# BHARATHIDASAN UNIVERSITY

## Tiruchirappalli- 620024

## Tamil Nadu, India.


## Programme: M.Sc. Statistics

## Course Title: Statistical Inference-I

## Course Code: (23ST06CC)

# Unit-I

# Estimation Theory


**Dr. T. Jai Sankar**

**Associate Professor and Head**

**Department of Statistics**

**Ms. N. Saranya**

**Guest Faculty**

**Department of Statistics**

## Introduction:

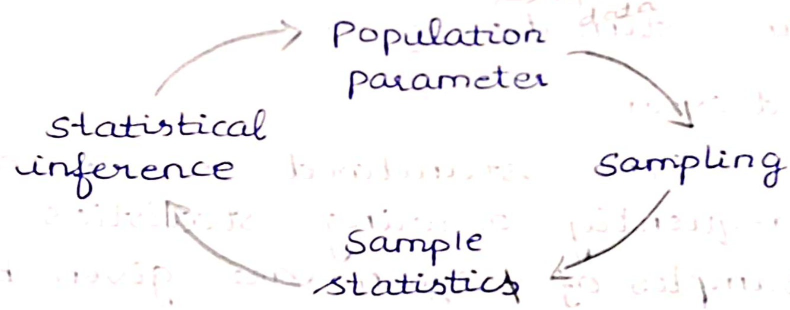
One of the main objectives of statistics is to drawn inference about a population from the analysis of a sample drawn from the population.

Two important problems in statistical inference are

* Estimation Theory
* Testing of Hypothesis

## Definition:

In statistics, we usually want to statistically analysis a population but collecting data for the whole population is usually impractical, expensive and unavailable. That is the collect samples of the population and analysis thus samples to drawn some conclusion about the population. It is called the inference.




## Parameter:

Any population constraints are called parameter. For example, A random variable $X \sim N(\mu, \sigma^2)$. Here $\mu$ and $\sigma^2$ are called the parameters or normal distribution.

## Parametric space:

The set of all possible values of the parameter is called parametric space.

i.e. $X \sim f(x:\theta) \;\forall\; \theta \in \Theta \rightarrow$ parameter

It is denoted as $\Theta$.

For example,

$$X \sim N(\mu, \sigma^2) \;\forall\; \Theta = \{\theta = \mu, \sigma^2\} \quad -\infty < \mu < \infty,\ \sigma > 0$$

Population:

The set of all possible observations under study is called the population. It is denoted by N.

Sample:

It is the subset of the population. It is denoted by n.

Estimator:

Any functions of the random sample $x_1, x_2 \dots x_n$ that are being observed say $T_n(x_1, x_2, \dots x_n)$ is called a statistic. Clearly a statistic is a random variable. If it is used to estimate an unknown parameter $\theta$ of the distribution, it is called an estimator. A particular value of the estimator say $T_n(x_1, x_2 \dots x_n)$ is called an estimate of $\theta$.

Standard error:

The standard errors of the some frequently occuring statistics for large samples of size n are given below where $\sigma^2$ is the population variance, P is the population proportion and $Q = 1 - P$ and $n_1, n_2$ respectively. Size of two independent random samples drawn from the given population.

| Statistic | Standard Error |
|---|---|
| Sample mean ($\bar{x}$) | $\sigma/\sqrt{n}$ |
| Sample Proportion p | $\sqrt{PQ/n}$ |
| Sample S.D (s) | $\sqrt{\sigma^2/2n}$ |
| Sample Variance ($s^2$) | $\sigma^2/\sqrt{2/n}$ |

| | |
|---|---|
| Difference of two sample means $(\bar{x}_1 - \bar{x}_2)$ | $\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ |
| Difference of two sample S.D $(S_1, S_2)$ | $\sqrt{\sigma_1^2/2n_1 + \sigma_2^2/2n_2}$ |
| Difference of two sample proportions $(P_1, P_2)$ | $\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$ |

**Definition of Standard Error:**

The standard deviation of the sampling distribution of a statistic is known as its standard error.

**Null Hypothesis and Alternative Hypothesis:**

**Null Hypothesis:**

The Null Hypothesis is the hypothesis which is tested for possible rejection under the assumption that it is true. It is denoted by $H_0$.

**Alternative Hypothesis:**

Any hypothesis which is complementary to the null Hypothesis is called the Alternative Hypothesis and it is denoted by $H_1$.

**Critical Region: (Rejection Region)**

A region corresponding to a statistic (t) in the sample space S which amounts to rejection of $H_0$ is termed as critical region or region of rejection.

i) $P(t \in \omega / H_0) = \alpha$

ii) $P(t \in \bar{\omega} / H_1) = \beta$

where,

$\bar{\omega}$; the complementary set of $\omega$ is called the acceptance region.

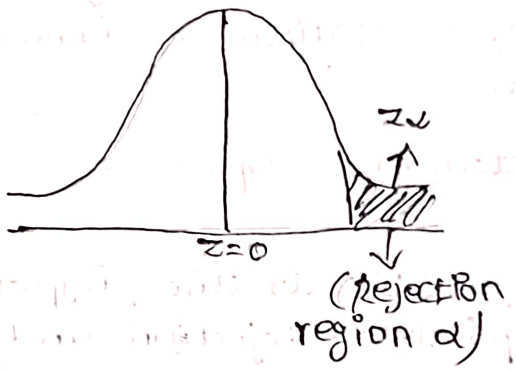
## One-tailed and two-tailed test:

A test of any statistical hypothesis where the alternative hypothesis is one kind (right tailed or left tailed) is called one-tailed test.

