Statistical Inference

Statistical inference is mainly concerned with making inferences regarding the unknown aspects of the distribution of the population based on samples taken from it. The unknown aspect may be the form of the distribution or values of the parameters involved or both. Statistical inference is broadly classified into two

1. Estimation of parameters
2. Testing of hypotheses

Estimation deals with methods of determining numbers which may be taken as the values of the unknown parameters (called the point Estimation) as well as with the determination of intervals which will contain the unknown parameters with a specified probability (known as interval estimation), based on samples taken from the population.

Testing of hypotheses deals with the methods for deciding either to accept or reject the hypotheses based on samples taken from the population, with the degree of validity of the decision indicated in terms of probability.

Statistics

Any function of the sample values is known as a statistics.

Estimation

The objective of estimation is to determine approximate value of a population parameter on the basis of a sample statistics.

In statistics, an estimator is a function of the data or sample that is used to infer the value of of an unknown parameter in population in a statistical model. So we can define a an estimator as

Any value or function of the sample values suggested as the value of the parameters is known as the ***Estimator***. Thus, an estimator is a rule, usually a formula, that tells you how to calculate the estimate based on the sample.

There are two type of estimators:

1. **Point estimator**
   In this case we find a single statistic value that is the best guess for the parameter value.
2. **Interval Estimator**
   In this case we find an interval in which the actual value of the parameter lies with some probability level called the confidence level

Point Estimation

An estimate of a population parameter given by a single number is called point estimate.

Point estimator

A point estimator is a statistic for estimating the population parameter θ and will be denoted by $\hat{\theta}$.

Methods of Estimation

Suppose we have a random sample $X_1, X_2, \ldots, X_n$ whose assumed probability distribution depends on some unknown parameter θ. Our primary goal here will be to find a point estimator $t(X_1, X_2, \ldots, X_n)$, such that $t(x_1, x_2, \ldots, x_n)$ is a "good" point estimate of θ, where $x_1, x_2, \ldots, x_n$ are the observed values of the random sample. For example, if we plan to take a random sample $X_1, X_2, \ldots, X_n$ for which the $X_i's$ are assumed to be normally distributed with mean μ and variance σ^2., then our goal will be to find a good estimate of μ, say, using the data $x_1, x_2, \ldots, x_n$ that we obtained from our specific random sample. There are several methods to find out the point estimate of the parameter θ based on a random sample $X_1, X_2, \ldots, X_n$ from the population. Among them most widely used methods are the methods of M

1. Method of moments
2. Method of maximum likelihood

Now we consider each of them in details

1. **Method of moments**

Let f($\theta_1, \theta_2 \ldots \theta_k$)be the p.d.f of the population and let $x_1, x_2, \ldots x_n$ be a random sample taken the population. In the method of moments we find the first k moments of the population and equate them to the corresponding moments of the sample to obtain k equations. Then the values of $\theta_1, \theta_2 \ldots \theta_k$ which are obtained as the solutions of these equations are taken as their estimates. In short, the method of moments involves equating sample moments with theoretical moments.

Definitions.

$\mu'_k = E(X^k)$ is the (theoretical) moment of the distribution or population (**about the origin**), for k=1,2,…. .

$\mu_k = E(X - \mu_1^1)^k$ is the (theoretical) moment of the distribution (**about the mean**), for k=1,2,…. .

$M'_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ is the sample moment, for k=1,2,…. .

$M_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$ is the sample moment about the mean, for k=1,2,…. .

The basic idea behind this form of the method is to:
Equate the first sample moment about the origin $M'_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ to the first theoretical moment $\mu'_1 = E(X) = \mu$.

Equate the second sample moment about the origin $M_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ to the second theoretical moment $\mu_2' = E(X^2)$.

.

Continue equating sample moments about the origin, M_k' with the corresponding theoretical moments $\mu_k' = E(X^k)$, k=3,4,….until you have as many equations as you have parameters. Solve for the parameters.

The resulting values are called **moments estimators**. It seems reasonable that this method would provide good estimates, since the empirical distribution converges in some sense to the probability distribution. Therefore, the corresponding moments should be about equal.

Example-1: Let $X_1, X_2, \dots, X_n$ be Bernoulli random variables with parameter *p*. What is the moments estimator of *p*?

$$\mu_1' = E(X) = p$$

So equating the first row theoretical moment with sample moment we get

$$\mu_1' = E(X) = p = M_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

or

$$\hat{p} = \bar{X}$$

Example-2: Let $X_1, X_2, \dots, X_n$ be normally variable with mean μ and variance σ². What are the moment estimate of μ and variance σ².

The first and second moment about the origin are

$$\mu_1' = E(X) = \mu \text{ and } \mu_2' = E(X^2) = \mu^2 + \sigma^2$$

Equating this with first and second raw moment we get

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

and

$$\mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \quad \text{or} \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \mu^2$$

So

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = S^2$$

Another form of the method is

Equate the first sample moment about the origin $M_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ to the first theoretical moment $\mu_1' = E(X) = \mu$.

Equate the second sample moment about the mean $M_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ to the second theoretical moment about the mean $\mu_2 = E(X-\mu)^2$

Continue equating sample moments about the mean $M_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$ with the corresponding theoretical moments about the mean $\mu_k = E(X-\mu)^k$ for k=1,2,…. . until you have as many equations as you have parameters. Solve for the parameters.

Consider the Example 2 above

Equating this with first and second raw moment we get

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

Equate the second sample moment about the mean $M_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ to the second theoretical moment about the mean we get

$\mu_2 = E(X-\mu)^2 = \sigma^2 = M_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

So we have

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S^2$$

2. **Method of maximum likelihood**

Let f(x, $\theta_1, \theta_2 \ldots \theta_k$) be the p.d.f of the population where $\theta_1, \theta_2 \ldots \theta_k$ are the parameters. Let $x_1, x_2, \ldots x_n$ be a random sample taken the population . The likelihood function of the sample is,

$L(x_1, x_2, \ldots x_n : \theta_1, \theta_2 \ldots \theta_k) = f(x_1, \theta_1, \theta_2 \ldots \theta_k)\, f(x_2, \theta_1, \theta_2 \ldots \theta_k) \ldots f(x_n, \theta_1, \theta_2 \ldots \theta_k)$
For any given sample this may regarded as a function of unknown parameters $\theta_1, \theta_2 \ldots \theta_k$. Those values of $\theta_1, \theta_2 \ldots \theta_k$ which maximizes the likelihood function are called maximum likelihood estimates of $\theta_1, \theta_2 \ldots \theta_k$).

Maximum likelihood estimate are found to be unique in most of the case and they can be obtained by the differential calculus. If there is only one parameter θ, then that value of θ which satisfies $\frac{dL}{d\theta} = 0$ and $\frac{d^2L}{d\theta^2} < 0$ will give the mle of θ. If there are *k* parameters we have to solve the equations $\frac{\partial L}{\partial \theta_1} = 0, \frac{\partial L}{\partial \theta_2} = 0, \ldots, \frac{\partial L}{\partial \theta_k} = 0$, together with the method of checking of the maximum will give the mle's.

Note : A value of θ which maximize L also maximize log L also. So we use the equations $\frac{\partial logL}{\partial \theta_i} = 0$, i=1,2, …,*k*. to find mle.

Method Of Finding Estimators :—

(A) Maximum Likelihood Estimators :—

To introduce the method of maximum likelihood estimation, consider a simple estimation problem:
Suppose an urn contains a number of black and white balls and it is known that the ratio of the number is 3:1 but it is unknown whether black or white ball are more numerous. The probability of drawing a black is either $\frac{1}{4}$ or $\frac{3}{4}$. If 3 balls are drawn WR, the distn of the number of black balls (X) is given by $f(x;p)=\binom{3}{x}p^x q^{3-x}$, $x=0(1)3$, where $p\in\Omega=\left\{\frac{1}{4},\frac{3}{4}\right\}$.

To estimate p, based on an observed value x of X. The possible outcomes and their probabilities are given below:

Outcome	0	1	2	3
$f(x;\frac{1}{4})$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$
$f(x;\frac{3}{4})$	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$

If $x=0$ is observed, then a sample with $x=0$ is more likely (in the sense of having larger probability) to arise from a popln. with $p=\frac{1}{4}$ than from one with $p=\frac{3}{4}$ and consequently $\hat{p}=\frac{1}{4}$ would be preferred over $\hat{p}=\frac{3}{4}$.
Hence, the estimate may be defined as:

$$\hat{p}(x)=\begin{cases} 1/4 & , x=0,1 \\ 3/4 & , x=2,3 \end{cases}$$

and then the estimator is $\hat{p}(x)$. The estimator $\hat{p}(x)$ selects the value of p, say $\hat{p}(x)$ such that $f(x,\hat{p})>f(x,p')$, where, p' is an alternative value of p $\forall x$.

<u>Likelihood Function</u>:— Let $(x_1,x_2,\dots,x_n)$ be an observed random sample from a popln with PDF or PMF $f(x;\theta)$, $\theta\in\Omega$. Then, for given $(x_1,x_2,\dots,x_n)$, $L(\underset{\sim}{x};\theta)=\prod_{i=1}^{n} f(x_i;\theta)$, as a function of θ, is called the Likelihood function or the Likelihood of the sample $\underset{\sim}{x}$.

[When x is discrete R.V.s, the larger the prob. $P[X=x;\theta]$ or $f(x;\theta)$, the more likely the value x to occur. Hence, $f(x;\theta)$, for given x, gives the likeliness of the value x, for different $\theta\in\Omega$.

When X is continuous RV with PDF $f(x;\theta)$, then $P[x-h/2<X<x+\frac{h}{2}]\simeq f(x;\theta)\cdot h$ for small $h>0$. Therefore, $f(x;\theta)$, for given x, represents the likeliness of the value x. Note that, the Likelihood function $f(x;\theta)$ is a point function, it can't be a probability function or set function,]

- Maximum Likelihood Estimators :-

If a sample $\underset{\sim}{x} = (x_1, \ldots, x_n)$ is observed from a popln, we believe that the sample is "most likely to occur". When a sample $\underset{\sim}{x}$ is observed, we want to find the value of $\theta \in \Omega$ which maximizes the likelihood function $L(x;\theta)$ or $L(\theta/\underset{\sim}{x})$. The value of $\theta \in \Omega$, ~~which maximizes likelihood function $L(\theta/x)$~~ a function of $\underset{\sim}{x}$, say $\hat{\theta}(\underset{\sim}{x})$, if it exists. Then the random variable $\hat{\theta}(\underset{\sim}{X})$ is called the Maximum Likelihood Estimator (MLE) of θ.

Ex.(1):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $Bin(1,p)$; $p \in (0,1) = \Omega$. Find MLE of p.

