(D_i - \overline{D}\right)^2}{n(n-1)}} \right)$$

where $t_{\alpha/2}$ is the $\alpha/2$ th quartile point of the t-distribution with (n-1) degrees of freedom.