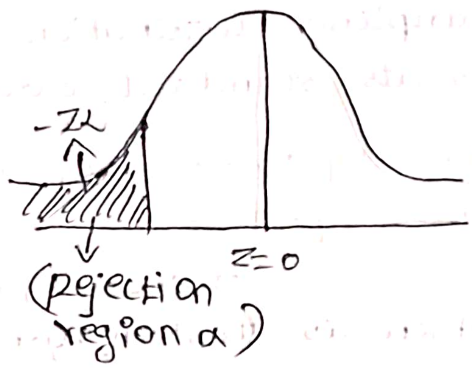
$H_0 : \mu > \mu_0$ (Right tailed)

$H_1 : \mu < \mu_0$ (Left tailed)

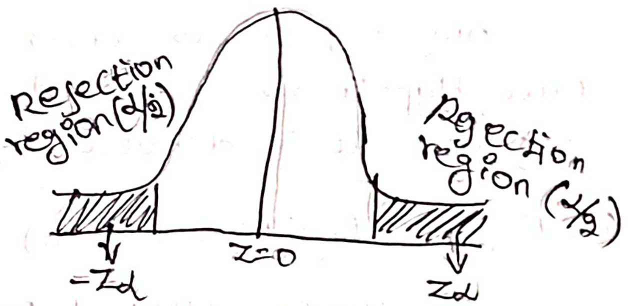
Right tailed test:-




Left tailed test




Two-tailed test




A test of statistical hypothesis where the alternative hypothesis is two tailed such as

$H_0 : \mu = \mu_0$ against the alternative hypothesis. $H_1 : \mu \neq \mu_0$ is known as two-tailed test.

## Characteristics of good estimator:.

* Unbiasedness
* Consistency
* Efficiency
* Sufficiency

* Unbiasedness:

An estimator is said to be unbiased if its expected value is equal to its population parameter (i.e) $E(\hat{\theta}) = \theta$. where $\hat{\theta}$ is point estimator and $\theta$ is population parameter.

For example, $\mu_{\bar{x}} = E(\bar{x}) = \mu$ here $\bar{x}$ is unbiased estimator and biased $E(\hat{\theta}) \neq \theta$

Example 1:

If $x_1, x_2 \dots x_n$ is random sample from a normal population $N(\mu, 1)$. Show that $t = \frac{1}{n} \sum_{i=1}^{n} x_i^2$ is an unbiased estimator of $\mu^2+1$

Soln:

We are given

$E(x_i) = \mu$ and $V(x_i) = 1$, $i = 1, 2 \dots n$

Now,

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

Hence t is an unbiased estimator of $1+\mu^2$.

Example 2:

If t is an unbiased estimator for $\theta$ show that $t^2$ is an biased estimator for $\theta^2$.

Soln:

Since T is an unbiased estimator of $\theta$ we have $E(T) = \theta$

Also, $V(T) = E(T^2) - [E(T)]^2$

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

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

Since $E(T^2) \neq \theta^2$

$T^2$ is a biased estimator for $\theta^2$.

Example 3:

Show that $\frac{1}{n(n-1)} \sum x_i (\sum x_i - 1)$ is an unbiased estimator of $\theta^2$ for the sample $x_1, x_2 \dots x_n$ drawn on $x$ which taken the value 0 or 1 with respective probabilities $\theta$ and $(1-\theta)$.

Soln:

Since $x_1, x_2 \dots x_n$ a random sample from Bernoulli population with parameter $\theta$.

$$T = \sum_{i=1}^{n} x_i \sim B(n, \theta)$$

$$\Rightarrow \quad E(T) = n\theta \text{ and } V(T) = n\theta(1-\theta)$$

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

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

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

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

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

$$= \theta^2$$

$\Rightarrow \left\{\sum x_i (\sum x_i - 1)\right\} / n(n-1)$ is an unbiased estimator of $\theta^2$.

Unbiasedness:

Definition:

An estimator $T_n = T(x_1, x_2 \dots x_n)$ is said to be an unbiased estimator of $\gamma(\theta)$ if $E(T_n) = \gamma(\theta) \ \forall \ \theta \in \Theta$

In sampling from a population with mean $\mu$ and variance $\sigma^2$, $E(\bar{x}) = \mu$ and $E(s^2) \neq \sigma^2$ but $E(S^2) = \sigma^2$. Hence, there is a reason to prefer

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

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

Consistency:

An estimator $T_n = T(x_1, x_2 \dots x_n)$ based on a random sample of size n, is said to be consistent estimator of $\gamma(\theta)$, $\theta \in \Theta$ the parameter space, if $T_n$ converges to $\gamma(\theta)$ in probability.

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

In other words, $T_n$ is a consistent estimator of $\gamma(\theta)$ if for every $\varepsilon > 0$, $\eta > 0$ there exists a positive integer $n \geq m(\varepsilon, \eta)$ such that

$P\{|T_n - \gamma(\theta)| < \varepsilon\} \to 1$ as $n$

$\Rightarrow P\{|T_n - \gamma(\theta)| < \varepsilon\} > 1 - \eta$ ; $\forall\, n \geq m$

where m is some very large value of n.