Solution:- The Likelihood function is

$$L(p/\underset{\sim}{x}) = \begin{cases} p^{\sum_{i=1}^{n} x_i} (1-p)^{n-\sum_{i=1}^{n} x_i} & ; x_i = 0,1, \forall i = 1(1)n. \\ 0 & ; ow \end{cases}$$

where, $p \in \Omega = (0,1)$.

when $\sum_{i=1}^{n} x_i \neq 0$ or $\neq n$, then

$$\ln L(p/\underset{\sim}{x}) = (\Sigma x_i) \ln p + (n - \Sigma x_i) \ln(1-p)$$

and
$$\frac{\partial}{\partial p} \ln L = \frac{\Sigma x_i}{p} + \frac{n - \Sigma x_i}{(1-p)}(-1)$$

$$= \frac{n\bar{x}}{p} + \frac{n(1-\bar{x})}{(1-p)}$$

$$= \frac{n(\bar{x}-p)}{p(1-p)} \begin{cases} > 0 & \text{iff } p < \bar{x} \\ < 0 & \text{iff } p > \bar{x} \end{cases}$$

Hence, $L(p/\underset{\sim}{x})$ first increases, then achives its maximum at $p = \bar{x}$ and finally decreases.

Hence $L(p/\underset{\sim}{x})$ is maximum at $p = \bar{x}$.

When $\sum_{i=1}^{n} x_i = 0$, i.e. $\underset{\sim}{x} = \underset{\sim}{0}$, then

$$L(p/\underset{\sim}{x} = 0) = (1-p)^n \downarrow p \text{ and it is maximum at } p = 0 \notin \Omega = (0,1).$$

When $\sum_{i=1}^{n} x_i = n$, i.e. $\underset{\sim}{x} = \underset{\sim}{1}$, then

$L(p/\underset{\sim}{x} = \underset{\sim}{1}) = p^n \uparrow p$ and it is maximum at $p = 1 \notin \Omega$.

Hence, when $\sum_{i=1}^{n} x_i \neq 0$, or, $\neq n$, the MLE of $p \in \Omega = (0,1)$ is $\hat{p} = \bar{X}$; ow the MLE of $p \in (0,1)$ does not exist when $\sum_{i=1}^{n} x_i = 0$ or n.

Remark:- Let $(x_1, x_2, \dots, x_n)$ be a r.s. from Bernoulli(p), $0<p<1$. If $(x_1, \dots, x_n) = (0, 0, \dots, 0)$ or $(1, 1, \dots, 1)$ then MLE of p does not exist.

Ex.(2):- Let $X_1, \dots, X_n$ be a r.s. from $P(\lambda)$, $\lambda>0$. Find the MLE of λ.

Solution:- Let $X_1, X_2, \dots, X_n$ be a r.s. from $P(\lambda)$, $\lambda>0$.

The Likelihood function is

$$L(\lambda/\underset{\sim}{x}) = e^{-n\lambda} \cdot \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i} \text{ ; } x_i = 0, 1, 2, \dots \text{ ; } \lambda>0$$

$$\ln L = \ln L(\lambda/\underset{\sim}{x}) = -n\lambda + \sum_{i=1}^{n} x_i \cdot \ln\lambda - \sum_{i=1}^{n} \ln x_i$$

$$\frac{\partial}{\partial\lambda}\ln L = -n + \frac{\Sigma x_i}{\lambda} = -n + \frac{n}{\lambda}\cdot\bar{x} = \frac{-n\lambda + n\bar{x}}{\lambda} = \frac{n}{\lambda}(\bar{x}-\lambda)$$

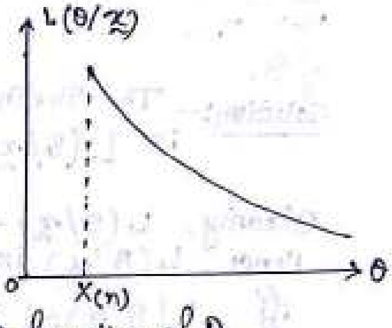
$$= \frac{n}{\lambda}(\bar{x}-\lambda) \begin{cases} >0 & \text{if } \bar{x}>\lambda \\ <0 & \text{if } \bar{x}<\lambda \end{cases}$$

Hence, $L(\lambda/\underset{\sim}{x})$ first increases, then achives it's maximum point at $\bar{x}=\lambda$ and then decreases.

Hence, $L(\lambda/\underset{\sim}{x})$ is maximum at $\lambda = \bar{x}$.

Ex.(3):- [An example of an MLE which is not unbiased]

Let $X_1, \dots, X_n$ be a r.s. from $U(0, \theta)$, $\theta>0$. Find MLE of θ. Show that it is not unbiased.



Solution:- The likelihood function is

$$L(\theta/\underset{\sim}{x}) = \begin{cases} \frac{1}{\theta^n}, & \text{if } 0 \le x_i \le \theta,\ i = 1(1)n \\ 0, & \text{ow} \end{cases}$$

$$= \begin{cases} \frac{1}{\theta^n} & \text{if } 0 \le x_{(1)} \le x_{(n)} \le \theta \\ 0, & \text{ow} \end{cases}$$

For $\theta \geq x_{(n)}$, $L(\theta/\underset{\sim}{x}) = \frac{1}{\theta^n}$ is a decreasing function of θ.

Hence, $L(\theta/\underset{\sim}{x})$ is maximum iff $\theta\ (\geq x_{(n)})$ is minimum iff $\theta = x_{(n)}$.

Hence, the MLE of θ is $\hat{\theta} = x_{(n)}$.

Note that, MLE $(\hat{\theta}) = x_{(n)}$ is consistent, complete sufficient but not unbiased.

Note that, for $X_{(n)}$; $f(x_{(n)}) = \frac{nx^{n-1}}{\theta^n}$ and $E[X_{(n)}] = \int_0^\theta \frac{nx^n}{\theta^n}\,dx = \frac{n\theta}{n+1}$,

i.e. $E(X_{(n)}) = \frac{n\theta}{n+1} \Rightarrow E\left(\frac{n+1}{n}\hat{\theta}\right) = \theta$

$\Rightarrow$ MLE $\hat{\theta}$ is not unbiased, but $\frac{n+1}{n}\hat{\theta}$ is unbiased for θ.

Ex.(4):- Let $X_1, \dots, X_n$ be a r.s. from $U(\alpha, \beta)$. Find the MLE of (α, β).

Solution:- The Likelihood function is

$$L(\alpha, \beta | \underset{\sim}{x}) = \begin{cases} \frac{1}{(\beta-\alpha)^n}, & \alpha \le x_i \le \beta. \\ 0 & , \text{OW} \end{cases}$$

$$= \begin{cases} \frac{1}{(\beta-\alpha)^n}, & \alpha \le x_{(1)} \le x_{(n)} \le \beta \\ 0 & , \text{OW} \end{cases}$$

Now, $L(\alpha, \beta / \underset{\sim}{x})$ is maximum iff

$L(\alpha, \beta / \underset{\sim}{x}) = \frac{1}{(\beta-\alpha)^n}$ is maximum subject to the restriction $\alpha \le x_{(1)} \le x_{(n)} \le \beta$, i.e. iff the length $(\beta - \alpha)$ is minimum subject to $\alpha \le x_{(1)}$ and $\beta \ge x_{(n)}$.

[Note that, $\alpha \le x_{(1)}$, $\beta \ge x_{(n)} \Rightarrow \beta - \alpha \ge x_{(n)} - x_{(1)}$
$\Rightarrow (\beta - \alpha)$ attains its minimum when $\beta = x_{(n)}$ & $\alpha = x_{(1)}$.]

i.e. iff $\beta = x_{(n)}$, $\alpha = x_{(1)}$.

Hence, the MLE of α, β is $(\hat{\alpha}, \hat{\beta}) = (x_{(1)}, x_{(n)})$.

Ex.(5):- [An example of MLE which is not unique]

Let $X_1, \dots, X_n$ be a r.s. from $U(\theta - \frac{1}{2}, \theta + \frac{1}{2})$. Find the MLE of θ.

Solution:- The likelihood function of the sample $\underset{\sim}{x} = (x_1, \dots, x_n)$ is $L(\theta / \underset{\sim}{x}) = \begin{cases} 1 & \text{if } \theta - 1/2 \le x_{(1)} \le x_{(n)} \le \theta + \frac{1}{2} \\ 0 & , \text{OW} \end{cases}$

Clearly, $L(\theta / \underset{\sim}{x})$ takes only two values 1 and 0.

Hence, $L(\theta / \underset{\sim}{x})$ is maximum

iff $L(\theta / \underset{\sim}{x}) = 1$ iff $\theta - \frac{1}{2} \le x_{(1)} \le x_{(n)} \le \theta + \frac{1}{2}$

iff $x_{(n)} - \frac{1}{2} \le \theta \le x_{(1)} + \frac{1}{2}$ ———— $(*)$

Hence, any statistic $T(\underset{\sim}{X})$ such that

$X_{(n)} - \frac{1}{2} \le T(\underset{\sim}{X}) \le X_{(1)} + \frac{1}{2}$. is an MLE of θ.

Clearly, for $0 \le \alpha \le 1$,

$$T_\alpha(\underset{\sim}{X}) = \alpha\left(X_{(n)} - \frac{1}{2}\right) + (1-\alpha)\left(X_{(1)} + \frac{1}{2}\right)$$

lies in the interval $(*)$, hence, for each $\alpha \in [0,1]$ $T_\alpha(\underset{\sim}{X})$ is an MLE of θ.

Hence, MLE of θ is not unique.

Ex.(8):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from one of the following two PDFs

If $\theta = 0$, $f(x/\theta) = \begin{cases} 1 & , 0<x<1 \\ 0 & , \text{ow} \end{cases}$

If $\theta = 1$, $f(x/\theta) = \begin{cases} \frac{1}{2\sqrt{x}} & , 0<x<1 \\ 0 & , \text{ow} \end{cases}$

Find the MLE of θ.

Solution:- The Likelihood function is

$L(\theta/\underset{\sim}{x}) = \prod_{i=1}^{n} f(x_i/\theta)$, $\theta \in \Omega = (0,1)$

When $\theta = 0$, $L(\theta/\underset{\sim}{x}) = \begin{cases} 1 & \text{if } 0<x_i<1 \ \forall i = 1(1)n \\ 0 & \text{ow} \end{cases}$

When $\theta = 1$, $L(\theta/\underset{\sim}{x}) = \begin{cases} \frac{1}{2^n \sqrt{\prod_{i=1}^{n} x_i}} & , 0<x_i<1 , i=1(1)n \\ 0 & , \text{ow} \end{cases}$

Now, $\frac{L(\theta=1/\underset{\sim}{x})}{L(\theta=0/\underset{\sim}{x})} \gtrless 1$

iff $\frac{1}{\sqrt{4^n G^n}} \gtrless 1$, where $G = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$

iff $4G \lesseqgtr 1$ iff $G \lesseqgtr \frac{1}{4}$

Hence MLE of θ is $\hat{\theta} = \begin{cases} 1 & \text{if } G < \frac{1}{4} \\ 0 & \text{if } G > \frac{1}{4} \\ 0,1 & \text{if } G = \frac{1}{4} \end{cases}$

Remark:- (1) When Ω is an open interval of $\mathbb{R}$ and $f(x;\theta)$ or $L(\theta/\underset{\sim}{x})$ is differentiable w.r.t. θ, the MLE is a solution of $\frac{\delta}{\delta\theta} L(\theta/\underset{\sim}{x}) = 0 \Leftrightarrow \frac{\delta}{\delta\theta} \ln L(\theta/\underset{\sim}{x}) = 0$ ——— (*)

This is known as Likelihood equation.