Invariance Property of consistent estimators:

Theorem

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

$T_n$

Proof:

Since $T_n$ is a consistent estimator of $\gamma(\theta)$, $T_n \xrightarrow{P} \gamma(\theta)$ as $n \to \infty$ i.e) for every $\varepsilon > 0$, $\eta > 0$, $\exists$ a positive integer $n \geq m(\varepsilon, \eta) \ni P\{|T_n - \gamma(\theta)| < \varepsilon\} > 1 - \eta$ $\forall\, n \geq m \to (1)$

Since $\psi(\cdot)$ is a continuous function for $\gamma(\theta)$ for every $\varepsilon > 0$, however small, $\exists$ a positive number $\varepsilon_1 \ni |\psi(T_n) - \psi(\gamma(\theta))| < \varepsilon_1$ whenever $|T_n - \gamma(\theta)| < \varepsilon$.

i.e) $|T_n - \gamma(\theta)| < \varepsilon \Rightarrow |\psi(T_n) - \psi(\gamma(\theta))| < \varepsilon_1 \to (2)$

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

$A \subseteq B \Rightarrow P(A) \leq P(B)$ or $P(B) \geq P(A) \to (3)$

From (2) & (3) we get,

$$P\left[|\psi(T_n)-\psi(\gamma(\theta))|<\xi_1\right]\geq P\left[|T_n-\gamma(\theta)|<\xi\right]$$

$$\Rightarrow P\left[|\psi(T_n)-\psi(\gamma(\theta))|<\xi\right]\geq 1-\eta \;;\; \forall\, n\geq m \quad \text{using } ①$$

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

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

**Sufficient Conditions for consistency:**

Theorem:

Let $\{T_n\}$ be a sequence of estimators such that for all $\theta \in \Theta$.

i) $E_\theta(T_n)\to\gamma(\theta)$ as $n\to\infty$ and

ii) $Var_\theta(T_n)\to 0$ as $n\to\infty$

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

Proof:

We have 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) $P\left[|T_n-\gamma(\theta)|<\xi\right]>1-\eta,\ \forall\, n\geq m(\xi,\eta)\to ①$

where $\xi$ and $\eta$ are arbitrarily small positive numbers and $m$ is some large value of $n$.

Applying chebyshev's inequality to the statistic $T_n$, we get

$$P\left[|T_n-E_\theta(T_n)|<\delta\right]\geq 1-\frac{Var_\theta(T_n)}{\delta^2}\to ②$$

we have,

$$|T_n-\gamma(\theta)|=|T_n-E(T_n)+E(T_n)-\gamma(\theta)|$$

$$\leq |T_n-E_\theta(T_n)|+|E_\theta(T_n)-\gamma(\theta)|\to ③$$

Now, $\quad |T_n-E_\theta(T_n)|\leq\delta$

$$\Rightarrow |T_n - \gamma(\theta)| \leq \delta + |E_\theta(T_n) - \gamma(\theta)| \rightarrow (4)$$

Hence on using, if $A \Rightarrow B$ then

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

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

$$\geq 1 - \frac{Var_\theta(T_n)}{\delta^2} \rightarrow (5) \quad (\text{from } (2))$$

We are given: $E_\theta(T_n) \rightarrow \gamma(\theta) \ \forall \ \theta \in \Theta$ as $n \rightarrow \infty$

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

$$|E_\theta(T_n) - \gamma(\theta)| \leq \delta_1 \quad \forall \ n \geq n_0(\delta_1) \rightarrow (6)$$

Also $var_\theta(T_n) \rightarrow 0$, as $n \rightarrow \infty$ (Given)

$$\therefore \frac{Var_\theta(T_n)}{\delta^2} \leq \eta, \quad \forall \ n \geq n_0'(\eta) \rightarrow (7)$$

where $\eta$ is already arbitrarily small positive number

Substituting (6) & (7) in (5), we get,

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

$$\Rightarrow P[|T_n - \gamma(\theta)| \leq \varepsilon] \geq 1 - \eta \ ; \ n \geq m$$

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

$\Rightarrow T_n \xrightarrow{P} \gamma(\theta)$ as $n \rightarrow \infty$ [using (1)]

$\therefore T_n$ is a constant estimator of $\gamma(\theta)$.

**Example:-**

If $X_1, X_2 \ldots X_n$ are random observations on a Bernoulli variate $X$ taking the value 1 with probability $p$ and the value 0 with probability $(1-p)$; show that $\frac{\Sigma x_i}{n}\left(1 - \frac{\Sigma x_i}{n}\right)$ is a consistent estimator of $P(1-P)$.

Soln:-

Since $X_1, X_2 \dots X_n$ are i.i.d Bernoulli variates with parameter 'p',

$$T = \sum_{i=1}^{n} X_i \sim B(n, p)$$

$$\Rightarrow E(T) = np \text{ and } var(T) = npq \rightarrow ①$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{T}{n} \Rightarrow E(\bar{x}) = \frac{1}{n} E(T)$$

$$= \frac{1}{n} \cdot np$$

$$= n \quad [\text{From } ①]$$

$$Var(\bar{x}) = Var\left(\frac{T}{n}\right) = \frac{1}{n^2} var(T)$$

$$= \frac{pq}{n} \rightarrow 0 \text{ as } n \rightarrow \infty$$

Since $E(\bar{x}) \rightarrow p$ and $var(\bar{x}) \rightarrow 0$ as $n \rightarrow \infty$

$\bar{X}$ is a consistent estimator of p.

Also, $\frac{\Sigma x_i}{n}\left(1 - \frac{\Sigma x_i}{n}\right) = \bar{x}(1-\bar{x})$, being a polynomial in $\bar{X}$, is a continuous function of $\bar{x}$. Since $\bar{x}$ is consistent estimator of p.

By the invariance property of consistent estimator of $p(1-p)$.

## Efficient estimator:-

There is a necessity of some further criterion which will enable us to choose between the estimators with the common property of consistency. Such a criterion which is based on the variances of the sampling distribution of estimators is usually known as efficiency.

If of the two consistent estimators $T_1, T_2$ of the certain parameter $\theta$, we have,

$$V(T_1) < V(T_2) \text{ for all } n.$$

Then, $T_1$ is more efficient than $T_2$ for all sample sizes.

## More efficient estimator:-

If in a class of consistent estimators for a parameter, there exists one whose sampling variance is less than that of any such estimator it is called the more efficient estimator. Whenever, such an estimator exists, it provides a criterion for measurement of efficiency of the other estimators.

## Efficiency (Defn):-

If $T_1$ is the most efficiency estimator with variance $V_1$ and $T_2$ is any other estimator with variance $V_2$, then efficiency E and $T_2$ is defined as

$$E = V_1/V_2$$

obviously, E cannot exceed unity

ie) $0 \leq E \leq 1$

If $T_1, T_2 \dots T_n$ are all estimators of $\gamma(\theta)$ and $var(T)$ is minimum, then the efficiency $E_i$ of $T_i$ $(i=1,2\dots n)$ is defined as

$$E_i = \frac{Var(T)}{Var(T_i)} \; ; \; i=1,2\dots n$$

obviously,

$$E_i \leq 1 \; ; \; i=1,2\dots n.$$

## Example:-

$X_1, X_2$ and $X_3$ is a random sample of size 3 from a population with mean $\mu$ and variance $\sigma^2$. $T_1, T_2, T_3$ are the estimators used to estimate mean value $\mu$ where $T_1 = X_1+X_2-X_3$.
$T_2 = 2X_1+3X_3-4X_2$ and $T_3 = (\lambda X_1+X_2+X_3)/3$

i) Are $T_1$ and $T_2$ unbiased estimators?

ii) Find the value of $\lambda$ such that $T_3$ is an unbiased estimator of $\mu$.

iii) With this value of $\lambda$, is $T_3$ a consistent estimator?

iv) Which is the best estimator?

Soln::

Since $X_1, X_2, X_3$ is a random sample from a population with mean $\mu$ and variance $\sigma^2$:

$$\left.\begin{array}{l} E(X_i) = \mu,\ Var(X_i) = \sigma^2 \text{ and} \\ cov(X_i, X_j) = 0\ (i \neq j = 1, 2, \ldots n \end{array}\right\} \rightarrow ①$$

i) we have [on using ①]

$$E(T_1) = E(X_1) + E(X_2) - E(X_3)$$
$$= \mu + \mu - \mu$$
$$= \mu$$

$\Rightarrow T_1$ is an unbiased estimator of $\mu$.

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

$\therefore T_2$ is an unbiased estimator of $\mu$.

ii) We are given: $E(T_3) = \mu$ [From ①]

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

iii) with $\lambda = 1$, $T_3 = \frac{1}{3}\left[X_1 + X_2 + X_3\right] = \bar{X}$

Since the sample mean is a consistent estimator of population mean $\mu$. By weak law of large numbers, $T_3$ is a consistent estimator of $\mu$.

iv) We have [on using ①]

$$Var(T_1) = Var(X_1) + Var(X_2) + Var(X_3)$$
$$= \sigma^2 + \sigma^2 + \sigma^2$$
$$= 3\sigma^2$$

$$Var(T_2) = 4\,Var(X_1) + 9\,Var(X_3) + 16\,Var(X_2)$$
$$= 4\sigma^2 + 9\sigma^2 + 16\sigma^2$$
$$= 29\sigma^2$$

$$Var(T_3) = \frac{1}{9}\left[Var(X_1) + Var(X_2) + Var(X_3)\right]$$
$$= \frac{1}{9}(3\sigma^2)$$
$$= \frac{\sigma^2}{3}$$

Since $var(T_3)$ is minimum,

$\therefore T_3$ is the best estimator of $\mu$ in the sense of minimum variance.

**Uniformly Minimum Variance Unbiased Estimator (UMVUE) or Minimum Variance Unbiased (MVU)**

If a statistic $T = T(x_1, x_2 \dots x_n)$ based on sample of size n is such that:

i) T is unbiased for $\gamma(\theta)$ $\forall$ $\theta \in \Theta$ and

ii) It has the smallest variance among the class of all unbiased estimators of $\gamma(\theta)$

then T is called minimum variance unbiased estimator (MVUE) of $\gamma(\theta)$.

More precisely, T is MVUE of $\gamma(\theta)$ if

$$E_\theta(T) = \gamma(\theta) \quad \forall \; \theta \in \Theta$$

and $Var_\theta(T) \leq Var_\theta(T') \;\forall\; \theta \in \Theta$

where $T'$ is any other unbiased estimator of $\gamma(\theta)$.

**Theorem:**

An MVU is unique in the sense that if $T_1$ and $T_2$ are MVU estimators for $\gamma(\theta)$, then $T_1 = T_2$ almost surely.