If Ω is an open interval of $\mathbb{R}$, there may still be many problems. Often, the likelihood equation has more than one roots or $L(\theta/\underset{\sim}{x})$ is not differentiable everywhere in Ω, the MLE $(\hat{\theta})$ is a terminated point, then the differentiation method of maximization is not applicable.

(2) When more than one parameters are involved in $f(x;\underset{\sim}{\theta})$, $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k) \in \Omega \subseteq \mathbb{R}^k$. If Ω is an open region of $\mathbb{R}^k$, then the MLE's of θ_i's are the solution of

$$\frac{\partial \ln L}{\partial \theta_i} = 0 \ \forall \ i = 1(1)k. \text{ and}$$

$$\left(\left(\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j}\right)\right)_{\underset{\sim}{\theta} = \hat{\underset{\sim}{\theta}}} \text{ is n.d.}$$

Ex.(9):- Let $X_1, \ldots, X_n$ be a r.s. from $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}, \sigma > 0$. Find the MLE of (μ, σ^2).

Solution:- Likelihood function:

$$L(\mu, \sigma^2 / \underset{\sim}{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \; ; x_i \in \mathbb{R}$$

where $\mu \in \mathbb{R}, \sigma > 0$.

$$\Rightarrow \ln L(\mu, \sigma^2/\underset{\sim}{x}) = \text{constant} \left(-\frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right)$$

$$0 = \frac{\partial \ln L}{\partial \mu} = -\frac{1}{2\sigma^2}\sum 2(x_i-\mu)(-1) = \frac{\sum x_i}{\sigma^2} - \frac{n\mu}{\sigma^2}$$

$$0 = \frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum(x_i-\mu)^2}{2\sigma^4}$$

$$\Rightarrow \begin{cases} \hat{\mu} = \overline{x} \\ \hat{\sigma}^2 = \frac{1}{n}\sum(x_i-\overline{x})^2 \end{cases}$$

, the likelihood function has a unique solution.

Note that, the matrix of second order partial derivatives at $(\hat{\mu}, \hat{\sigma}^2)$ is

$$\begin{pmatrix} \frac{\partial^2 \ln L}{\partial \mu^2} & \frac{\partial^2 \ln L}{\partial\mu\,\partial\sigma^2} \\ \frac{\partial^2 \ln L}{\partial\sigma^2\,\partial\mu} & \frac{\partial^2 \ln L}{\partial(\sigma^2)^2} \end{pmatrix}_{(\mu,\sigma^2) = (\hat{\mu}, \hat{\sigma}^2)}$$

$$= \begin{pmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4} \end{pmatrix}$$ is negative definite (n.d.).

Hence, $L(\mu, \sigma^2/\underset{\sim}{x})$ is maximum at $(\mu, \sigma^2) = (\hat{\mu}, \hat{\sigma}^2)$.

Therefore, the MLE of (μ, σ^2) is

$(\hat{\mu}, \hat{\sigma}^2) = (\overline{X}, S^2)$, where $nS^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2$.

Ex.(10):- Let $X_1, \ldots, X_n$ be a r.s. from $f(x; \mu, \sigma) = \frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}}$; $x \in \mathbb{R}$, where $\mu \in \mathbb{R}, \sigma > 0$. Find the MLE of μ and σ.

Solution:- The log-likelihood function is

$$\ln L(\mu, \sigma / \underset{\sim}{x}) = -n\ln 2 - n\ln\sigma - \frac{1}{\sigma}\sum|x_i - \mu| \; ; \mu \in \mathbb{R}, \sigma > 0$$

[As $\sum|X_i - \mu|$ is not differentiable w.r.t. μ, hence the derivative technique is not applicable for maximizing $\ln L$ w.r.t. μ]

We adopt two stage maximization :-

First fix σ, then maximize $\ln L$ for variation in μ.
For fixed σ, $\ln L$ is maximum,
iff, $\Sigma|x_i-\mu|$ is minimum
iff, $\mu = \tilde{x} =$ the sample median
$= \hat{\mu}$, say.

Now, we maximize $\ln L(\mu,\sigma/\underset{\sim}{x}) = -n\ln 2 - n\ln\sigma - \frac{1}{\sigma}\Sigma|x_i-\hat{\mu}|$. w.r.t. σ

Note that $\frac{\delta}{\delta\sigma}\ln L(\hat{\mu},\sigma/\underset{\sim}{x})$

$$= -\frac{n}{\sigma} + \frac{1}{\sigma^2}\Sigma|x_i-\hat{\mu}|$$

$$= -\frac{n}{\sigma^2}\left\{\sigma - \frac{1}{n}\Sigma|x_i-\hat{\mu}|\right\}$$

$$\begin{cases} >0, & \text{if } \sigma < \frac{1}{n}\Sigma|x_i-\hat{\mu}| \\ <0, & \text{if } \sigma > \frac{1}{n}\Sigma|x_i-\hat{\mu}| \end{cases}$$

By 1st derivative test, $\ln L(\hat{\mu},\sigma/\underset{\sim}{x})$ is maximum at $\hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}|x_i-\hat{\mu}|$.

Hence, the MLE of μ and σ are $\hat{\mu} = \tilde{x}$, $\hat{\sigma} = \frac{1}{n}\Sigma|x_i-\tilde{x}|$.

Ex.(11):- Let $x_1, x_2, \ldots, x_n$ be an r.s. from

$$f(x;\mu,\sigma) = \begin{cases} \frac{1}{\sigma}e^{-(x-\mu)/\sigma} & ; \text{if } x>\mu \\ 0 & ; \text{ow} \end{cases}$$

where, $\mu \in \mathbb{R}$, $\sigma>0$. Find the MLE of (i) μ and σ
(ii) μ when $\sigma = \mu (>0)$.

[JAM 2005]

Solution:-

(i) The likelihood function is

$$L(\mu,\sigma/\underset{\sim}{x}) = \begin{cases} \frac{1}{\sigma^n}\cdot e^{-\frac{\Sigma(x_i-\mu)}{\sigma}} & ; \text{if } x_{(1)} \geq \mu \\ 0 & ; \text{ow} \end{cases}$$

$\mu \in \mathbb{R}$, $\sigma>0$.

We adopt two stage maximization.

First, fix σ, then maximize $L(\mu,\sigma/\underset{\sim}{x})$ w.r.t. μ.
For fixed σ, $L(\mu,\sigma/\underset{\sim}{x})$ is maximum
iff $\Sigma(x_i-\mu)$ is minimum subject to $\mu \leq x_{(1)}$
iff μ is as large as possible subject to the restriction $\mu \leq x_{(1)}$.
iff $\mu = x_{(1)} = \hat{\mu}$ (say)

Now we shall maximize $L(\hat{\mu},\sigma/\underset{\sim}{x})$ w.r.t. σ.

Now, $\ln L(\hat{\mu},\sigma/\underset{\sim}{x}) = -n\ln\sigma - \frac{\Sigma(x_i-\hat{\mu})}{\sigma}$.

Note that, $\frac{\delta}{\delta\sigma} \ln L(\hat{\mu}, \sigma / \underset{\sim}{x}) = -\frac{n}{\sigma} + \frac{1}{\sigma^2} \Sigma (x_i - \hat{\mu})$

$$= -\frac{n}{\sigma^2} \left\{ \sigma - (\bar{x} - x_{(1)}) \right\}$$

$$\begin{cases} >0 & \text{if } \sigma < \bar{x} - x_{(1)} \\ <0 & \text{if } \sigma > \bar{x} - x_{(1)} \end{cases}$$

Hence, $L(\hat{\mu}, \sigma / \underset{\sim}{x})$ is maximum at $\sigma = \bar{x} - x_{(1)} = \hat{\sigma}$.

Therefore, the MLEs of μ and σ are $\hat{\mu} = X_{(1)}$ and $\hat{\sigma} = \bar{X} - X_{(1)}$.

ii) When $\sigma = \mu > 0$

$$L(\mu / \underset{\sim}{x}) = \begin{cases} \frac{1}{\mu^n} e^{-\frac{\Sigma (x_i - \mu)}{\mu}} & ; x_{(1)} \geq \mu \\ 0 & ; \text{ow} \end{cases}$$

$L(\mu / \underset{\sim}{x})$ is maximum iff

for $\mu \leq x_{(1)}$

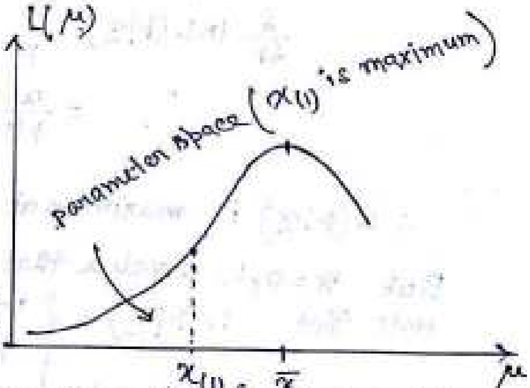
$$\frac{\delta}{\delta\mu} \ln L = \frac{\delta}{\delta\mu} \left\{ -n \ln \mu - \frac{1}{\mu} \Sigma (x_i - \mu) \right\}$$

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

$$\begin{cases} >0 & \text{if } \mu < \bar{x} \\ <0 & \text{if } \mu > \bar{x} \end{cases}$$

$\Rightarrow L(\mu / \underset{\sim}{x})$ is maximum at $\mu = \bar{x}$

From the graph for $\mu \leq x_{(1)}$, $L(\mu / \underset{\sim}{x})$ is maximum at $\mu = x_{(1)}$.



Therefore, $\hat{\mu} = x_{(1)}$ is the MLE of μ. [graph of $L(\mu / \underset{\sim}{x})$]

Ex(12): Let $X_1, \ldots, X_n$ be an r.s. from $U(\theta_1 - \theta_2, \theta_1 + \theta_2)$. Find the MLEs of θ_1 and θ_2.

Solution:–

Hints:–

$\theta_1 + \theta_2 = x_{(n)}$

$\theta_1 - \theta_2 = x_{(1)}$

$\Rightarrow \theta_1 = \frac{x_{(1)} + x_{(n)}}{2}$

$\theta_2 = \frac{x_{(n)} - x_{(1)}}{2}$.

Ex.(13):— (a) Let $X \sim Bin(1, p)$; $p \in [1/4, 3/4]$. Find the MLE of p. Explain the position of MLE w.r.t. the trivial estimation $\delta(X) = 1/2$, in terms of MSE.

(b) Let $X_1, \ldots, X_n$ be a r.s. from $B(1, p)$; $p \in [1/4, 3/4]$. Find the MLE of p.

Solution:— (a) $L(p|\underset{\sim}{x}) = p^x (1-p)^{1-x}$, if $x = 0, 1$.

$$\frac{\delta}{\delta p} \ln L(p|\underset{\sim}{x}) = \frac{x}{p} + \frac{1-x}{1-p}(-1)$$

$$= \frac{x-p}{p(1-p)} \begin{cases} >0 & \text{if } p < x \\ <0 & \text{if } p > x \end{cases}$$

$\therefore L(p|\underset{\sim}{x})$ is maximum at $p = x$.

But $x = 0, 1$, a value that does not lie in $\Omega = [\frac{1}{4}, \frac{3}{4}]$,

Note that $L(p|\underset{\sim}{x}) = \begin{cases} 1-p, & \text{if } x=0 \\ p, & \text{if } x=1 \end{cases}$

When $x=0$, $L(p|\underset{\sim}{x})$ is maximum,

iff $1-p$ is max, when $p \in [\frac{1}{4}, \frac{3}{4}]$

iff $p = 1/4$.