**Proof:**

$$\left.\begin{aligned} &E_\theta(T_1) = E_\theta(T_2) = \gamma(\theta) \;\forall\; \theta \in \Theta \\ &\text{and } Var_\theta(T_1) = Var_\theta(T_2) \;\forall\; \theta \in \Theta \end{aligned}\right\} \rightarrow (1)$$

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

$$E(T) = \frac{1}{2}\{E(T_1) + E(T_2)\} = \gamma(\theta) \quad [\text{From } ①]$$

$$\begin{aligned}
var(T) &= var\{\tfrac{1}{2}(T_1 + T_2)\} \\
&= \tfrac{1}{4}\{var(T_1 + T_2)\} \\
&= \tfrac{1}{4}\{var(T_1) + var(T_2) + 2\,cov(T_1, T_2)\} \\
&= \tfrac{1}{4}\{var(T_1) + var(T_2) + 2\rho\sqrt{var(T_1)\,var(T_2)}\} \\
&= \tfrac{1}{4}\{var(T_1) + var(T_1) + 2\rho\sqrt{var(T_1)\,var(T_1)}\} \quad [\text{From } ①] \\
&= \tfrac{1}{4}\{2\,var(T_1) + 2\rho\,var(T_1)\} \\
&= \tfrac{2}{4}\,var(T_1)\{1 + \rho\} \\
&= \tfrac{1}{2}\,var(T_1)(1 + \rho)
\end{aligned}$$

where $\rho$ is the Karl Pearson's coefficient of correlation between $T_1$ & $T_2$.

Since $T_1$ is the MVU estimator,

$$var(T) \geq var(T_1)$$

$$\Rightarrow \tfrac{1}{2}\,var(T_1)(1+\rho) \geq var(T_1)$$

$$\Rightarrow \tfrac{1}{2}(1+\rho) \geq 1$$

$$\Rightarrow 1 + \rho \geq 2$$

$$\Rightarrow \rho \geq 2 - 1$$

$$\Rightarrow \rho \geq 1$$

Since $|\rho| \leq 1$, we must have $\rho = 1$.

i.e) $T_1$ and $T_2$ must have a linear relation of form:

$$T_1 = \alpha + \beta T_2 \longrightarrow ②$$

where $\alpha$ and $\beta$ are constants, independent of $x_1, x_2 \ldots x_n$ but may depend on $\theta$.

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

Taking expectation on both sides in ② and using ① we get,

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

$$\theta = \alpha + \beta\theta \rightarrow ③$$

Also from ② we get,

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

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

$$Var(T_1) = 0 + \beta^2 Var(T_2)$$

$$1 = \beta^2$$

$$\beta = \pm 1 \qquad \because Var(T_1) = Var(T_2)$$

But since $\rho(T_1, T_2) = \pm 1$, the coefficient of regression of $T_1$ and $T_2$ must be positive.

$$\beta = 1 \Rightarrow \alpha = 0 \quad [\text{from } ③]$$

Substituting in ② we get,

$T_1 = T_2$, as desired.

Theorem:

If $T_1$ is an MVUE of $\gamma(\theta)$ and $T_2$ is any other unbiased estimator of $\gamma(\theta)$ with efficiency $e < 1$, then no unbiased linear combination of $T_1$ and $T_2$ can be an MVUE of $\gamma(\theta)$.

Proof:

A linear combination: $T = l_1 T_1 + l_2 T_2 \rightarrow ①$ will be an unbiased estimator of $\gamma(\theta)$ if

$$E(T) = l_1 E(T_1) + l_2 E(T_2) = \gamma(\theta) \quad \forall \ \theta \in \Theta$$

$$\Rightarrow l_1 + l_2 = 1 \rightarrow ②$$

Since we are given $E(T_1) = E(T_2) = \gamma(\theta)$

we have,

$$e = \frac{Var(T_1)}{Var(T_2)}$$

$$\Rightarrow Var(T_2) = \frac{Var(T_1)}{e} \rightarrow ③$$

$$\text{and } \rho = \rho(T_1, T_2) = \sqrt{e} \rightarrow ④$$

From ①

$$Var(T) = l_1^2 \, var(T_1) + l_2^2 \, var(T_2) + 2 l_1 l_2 \, cov(T_1, T_2)$$

$$= l_1^2 \, var(T_1) + l_2^2 \, var(T_2) + 2 l_1 l_2 \, \rho \sqrt{var(T_1) \, var(T_2)}$$

Using ③ & ④ we get,

$$Var(T) = l_1^2 \, var(T_1) + l_2^2 \frac{var(T_1)}{e} + 2 l_1 l_2 \, \rho \sqrt{var(T_1) \frac{var(T_1)}{e}}$$

$$= l_1^2 \, var(T_1) + l_2^2 \, var(T_1) + 2 l_1 l_2 \, \rho \frac{var(T_1)}{\sqrt{e}}$$

$$= var(T_1) \left[ l_1^2 + \frac{l_2^2}{e} + 2 l_1 l_2 \right] \quad [\text{from ④}]$$

$$> var(T_1) \left[ l_1^2 + 2 l_1 l_2 + l_2^2 \right] \quad \left[ \because 0 < e < 1 \Rightarrow 1/e > 1 \right]$$

$$Var(T) > var(T_1) (l_1 + l_2)^2$$

$$var(T) > var(T_1) \quad [\text{from ②}]$$

T cannot be a MVUE.

## Sufficiency:

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

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

## Factorization theorem (Neyman)

The necessary and sufficient condition for a distribution to admit sufficient statistic is provided by the factorization theorem due to Neyman.

Statement:

A statistic $T = t(x)$ is sufficient for $\theta$ if and only if the joint pmf or pdf $L$ (say), of the sample values can be expressed in the form:

$$L = g_\theta[t(x)] \cdot h(x)$$

where $g_\theta[t(x)]$ depends on $\theta$ and $x$ only through the values of $t(x)$ and $h(x)$ is independent of $\theta$.

Proof: [For discrete case]

Suppose that $T$ is sufficient.

Let $g_\theta[t(x)] = P_\theta[t(X) = T]$ and

$h(x) = P[X = x \mid t(X) = t(x)]$

then

$$f_\theta(x) = P_\theta(X = x) = P_\theta[X = x, t(X) = t(x)]$$

$$L = P_\theta[t(X) = t(x)] P[X = x \mid t(X) = t(x)]$$

$$L = g_\theta[t(x)] \cdot h(x)$$

Suppose now that $L = g_\theta[t(x)] h(x)$ for $x \in \mathcal{X}$

Let $q_\theta(t)$ be the pmf of $t(X)$ and

$A_x = \{y : t(y) = t(x)\}$ then for any $x \in \mathcal{X}$

$$L = g_\theta[t(x)] \cdot h(x)$$

$$f_\theta(x) = g_\theta[t(x)] h(x)$$

$$\frac{f_\theta(x)}{q_\theta[t(x)]} = \frac{g_\theta[t(x)] \cdot h(x)}{q_\theta[t(x)]} = \frac{g_\theta[t(x)] h(x)}{P_\theta[t(X) = t(x)]}$$

$$= \frac{g_\theta[t(x)] h(x)}{\sum_{y \in A_x} f_\theta(y)}$$

$$= \frac{g_\theta[t(x)] h(x)}{\sum_{y \in A_x} g_\theta[t(y)] h(y)}$$

$$= \frac{g_\theta[t(x)]\, h(x)}{g_\theta[t(x)] \sum_{y \in A_x} h(y)}$$

$$= \frac{h(x)}{\sum_{y \in A_x} h(y)}$$

which does not depend on $\theta$.

i.e) $T$ is sufficient for $\theta$.

Hence proved.

Invariance property of sufficient estimator:

If $T$ is a sufficient estimator for the parameter $\theta$ and if $\psi(T)$ is a one to one function of $T$, $\psi(T)$ is sufficient for $\psi(\theta)$.

Fisher - Neyman's Criterior:

A statistic $t = t(x_1, x_2 \dots x_n)$ is sufficient estimator of parameter $\theta$ if and only if the likelihood function (joint pdf of the sample) can be expressed as:

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

where $g_1(t_1, \theta)$ is the pdf of the statistic $t_1$ and $k(x_1 x_2 \dots x_n)$ is a function of sample observations only independent of $\theta$.

Example:

Let $x_1 x_2 \dots x_n$ be a random sample from a uniform population on $[0, \theta]$. Find a sufficient estimator of $\theta$.

Soln:

We are given:

$$f_\theta(x_i) = \begin{cases} 1/\theta & ; \ 0 \le x_i < \theta \\ 0 & ; \ \text{otherwise} \end{cases}$$

Let $k(a,b) = \begin{cases} 1 & \text{if } a \le b \\ 0 & \text{if } a > b \end{cases}$

then $f_\theta(x_i) = \dfrac{k(0, x_i) \cdot k(x_i, \theta)}{\theta}$

$$L = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \left[ \frac{k(0,x_i) \cdot k(x_i, \theta)}{\theta} \right]$$

$$= \frac{k\left(0, \min_{1 \le i \le n} x_i\right) k\left(\max_{1 \le i \le n} x_i, \theta\right)}{\theta^n}$$

$$= g_\theta\{t(x)\} h(x)$$

where

$$g_\theta[t(x)] = \frac{k[t(x), \theta]}{\theta^n}$$

$t(x) = \max_{1 \le i \le n} (x_i)$ and $h(x) = k\left(0, \min_{1 \le i \le n} (x_i)\right)$

Hence by factorization theorem, $T = \max_{1 \le i \le n} (x_i)$ is sufficient statistic for $\theta$.

Example:

Let $x_1, x_2 \dots x_n$ be a random sample from $N(\mu, \sigma^2)$ population. Find sufficient estimators for $\mu$ and $\sigma^2$.

Soln:

Let us write $\theta = (\mu, \sigma^2)$; $-\infty < \mu < \infty$
$0 < \sigma^2 < \infty$

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

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

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

$$= g_\theta[t(x)] h(x)$$

where,

$$g_\theta[t(x)] = \left\{\frac{1}{\sigma\sqrt{2\pi}}\right\}^n \exp\left\{-\frac{1}{2\sigma^2}\left[t_2(x) - 2\mu t_1(x) - n\mu^2\right]\right\}$$

$t(x) = \{t_1(x), t_2(x)\} = (\Sigma x_i, \Sigma x_i^2)$ and $h(x) = 1$

Thus $t_1(x) = \Sigma x_i$ is sufficient for $\mu$
and $t_2(x) = \Sigma x_i^2$ is sufficient for $\sigma^2$.