When $x=1$, $L(p|\underset{\sim}{x})$ is maximum,

iff p is max., $p \in [\frac{1}{4}, \frac{3}{4}]$

$\therefore$ MLE of p is $\hat{p} = \begin{cases} \frac{1}{4}, & \text{if } x=0 \\ \frac{3}{4}, & \text{if } x=1 \end{cases}$

Note that, $E(\hat{p}) \neq p$

and $MSE(\hat{p}) = E(\hat{p} - p)^2$

$$= \left(\frac{1}{4} - p\right)^2 \cdot P[X=0] + \left(\frac{3}{4} - p\right)^2 \cdot P[X=1]$$

$$= \left(\frac{1}{4} - p\right)^2 (1-p) + \left(\frac{3}{4} - p\right)^2 \cdot p$$

$$= 1/16.$$

Now, MSE of $\delta(X) = E[\delta(X) - p]^2$

$$= E\left(\tfrac{1}{2} - p\right)^2 \qquad \left[\because \tfrac{1}{4} \le p \le \tfrac{3}{4} \Rightarrow -\tfrac{1}{4} \le p - \tfrac{1}{2} \le \tfrac{1}{4}\right]$$

$$\le \frac{1}{16}$$

Hence, MSE $\{\delta(X)\} \le$ MSE $(\hat{p})$.

In terms of MSE, the MSE is worse than the trivial estimator.

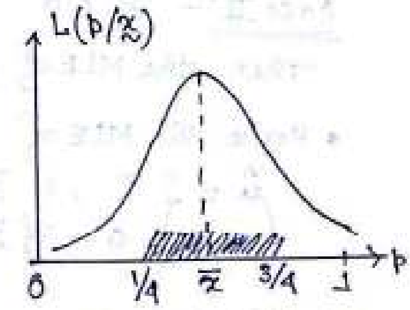
(b) The likelihood function:

$$L(p|\underset{\sim}{x}) = \begin{cases} p^{\Sigma x_i}(1-p)^{n-\Sigma x_i} & , \text{ if } x_i = 0,1,\ i = 1(1)n \\ 0 & , \text{ ow} \end{cases}$$

where, $p \in [\frac{1}{4}, \frac{3}{4}]$.

Note that, $\frac{\delta}{\delta p} \ln L(p|\underset{\sim}{x}) = \frac{n(\bar{x} - p)}{p(1-p)} \begin{cases} > 0 & \text{if } p < \bar{x} \\ < 0 & \text{if } p > \bar{x} \end{cases}$

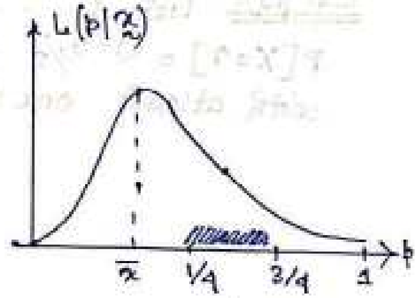
Hence, $L(p|\underset{\sim}{x})$ first increases, then achieves its maximum at $p = \bar{x}$ and finally decreases.



Case I :- Let, $\frac{1}{4} \le \bar{X} \le \frac{3}{4}$

For $p \in [\frac{1}{4}, \frac{3}{4}]$, $L(p|\underset{\sim}{x})$ is max. at $p = \bar{x}$

Hence, the MLE of p is $\hat{p} = \bar{X}$.



Case II :- Let, $\bar{X} < \frac{1}{4}$

Hence, the MLE of p is $\hat{p} = \frac{1}{4}$

Case III :- Let $\bar{x} > 3/4$

Then the MLE of p is $\hat{p} = \frac{3}{4}$

- Hence, the MLE of p is

$$\hat{p} = \begin{cases} \frac{1}{4} & \text{if } \bar{X} < \frac{1}{4} \\ \bar{X} & \text{if } \frac{1}{4} \le \bar{X} \le \frac{3}{4} \\ \frac{3}{4} & \text{if } \bar{X} > \frac{3}{4} \end{cases}$$

Properties of MLE:

We shall consider here some properties of MLE for samples of small size n and some asymptotic behavior of MLE for large n will be investigated. The importance of the method is clearly shown by the following properties:

(I) If a non-trivial sufficient statistic T of θ exists, any solution of the likelihood equation will be a function of T or the MLE, if exists, will be a function of T.

Proof:- For a non-trivial sufficient statistic T, we have $L(\underline{x};\theta) = g(T(\underline{x}),\theta)\cdot h(\underline{x})$; where, $h(\underline{x})$ is independent of θ, by factorization criterion.

Then, $\ln L(\underline{x};\theta) = \ln g(T(\underline{x}),\theta) + \ln h(\underline{x})$

Now, the likelihood equation is

$$0 = \frac{\partial}{\partial\theta}\ln L(\underline{x};\theta)$$

$$\Rightarrow 0 = \frac{\partial}{\partial\theta}\ln g(T(\underline{x}),\theta) + 0$$

and the function $g(T(\underline{x}),\theta)$ depends only on $T(\underline{x})$ and θ. Hence, any solution of the likelihood equation

$$0 = \frac{\partial}{\partial\theta}\ln L(\underline{x},\theta)$$

$$= \frac{\partial}{\partial\theta}\ln g(T(\underline{x}),\theta) \text{ will be a function of } T.$$

[Maximizing $\ln L(\underline{x};\theta)$ w.r.t. θ is equivalent to maximizing $\ln g(T(\underline{x}),\theta)$ w.r.t. θ. Here, $g(T(\underline{x}),\theta)$ depends only on θ and $T(\underline{x})$. The MLE of θ is the value of θ for which $\ln L(\underline{x};\theta)$ or $\ln g(T(\underline{x}),\theta)$ is maximum. Clearly, the MLE of θ will be a function of T.]

(II) Under the regularity condition in CR inequality, if MVBUE T of θ exists, then T is the MLE of θ.

Proof:- If MVBUE T of θ exists, then T attains CRLB.

$$\Leftrightarrow \frac{\partial \ln L(\underline{x};\theta)}{\partial\theta} = \Lambda(\theta)\{T-\theta\}$$

The likelihood equation is

$$\frac{\partial}{\partial\theta}\ln L(\underline{x};\theta) = 0$$

$$\Rightarrow \Lambda(\theta)\{T-\theta\} = 0$$

$\Rightarrow \theta = T$ is the unique solution.

Note that, $\frac{\partial^2}{\partial\theta^2}\ln L(\underline{x};\theta)$

$$= \Lambda(\theta)\cdot(-1) + (T-\theta)\Lambda'(\theta)$$

and $\frac{\partial^2}{\partial\theta^2}\ln L(\underline{x};\theta)\Big|_{\theta=T} = -\Lambda(T) < 0$

$$\left[\text{Now, } 0 < I(\theta) = E\left(-\frac{\partial^2}{\partial\theta^2}\ln L(\underset{\sim}{x};\theta)\right) \right.$$

$$= \Lambda(\theta) + \Lambda'(\theta)\{E(T)-\theta\}$$

$$\left. = \Lambda(\theta) \right]$$

Hence, $L(x;\theta)$ is maximum at $\theta = T$.

$\Rightarrow T$ is the MLE of θ.

(III) Bias of MLE :— MLE's are not in general unbiased and when MLE's are biased, then it is possible to modify them slightly so that they will be unbiased. e.g. The MLE of σ^2 in $N(\mu, \hat{\sigma}^2)$ popln., $\hat{\sigma}^2 = \frac{1}{n}\Sigma(x_i - \bar{x})^2$, which is biased but $E\left(\frac{n}{n-1}\hat{\sigma}^2\right) = \sigma^2$, i.e. $\frac{n}{n-1}\cdot\hat{\sigma}^2$ is unbiased.

(IV) Invariance of MLE :— If $\hat{\theta}$ is the MLE of θ, the $h(\hat{\theta})$ is the MLE of $h(\theta)$; provided $h(\theta)$ is a function of θ.

Proof :— If $h(\theta) = \lambda$ is a one-to-one function of θ, the inverse function $h^{-1}(\lambda) = \theta$ is well defined and we can write the likelihood function as a function of λ. We have

$$L^*(\lambda; \underset{\sim}{x}) = L(h^{-1}(\lambda); \underset{\sim}{x})$$

so that $\sup\limits_{\lambda} L^*(\lambda;\underset{\sim}{x}) = \sup\limits_{\lambda} L(h^{-1}(\lambda);\underset{\sim}{x}) = \sup\limits_{\theta} L(\theta;\underset{\sim}{x})$

It is followed that the supremum of L^* is achieved at $\lambda = h(\hat{\theta})$. Thus $h(\hat{\theta})$ is the MLE of $h(\theta)$.

In many applications, $\lambda = h(\theta)$ is not one-to-one, It is still tempting to take $\hat{\lambda} = h(\hat{\theta})$ as the MLE of λ.

e.g. (i) Let $X \sim b(1,p)$; $0 \le p \le 1$, let $h(p) = var(X) = p(1-p)$. We wish to find the MLE of $h(p)$. Note that $\Lambda = [0, \frac{1}{4}]$. The function h is not one-to-one. The MLE of p based on a sample of size n is $\hat{p}(x_1, \ldots, x_n) = \bar{X}$. Hence, the MLE of parameter $h(p)$ is $h(\bar{X}) = \bar{X}(1-\bar{X})$.

(ii) The MLE of σ^2 based on a r.s. from $x_1, \ldots, x_n$ from $N(\mu, \sigma^2)$ is $\hat{\sigma}^2 = \frac{1}{n}\Sigma(x_i - \bar{x})^2 = s^2$, then by invariance property, the MLE of $\mu_4 = 3(\sigma^2)^2$ is $\hat{\mu}_4 = 3(\hat{\sigma}^2)^2 = 3(s^2)^2$.

(v) Asymptotic Properties of MLE:—

(a) Under certain regularity conditions, the likelihood equation has a solution which is consistent for θ. Then the solution $\hat{\theta}$ is asymptotically normal and

$$\sqrt{n}(\hat{\theta}-\theta) \overset{a}{\sim} N\left(0, \frac{1}{I_{X_1}(\theta)}\right)$$

$$\Leftrightarrow \hat{\theta} \overset{a}{\sim} N\left(\theta, \frac{1}{I_n(\theta)}\right)$$

where, $I_n(\theta) = n I_{X_1}(\theta)$

$$= n.E\left(\frac{\partial}{\partial\theta} \ln f(x_1;\theta)\right)^2$$

i.e. $\hat{\theta}$ is the Based Asymptotical Normal (BAN) estimator.

In particular, for OPEF, the MLE $\hat{\theta}$ is consistent for θ and $\sqrt{n}(\hat{\theta}-\theta) \overset{a}{\sim} N\left(0, \frac{1}{I_{X_1}(\theta)}\right)$.