Example:

Let $X_1, X_2, \ldots X_n$ be a random sample from a population with p.d.f $f(x, \theta) = \theta x^{\theta-1}; 0 < x < 1$, $\theta > 0$. S.T $t_1 = \prod_{i=1}^{n} X_i$ is sufficient for $\theta$.

Soln:

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

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

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

$$= g(t, \theta)\, h(x_1, x_2 \ldots x_n)$$

Hence by factorization theorem,
$t_1 = \prod_{i=1}^{n} (X_i)$ is sufficient estimator for $\theta$.

Minimal Sufficiency:

A sufficient statistic $T(x)$ is a minimal sufficient statistic, if for any other sufficient statistic
$U(x) = T(x)$ is a function of $U(x)$.

$T(x)$ is a function of $U(x)$ if and only if $U(x) = U(y)$ implies that $T(x) = T(y)$ for any point $(x, y)$.

Theorem:

Let $f_\theta(x)$ be the p.m.f or p.d.f of X. Suppose that $T(x)$ is sufficient for $\theta$ and that for every pair $x$ and $y$ which at least one of $f_\theta(x)$ and $f_\theta(y)$ is not $\theta$, $f_\theta(x) | f_\theta(y)$ does not depend on $\theta$ implies $T(x) = T(y)$. Then $T(X)$ is minimum sufficient for $\theta$.

Proof:

Let $U(x)$ be another sufficient statistic. By factorization theorem, there are functions $h$ and $g_\theta$ such that $f_\theta(x) = g_\theta[U(x)]h(x)$ $\forall x$ and $\theta$. For $x$ and $y$ such that atleast one of $f_\theta(x)$ and $f_\theta(y)$ is not $\theta$.

i.e) atleast one of $h(x)$ and $h(y)$ is not 0.

if $U(x) = U(y)$ then

$$\frac{f_\theta(x)}{f_\theta(y)} = \frac{g_\theta[(U(x))]h(x)}{g_\theta[U(y)]h(y)} = \frac{h(x)}{h(y)}$$

which does not depends on $\theta$. By the assumption of the theorem

$$T(x) = T(y).$$

This shows that there is a function $\psi$ such that $T(x) = \psi[S(x)]$

Hence $T$ is minimal sufficient for $\theta$.

Complete family of distribution:

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

$$E_\theta[h(t)] = 0 \ \forall \theta$$

$$\Rightarrow P_\theta[h(t) = 0] = 1$$

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

or $\Sigma h(t)\, g(t,\theta) = 0 \quad \forall \theta \in \Theta$

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

Example:-

Let $X_1, X_2 \ldots X_n$ be a random sample of size $n$ from $N(\theta, 1)$ population. Examine if $T = t(x) = X_1$ is complete for $\theta$.

Soln:

We have $T = X_1$; $\Theta = \{\theta;\ -\infty < \theta < \infty\}$ so that $E_\theta[h(t)] = 0 \ \forall \theta$.

$$\Rightarrow \int_{-\infty}^{\infty} h(u)\, e^{-(u-\theta)^2/2}\, du = 0 \quad \forall \theta \in \Theta$$

$[\because T = X_1 \sim N(\theta, 1)]$

$$\Rightarrow \int_{-\infty}^{\infty} \left\{ h(u)\, e^{-u^2/2} \right\} e^{\theta u} = 0 \ \forall \theta \in \Theta$$

This is a bilateral Laplace Transform in $\theta$. Since these are unique:

$$h(u) \cdot e^{-u^2/2} = 0 \quad a.s.$$

$$\Rightarrow h(u) = 0 \quad a.s.$$

$$\Rightarrow P\{h(T) = 0\} = 1 \quad \forall \theta \in \Theta$$

$\therefore T = X_1$ is complete statistic for $\theta$.

**Example:**

Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(0, \theta)$. Prove that $T = X_1$ is not complete statistic for $\theta$ but $T_1 = X_1^2$ is complete for $\theta$.

**Soln:**

Here $T = t(x) = X$ ; $\Theta = \{\theta : 0 < \theta < \infty\}$

$$E_\theta[h(t)] = 0 \ \forall \ \theta \in \Theta$$

$$\Rightarrow \int_{-\infty}^{\infty} h(u) \exp\left\{ -u^2/(2\theta) \right\} du = 0 \ \forall \theta \in \Theta$$

This holds only for all odd functions $h(u)$ for $u$, for which the integral exists.

i.e., for all functions

S.T : $h(u) = -h(-u) \ \forall \ u$

$\Rightarrow h(u) \neq 0$ a.s. or

$T = X_1$ is not complete statistic for $\theta$.

Let us consider the statistic,

$$T_1 = X_1^2$$

$$E_\theta[h(T_1)] = 0 \ \forall \ \theta \in \Theta$$

$$\Rightarrow \int_{-\infty}^{\infty} h(x^2) \exp\left\{ -x^2/2\theta \right\} dx = 0 \quad \forall \theta \in \Theta$$

$$\Rightarrow \int_{-\infty}^{\infty} \frac{h(u)}{\sqrt{u}} \cdot \exp\left\{ \frac{-u}{2\theta} \right\} du = 0 \ \forall \ \theta \in \Theta$$

This being Laplace transform in $(1/\theta)$, we have

$$\frac{h(u)}{\sqrt{u}} = 0, \text{ a.s}$$

$$\Rightarrow h(u) = 0 \text{ a.s.}$$

$T_1 = X_1^2$ is complete statistic for $\theta$.

## MVUE and Blackwellisation:

How to obtain MVU estimator from any unbiased estimator through the use of sufficient statistic. This technique is called Blackwellization after D. Blackwell. The result is contained in the following theorem due to C.R. Rao and D. Blackwell.

## Rao- Blackwell Theorem:

Let $X$ and $Y$ be random variables such that $E(Y) = \mu$ and $Var(Y) = \sigma_Y^2 > 0$.

Let $E(Y|X = x) = \phi(x)$. Then

i) $E[\phi(x)] = \mu$ and

ii) $Var[\phi(x)] \leq Var(Y)$

**Proof:**

Let $f_{XY}(x,y)$ be the joint p.d.f of random variable $X$ and $Y$, $f_1(\cdot)$ and $f_2(\cdot)$ the marginal pdf's of $X$ and $Y$ respectively and $h(y|x)$ be the conditional pdf of $Y$ for given $X = x$ such that $h(y|x) = f(x,y) | f(x)$

$$E[Y | X = x] = \int_{-\infty}^{\infty} y . h(y|x) dy$$

$$= \int_{-\infty}^{\infty} y \frac{f(x,y)}{f_1(x)} dy$$

$$= \frac{1}{f_1(x)} \int_{-\infty}^{\infty} y f(x,y) dy$$

$$E(Y|X = x) = \frac{\phi(x) . f_1(x)}{f_1(x)}$$

$$E(Y|X = x) = \phi(x) \rightarrow (1)$$

$$\left[\because \int_{-\infty}^{\infty} y f(x,y) dy = \phi_x f_1(x)\right]$$

From ① we observe that the conditional distribution of $Y$ given $X = x$ does not depend on the parameter $\mu$. Hence $X$ is sufficient statistic for $\mu$. Also,

$$E[\phi(X)] = E\{E(Y|X=x)\} = E(Y) = \mu$$

$$\therefore E[\phi[X]] = \mu$$

Part i) established.

$$\text{Now, } Var(Y) = E[Y - E(Y)]^2 = E[Y-\mu]^2$$

$$= E(Y + \phi(X) - \phi(X) - \mu]^2$$

$$= E[Y - \phi(X)]^2 + E[\phi(X) - \mu]^2$$

$$+ 2E[X - \phi(x)] E[\phi(X) - \mu] \rightarrow ②$$

The product term gives

$$E\{[Y - \phi(X)][\phi(X) - \mu]\} =$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{[y - \phi(x)][\phi(x) - \mu]\} f(x,y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [y - \phi(x)][\phi(x) - \mu] f_1(x) h[y|x]\, dx\, dy$$

$$= \int_{-\infty}^{\infty} [\phi(x) - \mu] \left\{\int_{-\infty}^{\infty} [y - \phi(x)] h(y|x)\, dy\right] f_1(x)\, dx$$

$$\text{But } \int_{-\infty}^{\infty} [y - \phi(x)] h(y|x)\, dy = 0$$

$$\because E(Y|X=x) = \phi(x)$$

$$\therefore E\{[Y - \phi(x)] - [\phi(x) - \mu]\} = 0 \rightarrow ③$$

Sub ③ in ②

$$Var(Y) = E(Y - \phi(X)]^2 + E[\phi(X) - \mu]^2$$

WKT

$$Var[\phi(X)] = E[\phi(X)]^2 - \{E[\phi(x)]\}^2$$

$$= E[\phi(x)]^2 - \mu^2$$

$$\therefore Var[\phi(X)] = E[\phi(x) - \mu]^2$$

$$\therefore Var(X) = E[Y - \phi(X)]^2 + Var[\phi(X)]$$

$$Var(Y) \geq var[\phi(X)]$$

$$[\because E(Y - \phi(x)] \geq 0]$$

$\therefore Var[\phi(x)] \leq Var(Y)$

$\therefore$ Part ii) established.

Hence the theorem.

Lehmann - Scheffe Theorem:

Let $Y = (Y_1, Y_2 \dots Y_n)^T$ be random sample. If $S(Y)$ is a jointly complete sufficient statistic and $T(Y)$ is an unbiased estimator for $\phi = g(\theta)$ then $U = E[T|S]$ is probability with 1. a unique MVUE of $\phi$.

Proof:

First to prove that U is a MVUE of $g(\theta)$ We show that whatever unbiased estimator $T(Y)$. We take, we obtain the same $E[T|S]$ is the same U.

Then by Rao-Blackwell Theorem, condition(b): U must be MVUE of $g(\theta)$.

Suppose that $T(Y)$ and $T'(Y)$ are any two unbiased estimator of $g(\theta)$.

Let $U = E[T|S]$ ; $U' = E[T'|S]$

Then we have

$$E[U-U'] = E\{E[T|S] - E[T'|S]\}$$
$$= E(T) - E(T')$$
$$= g(\theta) - g(\theta)$$
$$= 0$$

Hence by completeness of $S(Y)$, we get

$$P[U(S|Y) = U'(S|Y)] = 1 \quad \forall \theta$$

This proves the 1st part of theorem.

Now uniqueness,

Suppose that U and T are MVUE of $g(\theta)$ then if $T^*$ is a function of sufficient statistic $S(Y)$, then as shown above, it must be equal to U.

If $T^*$ is not a function of $S(Y)$ the

$var(U) < var(T^*)$

Hence $T^*$ cannot be MVUE.

Hence, $U$ is a unique MVUE of $g(\theta)$.