(b) Asymptotic Invariance:—

In OPEF, if $\hat{\theta}$ is the MLE of θ, then

$$\sqrt{n}(\hat{\theta}-\theta) \overset{a}{\sim} N\left(0, \frac{1}{I_{X_1}(\theta)}\right)$$

implies $\sqrt{n}\{\psi(\hat{\theta}) - \psi(\theta)\} \overset{a}{\sim} N\left(0, \frac{\{\psi'(\theta)\}^2}{I_{X_1}(\theta)}\right)$

Ex. (1):— Let $X_1, \ldots, X_n$ be a r.s. from $B(1,p)$, $p \in (0,1)$. Find the MLE of (i) $\psi(p) = e^{-p}$, (ii) $\psi(p) = Var(X_1)$.

Solution:— The MLE of $p \in (0,1)$ is $\hat{p} = \bar{X}$, provided $\bar{X} \neq 0$ or 1.

(i) Note that $\psi(p) = e^{-p}$ is a function from $\Omega = (0,1)$ onto $\Lambda = (e^{-1}, 1)$.

By invariance property, $\psi(\hat{p}) = e^{-\bar{X}}$ is the MLE of $\psi(p) = e^{-p}$.

(ii) $\psi(p) = Var(X_1) = p(1-p)$ is a function from $\Omega = (0,1)$ onto $\Lambda = (0, \frac{1}{4})$.

By invariance property, $\psi(\hat{p}) = \hat{p}(1-\hat{p}) = \bar{X}(1-\bar{X})$ is the MLE of $\psi(p) = p(1-p)$.

*Ex. (2):— Let $X_1, \ldots, X_n$ be a r.s. from $P(\lambda)$. Find the MLE of (i) $\psi(\lambda) = e^{-\lambda}$, (ii) $\psi(\lambda) = P[X \geq 2]$.

Also find the SE of $\psi(\lambda) = e^{-\lambda}$ and its MLE.

Solution:— The MLE of λ is $\hat{\lambda} = \bar{X}$, provided $\bar{X} > 0$.

(i) Note that $\psi(\lambda) = e^{-\lambda}$ is a function from $\Omega = \{\lambda : \lambda > 0\}$ onto $\Lambda = (0,1)$.

By invariance property, the MLE of $\psi(p) = e^{-\lambda}$ is $\psi(\hat{\lambda}) = e^{-\hat{\lambda}} = e^{-\bar{X}}$

(ii) $\psi(\lambda) = 1 - P[X=0] - P[X=1]$

$$= 1 - e^{-\lambda}(1+\lambda)$$

$\therefore \psi(\hat{\lambda}) = 1 - e^{-\hat{\lambda}}(1+\hat{\lambda})$ is the MLE of $\psi(\lambda) = 1 - e^{-\lambda}(1+\lambda)$.

Using asymptotic property,

$$\sqrt{n}\left\{\psi(\hat{\lambda})-\psi(\lambda)\right\} \overset{a}{\sim} N\left(0, \frac{\{\psi'(\lambda)\}^2}{I_{X_1}(\lambda)}\right)$$

$$\Leftrightarrow \psi(\hat{\lambda}) \overset{a}{\sim} N\left(\psi(\lambda), \frac{\{\psi'(\lambda)\}^2}{nI_{X_1}(\lambda)}\right)$$

Here, $\psi(\lambda)=e^{-\lambda}$ and $nI_{X_1}(\lambda)=\frac{n}{\lambda}$ $\left[V(\bar{X})=\frac{1}{nI_{X_1}(\lambda)}\right]$

$\therefore e^{-\hat{\lambda}} \overset{a}{\sim} N\left(e^{-\lambda}, \frac{\lambda e^{-2\lambda}}{n}\right)$ is the asymptotic distn. of the MLE of $e^{-\hat{\lambda}}$.

For large n,

$$V(e^{-\hat{\lambda}}) \simeq \frac{\lambda e^{-2\lambda}}{n}$$

$$\Rightarrow SE\left(e^{-\hat{\lambda}}\right) \simeq e^{-\lambda}\cdot\sqrt{\frac{\lambda}{n}}$$

By invariance property, MLE of S.E. $(e^{-\hat{\lambda}})$ is

$$\hat{SE}\left(e^{-\hat{\lambda}}\right) = e^{-\hat{\lambda}}\sqrt{\frac{\hat{\lambda}}{n}} = e^{-\bar{X}}\sqrt{\frac{\bar{X}}{n}}, \text{ for large } n.$$

*__Ex.(3)__:- Let $X_1, X_2, \ldots, X_n$ be a r.s. from

$$f(x;\theta)=\begin{cases}\theta e^{-\theta x} & , x>0\\ 0 & , \text{ow}\end{cases}$$

where, $\theta>0$.

Find the MLE of θ. S.T. the MLE is biased but consistent. State its asymptotic distribution. Also, find the MLE of $S(t)=P[X>t]$ and its asymptotic distn. Also find the SE of $S(t)$ & its MLE.

Ex.(4):- Let $X_1, \ldots, X_n$ be an r.s. from $U(0,\theta)$. Find the asymptotic distribution of MLE of θ and comment.

Solution:- The MLE of θ is $\hat{\theta} = X_{(n)}$. (prove it)

Define, $Y_n = n(\theta - X_{(n)})$

The D.F. of Y_n is

$$G_n(y) = P[Y_n \le y] = P\left[X_{(n)} \ge \theta - \frac{y}{n}\right]$$

$$= F_{X_{(n)}}\left(\theta - \frac{y}{n}\right)$$

$$= \begin{cases} 1 - 0, & \text{if } \theta - \frac{y}{n} \le 0 \\ 1 - \left(\frac{\theta - \frac{y}{n}}{\theta}\right)^n & \text{if } 0 < \theta - \frac{y}{n} < \theta \\ 1 - 1 & \text{if } \theta - \frac{y}{n} \ge \theta \end{cases}$$

$$= \begin{cases} 0, & \text{if } y \le 0 \\ 1 - \left(1 + \frac{-\frac{y}{\theta}}{n}\right)^n, & \text{if } 0 < y < n\theta \\ 1, & \text{if } y \ge n\theta \end{cases}$$

$$\longrightarrow \begin{cases} 0 & \text{if } y \le 0 \\ 1 - e^{-y/\theta} & \text{if } 0 < y < \infty \end{cases}$$

which is the DF of the Exp. distn. with mean θ.

Hence, $Y_n = n(\theta - X_{(n)}) \xrightarrow{L} Y \sim$ Exponential distribution (θ).

Therefore, the MLE $\hat{\theta} = X_{(n)}$ is not an asymptotic normal.

Note that $U(0,\theta)$ distn. does not satisfy the regularity conditions required for CR inequality and the CRLB does not exist. consequently, the asymptotic property of MLE $\hat{\theta} \overset{a}{\sim} N\left(\theta, \frac{1}{I_n(\theta)}\right)$ does not hold.

*Ex.(5):- Find the MLE of $g(\theta)=2\theta+1$ based on a r.s. $X_1,\ldots,X_n$ from $f(x;\theta)=\frac{1}{2}e^{-|x-\theta|}$; $x\in\mathbb{R}$, where $\theta\in\mathbb{R}$. Find a consistent estimator of θ and $g(\theta)$.

Solution:- The MLE of θ is $\hat{\theta}=\tilde{x}$ = the sample median (prove it).
By invariance property, the MLE of $g(\theta)=2\theta+1$ is $g(\hat{\theta})=2\tilde{x}+1$
We have $\hat{\xi}_p \overset{a}{\sim} N\left(\xi_p, \frac{p(1-p)}{nf^2(\xi_p)}\right)$

$$\Rightarrow \hat{\xi}_{1/2} \overset{a}{\sim} N\left(\xi_{1/2}, \frac{1}{4nf^2(\xi_{1/2})}\right)$$

Here, $\tilde{X} \overset{a}{\sim} N\left(\theta, \frac{1}{4n\left(\frac{1}{2}\right)^2}\right)$

$$\Rightarrow \tilde{X} \overset{a}{\sim} N\left(\theta, \frac{1}{n}\right)$$

For large n, $E(\tilde{X})\simeq\theta$ and $Var(\tilde{X})\simeq\frac{1}{n}\to 0$ as $n\to\infty$.
Hence, $\tilde{X}$ is consistent for θ and $g(\tilde{X})$ is consistent for $g(\theta)$, by invariance property.

Ex.(6):- Let $X_1,\ldots,X_n$ be a r.s. from $N(\theta,\theta)$, $\theta>0$. Find the MLE of θ. Is it unique? Also, suggest a sufficient statistic for θ.

Solution:- The likelihood function :-

$$L(\theta|\underset{\sim}{x})=\left(\frac{1}{\sqrt{2\pi\theta}}\right)^n\cdot e^{-\frac{1}{2\theta}\Sigma(x_i-\theta)^2}\text{; where }\theta>0.$$

$$\therefore \ln L(\theta|\underset{\sim}{x}) = \text{constant} - \frac{n}{2}\ln\theta - \frac{\Sigma x_i^2-2\theta\Sigma x_i+n\theta^2}{2\theta}$$

Likelihood Equation: -

$$0=\frac{\partial}{\partial\theta}\ln L = -\frac{n}{2\theta}+\frac{1}{2\theta^2}\Sigma x_i^2-\frac{n}{2}$$

$$= -\frac{n}{2\theta^2}\left\{\theta^2+\theta-\frac{1}{n}\Sigma x_i^2\right\}$$

$$\Rightarrow \theta^2+\theta-\frac{1}{n}\Sigma x_i^2=0$$

$$\Rightarrow \theta=\frac{-1\pm\sqrt{1+\frac{4}{n}\Sigma X_i^2}}{2}=\alpha,\beta$$

$$\Rightarrow \theta=\beta=\frac{-1+\sqrt{1+\frac{4}{n}\Sigma X_i^2}}{2}\text{ ; neglecting negative sign as }\theta>0.$$

Note that, $\frac{\partial\ln L}{\partial\theta}=-\frac{n}{2\theta^2}(\theta-\alpha)(\theta-\beta)$

$$=\begin{cases} >0, & \theta<\beta \\ <0, & \theta>\beta \end{cases}$$

$\Rightarrow L(\theta|\underset{\sim}{x})$ is maximum at $\theta=\beta$.

$$\Rightarrow \hat{\theta}=\frac{-1+\sqrt{1+\frac{4}{n}\Sigma X_i^2}}{2}\text{ is the unique MLE of }\theta.$$

As MLE is a function of a sufficient statistic. hence $T=\sum_{i=1}^{n}X_i^2$ is sufficient for θ.

Ex.(7):- Let X denotes the no. of white balls in a sample of n balls drawn without replacement (WOR) from an urn containing N white and M − N black balls where M is unknown and N is known. Find the MLE of M.

Solution:- The Likelihood function is; —

$$p(M|x) = \begin{cases} \dfrac{\binom{N}{x}\binom{M-N}{n-x}}{\binom{M}{n}} & ; \text{if } x = 0(1)n. \\ 0 & ; \text{ow} \end{cases}$$

Note that, $\dfrac{p(M|x)}{p(M-1|x)} = \dfrac{M-n}{M} \cdot \dfrac{M-n}{M-N-n+x} \gtrless 1$

according as $M \lessgtr \dfrac{nN}{x}$.

It follows that $p(M/x)$ reaches its maximum at $M \simeq \dfrac{nN}{x}$, i.e. at $M = \left[\dfrac{nN}{x}\right]$.

Hence, $\hat{M} = \left[\dfrac{nN}{x}\right]$ is the MLE of M.

Ex.(7):- One observation is taken on a discrete r.v. with RVX with PMF $f(x;\theta)$; where $\theta \in [1,2,3]$. Find the MLE of θ.

x	0	1	2	3	4
$f(x;1)$	1/3	1/3	0	1/6	1/6
$f(x;2)$	1/4	1/4	1/4	1/4	0
$f(x;3)$	0	0	1/4	1/2	1/4

Solution:- For each value of x, the MLE $(\hat{\theta})$ in the value of θ that maximizes $f(x;\theta)$. These values are given in the following table:

x	0	1	2	3	4
$\hat{\theta}$	1	1	2 or 3	3	3

When $x=2$ is observed, $f(x;2) = f(x;3)$ are both maxima, so both $\hat{\theta}=2$ or $\hat{\theta}=3$ are MLEs of θ.

Of these two methods suggested, the M.L method is superior because of the following desirable properties

1. M.L estimates are asymptotically unbiased
2. M.L estimates are consistent
3. M.L estimates are most efficient
4. M.L estimates are sufficient if sufficient estimates exist.
5. M.L estimates are asymptotically normally distributed

Since the point estimate is a statistic it not unique always. So, a natural question arises which estimate is the best?. The answer of this is question is that estimate which satisfies the desirable properties of an estimate is a good estimator. According to R.A. Fisher, most important desirable properties of a good estimate are

1. Unbiasedness
2. Consistency
3. Efficiency
4. Sufficiency

Now we look this one by one

Unbiasedness

An unbiased estimator of a population parameter is an estimator whose expected value is equal to that parameter. Let ‘t’ be a statistic suggested as an estimate of the parameter θ. ‘t ‘ is said to be unbiased estimate of θ if E(t)= θ.

If E(t) ≠ θ, then the estimator is called a biased estimator and the bias of the estimate is given by E(t) – θ.

Example-1

Let X_i is a binomial random variable with parameter *p*, then $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an unbiased estimate of *p*. Since X_i is a binomial random variable with petameter *p*, we have E(X_i) =*p*. So

$$E(\hat{p}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} p = \frac{1}{n}(np) = p.$$

Hence $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an unbiased estimate of *p*.

Let X_i is a normal random variable with mean μ and variance σ^2. Check whether the estimate

$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ are unbiased or not

Recall that if X_i is a normally distributed random variable with mean μ and variance σ^2, then $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, therefore:

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}(n\mu) = \mu$$

In the case of σ^2 we have

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X})^2$$

So

$$E(\hat{\sigma}^2) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X})^2\right] = \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - E(\bar{X})^2$$

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

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

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

Since

$$Var(X_i) = \sigma^2 = E(X^2) - \mu^2 \text{ and } Var(\bar{X}) = \frac{\sigma^2}{n} = E(\bar{X}^2) - \mu^2.$$

So $\hat{\sigma}^2$ is not an unbaised estimate of σ^2 as $E(\hat{\sigma}^2) = \frac{(n-1)}{n}\sigma^2 \neq \sigma^2$.

X_i is a normal random variable with mean μ and variance σ^2, what is an unbiased estimator of σ^2?

Let $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{n}{n-1}\hat{\sigma}^2$

So $E(S^2) = \frac{n}{n-1}E(\hat{\sigma}^2) = \frac{n}{(n-1)}\frac{(n-1)}{n}\sigma^2 = \sigma^2$

Hence $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an unbiased estimate of σ^2.

Consistency

Let ' t_n' be a statistic where n is the sample size. t_n is said to be a consistent estimate of parameter θ , if for any two positive numbers ' ϵ' and ' η' (however small they may be) a N can be found out such that n≥N.

$$P(|t_n - \theta| < \in) > 1 - \eta$$

This idea is sometimes expressed by saying that t_n is a consistent estimate of θ if it tends 'in probability' to θ as $n \to \infty$. This means that if t_n is a consistent estimate of θ, the difference between t_n and θ can be made however small we please in almost all samples of size n if the sample size is sufficiently large. The definition may also be stated as t_n is a consistent estimate of θ if,

$$P(|t_n - \theta| < \in) \to 1 \text{ as } n \to \infty$$

It is to be noted that consistency is a large sample property while unbiasedness is defined for all sample sizes. A consistent estimate need not ne unbiased for small samples, but it becomes unbiased as $n \to \infty$. The following is a sufficient set of condition for the consistency of an estimate which will be found very useful in examining whether an estimate is consistent or not.

Result-1: If $E(t_n) = \theta$ or $E(t_n) \to \theta$ and $V(t_n) \to \theta$ as $n \to \infty$, t_n is a consistent estimate of θ. This may be proved as follows.

By Techebycheff's inequality

$$P(|t_n - \theta| < k\sqrt{V(t_n)}) > 1 - \frac{1}{k^2}$$

Let $\in > 0$ and $\eta > 0$ be two given numbers. Since $V(t_n) \to \theta$ we can find an N such that for $n \geq N$, and

$$k\sqrt{V(t_n)} < \in \quad \text{and} \quad \frac{V(t_n)}{t^2} < \eta$$

$\therefore \quad P(|t_n - \theta| < \in) > 1 - V(t_n) > 1 - \eta^2$

$\therefore$ t_n is a consistent estimate of θ

When $E(t_n) \to \theta$, $|t_n - E(t_n)| \to |t_n \to \theta|$ and the result follows

Result-2

If t_n is a biased estimate of the parameter θ based on a sample of size n and

$E(t_n) = \theta + b_n$ and if $b_n \to 0$ and $V(t_n) \to 0$ as $n \to \infty$, Show that t_n is a consistent estimator of θ.

By Techebycheff's inequality

$$P(|t_n - E(t_n) \geq t\sqrt{V(t_n)}|) \leq \frac{1}{t^2}$$

Ie, $\quad P(|t_n - \theta - b_n| > \in) \leq \frac{V(t_n)}{\epsilon^2}$

When $n \to \infty$, $b_n \to 0$ and $V(t_n) \to 0$. So ineqyality becomes,

$P(|t_n - \theta| \geq \in|) \leq 0$ when $n \to \infty$

$\therefore$ t_n is a consistent estimator of θ

Example:

If T is an unbiased estimate of the parameter θ. T^2 is a biased estimator of θ^2 but if T is a consistent estimator of θ the T^2 is also a consistent estimator of θ^2

E(T) = θ, V(t) = E(T- θ)^2 = E(T)^2 -2 θE(T) + θ^2

= E(T)^2 - $\theta^2 \geq 0$ (Var is always positive)

$\therefore$ E(T^2) $\geq \theta^2$ and hence T^2 is not an unbiased estimator of θ^2. If T is a consistent estimator of θ we can find an n_0 such that for all values of $n \geq n_0$, P{|T – θ|≤ε|≥1- η where ε >0 and η>0 are any twi constants however small they may be. Since T^2 is a continuous function of T we can find an ε such that, |T^2 - θ^2|≤ e when |T – θ| ≤ $\varepsilon_1$and for this ε_1>0 and η >0 we can find an n_0 such that for $n \geq n_0$, P{|T – θ|≤ε_1| and hence |T^2 - θ^2|≤ε|≥1- η. So T^2 is a consistent estimator of θ^2

The second part of this question can also be proved as follows:

Since T is a consistent estimator of θ, for $n \geq n_0$

P{|T- θ|≤∈ }≥1- η ie, P{-∈ ≤ T- θ ≤+ ∈} ≥1- η

ie, P{θ-∈ ≤ T ≤ θ + ∈} ≥1- η

ie, $P\{(\theta - \in)^2 \leq T^2 \leq (\theta + \in)^2\}$ ≥1- η ($\theta - \in \geq 0$)

ie, $P\{\in^2 - 2\theta \in \leq T^2 - \theta^2 \leq \in^2 + 2\theta \in\}$ ≥1- η

we know that if a<b, then a-c<b for any c>0, So by substituting $2\in^2$ for $\in^2 - 2\theta \in$ the inequality in not affected. Hence

$P\{-\in^2 \leq T^2 - \theta^2 \leq \in^2 + 2\theta \in\}$ ≥1- η

If we put $\in^2 + 2\theta \in = \in_1$

$P\{-\in_1 \leq T^2 - \theta^2 \leq \in_1\}$ ≥1- η

Or

$P\{|T^2 - \theta^2| \leq \in_1\}$ ≥1- η

So T^2 is a consistent estimate of θ^2.

Ex. (3):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $U(0, \theta)$, $\theta > 0$. Which of the following estimators are consistent for θ?

(i) $T_1 = \max_i \{x_i\}$, (ii) $T_2 = \frac{n+1}{n} T_1$, (iii) $T_3 = 2\bar{X}$.

ANS:- (i) $F_{T_1}(t_1) = \begin{cases} 0 & , t_1 \le 0 \\ \left(\frac{t_1}{\theta}\right)^n & , 0 < t_1 < \theta \\ 1 & , t_1 \geq \theta \end{cases}$

Now, $P\left[|T_1 - \theta| < \epsilon\right] = P\left[\theta - \epsilon < T_1 < \theta + \epsilon\right]$

$= F_{T_1}(\theta+\epsilon) - F_{T_1}(\theta - \epsilon)$

$= \begin{cases} 1 - \left(\frac{\theta - \epsilon}{\theta}\right)^n & ; \text{if } 0 < \epsilon < \theta \\ 1 & ; \text{if } \epsilon \geq \theta \end{cases}$

$\longrightarrow 1$ as $n \to \infty$, for every $\epsilon > 0$.

Hence T_1 is consistent for θ.

(ii) $T_2 = \frac{n+1}{n} T_1$

$= b_n T_1$, where $b_n = \frac{n+1}{n} \longrightarrow 1$ as $n \to \infty$

Clearly, T_n is consistent for θ, since for every $\epsilon > 0$,

$P\left[|T_2 - \theta| < \epsilon\right]$

$= P\left[\left|\frac{n+1}{n} T_1 - \theta\right| < \epsilon\right]$

$\simeq P\left[|T_1 - \theta| < \epsilon\right]$, for large n.

$\longrightarrow 1$ as $n \to \infty$.

(iii) Note that, $E(\bar{X}) = E(X_1) = \frac{\theta}{2}$

& $V(\bar{X}) = \frac{V(X_1)}{n} = \frac{\theta^2}{12n}$

For every $\epsilon > 0$, $P\left[|T_3 - \theta| > \epsilon\right]$

$= P\left[|2\bar{X} - \theta| > \epsilon\right]$

$< \frac{V(2\bar{X})}{\epsilon^2} = \frac{4V(\bar{X})}{\epsilon^2} = \frac{4 \times \theta^2}{12n\epsilon^2}$

$\longrightarrow 0$ as $n \to \infty$

So, T_3 is consistent for θ.

A sufficient condition for consistency :—

The direct verification of consistency from the definition may not always be an easy task. The following theorem helps in determining the consistency of $\{T_n\}$ for θ.

Theorem:— If $\{T_n\}$ is a sequence of estimators such that
$$E(T_n) \to \theta \quad \text{and} \quad V(T_n) \longrightarrow 0 \quad \text{as } n \to \infty.$$
Then $\{T_n\}$ is consistent for θ.

Proof:— For $\epsilon > 0$,
$$0 \le P\left[|T_n - \theta| > \epsilon\right] < \frac{E(T_n - \theta)^2}{\epsilon^2}$$
$$= \frac{V(T_n) + \{E(T_n) - \theta\}^2}{\epsilon^2}$$
$$\longrightarrow 0 \text{ as } n \to \infty,$$
provided $E(T_n) \to \theta$ and $V(T_n) \to 0$ as $n \to \infty$.

[Markov's inequality : $P\left[|X| > \epsilon\right] < \frac{E|X|^n}{\epsilon^n}$, $\epsilon > 0, n > 0$]

Remark:— The above theorem can also be stated as follows :
' If $\{T_n\}$ is a sequence of estimators such that $E(T_n - \theta)^2 \longrightarrow 0$ as $n \to \infty$, then $\{T_n\}$ is consistent for θ.'

Ex. (4). Let $X_1, X_2, \dots, X_n$ be a r.s. from a popln with mean μ and variance σ^2. Which of the following estimators are consistent for μ?

(i) $T_1 = \frac{2}{n(n+1)} \sum_{i=1}^{n} i.X_i$, (ii) $T_2 = \frac{X_1 + X_2 + \dots + X_n}{\frac{n}{2}}$

(iii) $T_3 = \frac{6 \sum_{i=1}^{n} i^2.X_i}{n(n+1)(2n+1)}$

Soln.:—

(i) $E(T_1) = E\left\{\frac{2\sum_{i=1}^{n} i.X_i}{n(n+1)}\right\}$
$$= \frac{2}{n(n+1)} \sum_{i=1}^{n} iE(X_i).$$
$$= \frac{2}{n(n+1)} \left(\sum_{i=1}^{n} i\right)\mu$$
$$= \mu$$

$\text{Var}(T_1) = \text{Var}\left\{\frac{2}{n(n+1)} \sum_{i=1}^{n} i.X_i\right\}$
$$= \frac{4}{\{n(n+1)\}^2} \sum_{i=1}^{n} i^2.\sigma^2$$
$$= \frac{4\sigma^2 n(n+1)(2n+1)}{6n^2(n+1)^2}$$
$$= \frac{2\sigma^2(2n+1)}{3n(n+1)}$$
$$\longrightarrow 0 \text{ as } n \to \infty$$

Hence, T_1 is consistent for μ.

(ii) $E(T_2) = \dfrac{n\mu}{n/2} = 2\mu$

$\Rightarrow E(T_2) \not\to \mu$

but $E\left(\dfrac{T_2}{2}\right) = \mu$

$\therefore$ T_2 is not consistent for μ.

(iii) $E(T_3) = E\left\{\dfrac{6\sum_{i=1}^{n} i^2 . x_i}{n(n+1)(2n+1)}\right\} = \dfrac{6\mu}{n(n+1)(2n+1)}\sum_{i=1}^{n} i^2$

$= \mu$

$\text{var}(T_3) = \dfrac{6\sigma^2}{n(n+1)(2n+1)}\sum_{i=1}^{n} i^4$

$= \dfrac{36 . \cancel{n^5}^{\,n^3} . \sigma^2}{5\,\cancel{n^2}(n+1)^2(2n+1)^2}$

$= \dfrac{36n^3\sigma^2}{5(n+1)^2(2n+1)^2}$

$\longrightarrow 0$ as $n \to \infty$

$\left[\ \dfrac{1}{n}\sum_{i=1}^{n}\left(\dfrac{i}{n}\right)^4 \simeq \int_0^1 x^4 dx = \dfrac{1}{5},\right.$

$\Rightarrow \sum_{i=1}^{n} i^4 = \dfrac{n^5}{5}$

OR, $\left.\sum_{i=1}^{n} i^4 \simeq \int_0^n x^4 dx = \dfrac{n^5}{5}\ \right]$

$\therefore$ T_3 is consistent for μ.

Ex.(5). Let $X_1, X_2, \ldots, X_n$ be a r.s. from $U(\theta, \theta+1)$. S.T. (i) $T_1 = \bar{X} - \frac{1}{2}$, (ii) $T_2 = X_{(n)} - \frac{n}{n+1}$ are both consistent for θ.

ANS:-

$E(\bar{X}) = E(X_1) = \theta + \frac{1}{2}$

$\Rightarrow E(T_1) = \theta$,

$V(\bar{X}) = \dfrac{\sigma^2}{n} = \dfrac{1}{12n}$

$\Rightarrow V(T_1) = \dfrac{1}{12n} \longrightarrow 0$ as $n \to \infty$

$\therefore$ T_1 is consistent for θ.

Ex.(6). Let $X_1, X_2, \ldots, X_n$ be a r.s. from $U(0, \theta)$. S.T. $G = \left(\prod_{i=1}^{n} X_i\right)^{1/n}$ is consistent for θ/e.

ANS:-

$$E(G) = E\left(\prod_{i=1}^{n} X_i\right)^{1/n}$$

$$= E\left\{\prod_{i=1}^{n} (x_i)^{1/n}\right\}$$

$$= \prod_{i=1}^{n} E(X_i^{1/n})$$

$$= \prod_{i=1}^{n} \left\{\int_0^\theta x_i^{1/n} \cdot \frac{1}{\theta}\, dx_i\right\}$$

$$= \prod_{i=1}^{n} \left[\frac{x_i^{1/n+1}}{1/n+1}\right]_0^\theta \cdot \frac{1}{\theta}$$

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

$$= \frac{\theta}{\left(1+\frac{1}{n}\right)^n} \quad [\because X_i\text{'s are i.i.d. RV's}]$$

$$\longrightarrow \frac{\theta}{e} \text{ as } n \to \infty.$$

$$V(G) = E(G^2) - E^2(G)$$

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

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

$$\longrightarrow \frac{\theta^2}{e^2} - \frac{\theta^2}{e^2} = 0 \text{ as } n \to \infty.$$

Hence, G is consistent for $\frac{\theta}{e}$.

Ex.(7). Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(0, \sigma^2)$, S.T. some multiple of $\sum_{i=1}^{n} |X_i|$ is consistent for σ.

ANS:-

$$E\left(\sum_{i=1}^{n} |X_i|\right) = \sum_{i=1}^{n} E|X_i| = n \cdot \sigma \cdot \sqrt{\frac{2}{\pi}}$$

$$\Rightarrow E\left(\frac{1}{n} \cdot \sqrt{\frac{\pi}{2}} \sum_{i=1}^{n} |X_i|\right) = \sigma$$

$$\Rightarrow E(T_1) = \sigma, \text{ where } T_1 = \frac{1}{n}\sqrt{\frac{\pi}{2}} \sum_{i=1}^{n} |X_i|$$

$$\text{Var}(T_1) = \frac{\pi}{2n^2} \sum_{i=1}^{n} \left\{ E(x_i^2) - n^2 \cdot \sigma^2 \cdot \frac{2}{\pi} \right\}$$

$$= \frac{\pi}{2n^2} \sum_{i=1}^{n} \left\{ \sigma^2 - n^2 \cdot \sigma^2 \cdot \frac{2}{\pi} \right\}$$

$$= \frac{\pi}{2n} \sigma^2 \left(1 - \frac{2n^2}{\pi}\right) \longrightarrow 0 \text{ as } n \to \infty$$

Hence $T_1 = \frac{1}{n}\sqrt{\frac{\pi}{2}} \sum_{i=1}^{n} |x_i|$ is consistent for σ.

<u>Remark:-</u> We have the theorem:

"If $\{T_n\}$ is a sequence of estimators such that $E(T_n - \theta)^2 \to 0$ as $n \to \infty$, then $\{T_n\}$ is consistent for θ."

"The converse of the theorem is not necessarily true", i.e. we have situations where $T_n \xrightarrow{P} \theta$ but $E(T_n - \theta)^2 \not\to 0$ as $n \to \infty$.

<u>For example:-</u>

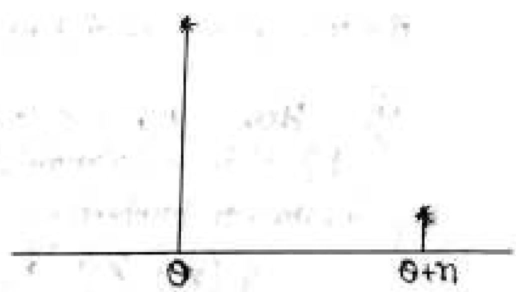
$$T_n = \begin{cases} \theta & \text{with probability } \left(1 - \frac{1}{n}\right) \\ \theta + n & \text{with probability } \frac{1}{n} \end{cases}$$

Now, $P\left[|T_n - \theta| > \epsilon\right]$

$= P\left[T_n = \theta + n\right]$

$= \frac{1}{n} \longrightarrow 0$ as $n \to \infty$

$\Rightarrow T_n \xrightarrow{P} \theta$



But, $E(T_n - \theta)^2 = (\theta - \theta)^2 \cdot \left(1 - \frac{1}{n}\right) + (\theta + n - \theta)^2 \cdot \frac{1}{n}$

$$= \frac{n^2}{n} = n \not\longrightarrow 0 \text{ as } n \to \infty$$

Hence, $T_n \xrightarrow{P} \theta$ but $E(T_n - \theta)^2 \not\to 0$ as $n \to \infty$.

Sufficiency

An estimate of a parameter θ is called a sufficient estimate if it contains all the information about θ contained in the sample.

Let t' be any other estimate of , let P(t/t') be the joint p.d.f of t and t'. Let P(t/t') be the conditional p.d.f of t' and t. If this is independent of let θ for all t; we say that t is a sufficient estimate.

A necessary and sufficient condition for an estimate of a parameter θ to be sufficient is given by Neyman. This will be found very useful in examining whether a given estimate is sufficient for θ or not.

Neyman's condition for sufficiency

Let $x_1, x_2, \ldots x_n$ be a sample from a population with p.d.f f(x, θ). The joint p.d.f of the sample (usually called the likelihood of the sample) is,

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

t is a sufficient estimate of θ if and only if it is possible to write

$$L(x_1, x_2, \ldots x_n, \theta) = L_1(t,\ \theta)\ L_2(x_1, x_2, \ldots x_n),$$

Where $L_1(t, , \theta)$ is a function of t and θ alone and L_2 is independent of θ.

Show that if σ^2 is known $\bar{x}$ is a sufficient estimate of μ is known σ^2 is not a sufficient estimate of s^2 in the case of samples from N(μ, σ), $\bar{x}$ and s^2 being the mean and variance of a sample.

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

$$L(x_1, x_2, \ldots x_n,) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \quad \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_2-\mu)^2}{2\sigma^2}} \quad \ldots\ldots.. \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}$$

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

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

$$[\textstyle\sum_{i=1}^{n}(x_1-\mu)^2 = \sum_{i=1}^{n}(x_1-\bar{x}+\bar{x}-\mu)^2 = \sum_{i=1}^{n}(x_1-\bar{x})^2 + n(\bar{x}-\mu)^2]$$

Where σ^2 is given, $L(x_1, x_2, \ldots x_n,) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e - \frac{ns^2-}{2\sigma^2}\ e^{\frac{-n(x-\mu)^2}{2\sigma^2}}$

If $L_1 = e^{\frac{-n(\bar{x}-\mu)^2}{2\sigma^2}}$ and $L_2 = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \ -\frac{ns^2}{2\sigma^2}$

The condition for sufficiency is satisfied and so $\bar{x}$ is a sufficient estimate of μ.
But when μ is given it is not possible to write $L(x_1, x_2, \ldots x_n,)$ as the product of two factors such that one is a function of σ and s^2 alone and the other is independent of σ. So s^2 is not a sufficient estimate of σ^2.

Ex. (1). Sufficient statistics for $P(\lambda)$ distribution :—

Let $(X_1, X_2, \dots, X_n)$ be a r.s. from $P(\lambda)$.

Then $\prod_{i=1}^{n} f(x_i;\lambda) = e^{-n\lambda} \cdot \dfrac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} \lfloor x_i}$, if $x_i = 0, 1, 2, \dots$

$$= g\left(T(\underset{\sim}{x}), \lambda\right) \cdot h(\underset{\sim}{x}) ;$$

where $h(\underset{\sim}{x}) = \dfrac{1}{\prod_{i=1}^{n} \lfloor x_i}$ and $T(\underset{\sim}{x}) = \sum_{i=1}^{n} x_i$

Hence, by factorization criterion, $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ is sufficient for λ.

Also note that,—

(i) $\underset{\sim}{T_1} = (X_1, X_2, \dots, X_n)$ is sufficient for λ, as

$$\underset{\sim}{1}' \underset{\sim}{T_1} = \sum_{i=1}^{n} X_i$$

(ii) $\underset{\sim}{T_2} = (X_1, \dots, X_{n-2}, X_{n-1} + X_n)$ is sufficient for λ, as

$$\underset{\sim}{1}' \underset{\sim}{T_2} = \sum_{i=1}^{n} X_i$$

⋮

(iii) $\underset{\sim}{T_{n-1}} = (X_1, X_2 + X_3 + \dots + X_n)$ is sufficient for λ.

It is clear that $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ reduces the space most and is to be preferred.

We should always looking for a sufficient statistic that results in the greatest reduction of the space.

Ex.(2). If $(X_1, X_2, \dots, X_n)$ be a r.s. from Bin$(1,p)$ or Bernoulli(p) distn. then find a one-dimensional sufficient statistic for p.

Soln. :—

$$\prod_{i=1}^{n} f(x_i; p) = \left\{ \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum x_i} \right\} \times 1$$

$$= g\{T(\underset{\sim}{x}), \theta\} \cdot h(\underset{\sim}{x}), \text{ where } h(\underset{\sim}{x}) = 1$$

and $T(\underset{\sim}{x}) = \sum_{i=1}^{n} x_i$

Hence $T = \sum_{i=1}^{n} X_i$ is sufficient estimator of θ.

$\therefore$ $\sum_{i=1}^{n} X_i$ is sufficient for θ, by factorization criterion.

Ex. (3). If $(x_1, x_2, \ldots, x_n)$ be a r.s. from $N(\mu, \sigma^2)$. Then find a two-dimensional sufficient statistic for (μ, σ).

Solution:- The PDF of $\underset{\sim}{x}$ is

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

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n . e^{\left(-\frac{\sum x_i^2}{2\sigma^2} + \frac{\mu\sum x_i}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}\right)}$$

$$= g\left(T(\underset{\sim}{x}); \mu, \sigma\right) . h(\underset{\sim}{x})$$

where, $h(\underset{\sim}{x}) = 1$ and $T(\underset{\sim}{x}) = \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right)$

$\therefore$ By factorization criterion, $T(\underset{\sim}{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is sufficient for (μ, σ).

Alternative :—

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

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

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

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

$$= g(\bar{x}, s^2; \mu, \sigma)\, h(\underset{\sim}{x}), \text{ where } h(\underset{\sim}{x}) = 1.$$

Hence $T(\underset{\sim}{X}) = (\bar{X}, S^2)$ is sufficient for (μ, σ).

Remark:- (1). If σ is unknown, then $\bar{X}$ is not sufficient for μ. But if σ is known $\bar{X}$ is sufficient for μ.

(2). If μ is unknown, then s^2 is not sufficient for σ but if μ is known then $T = \sum_{i=1}^{n}(X_i-\mu)^2 = (n-1)S^2 + n(\bar{X}-\mu)^2$ or $(\bar{X}, S^2)$ is sufficient for σ.

Ex. (1). Let $X_1, X_2, \ldots, X_n$ be a r.s. from Geometric (p). Suggest a one-dimensional sufficient statistic for p. Is $e^{\bar{X}}$ sufficient for p.

Hints:- $e^{\bar{x}}$ is a one-to-one function of $\bar{x}$.

Ex. (5). Uniform Distribution :—

Let $X_1, X_2, \ldots, X_n$ be a r.s. from ~~Rectangular~~ $U(0,\theta), \theta>0$. Find a one-dimensional sufficient statistic for θ. [ISI]

Soln. :- Here the domain of definition of $f(x;\theta)$, i.e. the range of the RV depends on θ, great care is needed.

The pdf of $\underset{\sim}{X}$ is

$$\prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} \frac{1}{\theta^n}, & \text{if } 0<x_i<\theta \ \forall\, i=1(1)n \\ 0, & \text{ow} \end{cases}$$

$$= \begin{cases} \frac{1}{\theta^n} & \text{if } 0 < x_{(1)} \le x_{(n)} < \theta \\ 0 & \text{ow} \end{cases}$$

$$= \begin{cases} \frac{1}{\theta^n} \cdot I(0, x_{(1)})\, I(x_{(n)}, \theta) ; & \text{where } I(a,b) = \begin{cases} 1 & \text{if } a<b \\ 0 & \text{if } a \ge b \end{cases} \\ 0 & ; \text{ow} \end{cases}$$

$$= \frac{1}{\theta^n} \cdot I(x_{(n)}, \theta) \cdot I(0, x_{(1)})$$

$$= g(T(\underset{\sim}{x}), \theta) \cdot h(\underset{\sim}{x}) \text{ ; where } h(\underset{\sim}{x}) = I(0, x_{(1)}) \text{ and }$$

$X_{(n)} = \{\max_{1\le i\le n} X_i\}$. $\qquad T(\underset{\sim}{x}) = x_{(n)}$.

$\therefore$ By factorization criterion, $T(\underset{\sim}{x}) = x_{(n)}$ is sufficient for θ.

Ex. (6) :- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $U(\theta_1, \theta_2)$; $\theta_1 < \theta_2$. Find a non-trivial sufficient statistic for (θ_1, θ_2).

Soln. :— Here the domain of definition of $f(x; \underset{\sim}{\theta})$ depends on θ_1 and θ_2, so great care is needed.

The PDF of $\underset{\sim}{X}$ is $\displaystyle\prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} \frac{1}{(\theta_2-\theta_1)^n} & \text{if } \theta_1 \le x_i \le \theta_2 \ \forall\, i=1(1)n \\ 0 & \text{ow} \end{cases}$

$$= \begin{cases} \frac{1}{(\theta_2-\theta_1)^n} & \text{if } \theta_1 \le x_{(1)} \le x_{(n)} \le \theta_2 \\ 0 & \text{ow} \end{cases}$$

$$= \frac{1}{(\theta_2-\theta_1)^n} I(\theta_1, x_{(1)})\, I(x_{(n)}, \theta_2) \text{, where}$$

$$I(a,b) = \begin{cases} 1 & \text{if } a \le b \\ 0 & \text{ow} \end{cases}$$

$$= g(T(\underset{\sim}{x}); \theta_1, \theta_2)\, h(\underset{\sim}{x})$$

where $h(\underset{\sim}{x}) = 1$ and $T(\underset{\sim}{x}) = (x_{(1)}, x_{(n)})$.

Hence, by fisher's factorization criterion, $T(\underset{\sim}{x}) = (X_{(1)}, X_{(n)})$ is sufficient for (θ_1, θ_2).

Remark:- The following examples are the particular cases of Ex.(6):-

Let $x_1, x_2, \ldots, x_n$ be a r.s. from

(i) $U(\theta - 1/2, \theta + 1/2)$

(ii) $U(\theta, \theta+1)$

(iii) $U(-\theta, \theta)$

Find a non-trivial sufficient statistic in each case.

Note:- As algebra says, for solving two unknowns, it is needed to have at least two equations.
For a single component parameter, it must contain at least one sufficient statistic.

Ex.(7). Let $(X_1, \ldots, X_n)$ be a r.s. from $U(-\theta, \theta)$, $\theta>0$. Find a one-dimensional sufficient statistic for θ.

Soln:→ The PDF of $\underset{\sim}{X}$ is

$$\prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} \left(\frac{1}{2\theta}\right)^n & \text{if } -\theta \le x_i \le \theta \ \forall\, i=1(1)n \\ 0 & \text{ow} \end{cases}$$

$$= \begin{cases} \left(\frac{1}{2\theta}\right)^n & \text{if } 0 \le |x_i| \le \theta \ \forall\, i=1(1)n \\ 0 & \text{ow} \end{cases}$$

$$= \begin{cases} \left(\frac{1}{2\theta}\right)^n, & 0 \le \min_i\{|x_i|\} \le \max_i\{|x_i|\} \le \theta \\ 0 & \text{ow} \end{cases}$$

$$= \left(\frac{1}{2\theta}\right)^n I\left(0, \min_i\{|x_i|\}\right) I\left(\max_i\{|x_i|\}, \theta\right);$$

where $I(a,b) = \begin{cases} 1 & \text{if } a \geq b \\ 0 & \text{ow} \end{cases}$

$$= g\left(T(\underset{\sim}{x}), \theta\right) h(\underset{\sim}{x}), \text{ where } h(\underset{\sim}{x}) = I\left(0, \min\{|x_i|\}\right)$$

Here, $T(\underset{\sim}{X}) = \max_i\{|x_i|\}$ is sufficient for θ.

Alt: Note that, here $X_i \overset{iid}{\sim} U(-\theta,\theta)\ \forall\, i=1(1)n$

$$\Rightarrow Y_i = |X_i| \overset{iid}{\sim} U(0,\theta)\ \forall\, i=1(1)n$$

By Ex.(5); $Y_n = \max_i\{|X_i|\}$ is sufficient for θ.

Remark:- Let T be sufficient for a family of distribution $\{f_i(x);\ i=1,2,\ldots\}$.
Here $f_i(x)$ may have the ~~same~~ different probability laws.
If $f_i(x)$ have the same probability law with an unknown constant (parameter) θ [e.g. $f_\theta(x) = N(\theta,1), \theta \in \mathbb{R}$] then we say that T is sufficient for θ.

Ex.(8). Let X be a single observation from a popln. belong to the family $\{f_0(x), f_1(x)\}$, where,

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ and } f_1(x) = \frac{1}{\pi(1+x^2)} \;;\; x \in \mathbb{R}$$

Find a non-trivial sufficient statistic for the family of distribution.

Solution:- Writing the family as $\{f_\theta(x) : \theta \in \Omega = \{0,1\}\}$

[Here the parameter θ is called labelling parameter]

Define, $I(\theta) = \begin{cases} 0 & \text{if } \theta = 0 \\ 1 & \text{if } \theta = 1 \end{cases}$

The PDF of X is

$$f_\theta(x) = \{f_0(x)\}^{1-I(\theta)} \{f_1(x)\}^{I(\theta)}$$

$$= \left\{\frac{f_1(x)}{f_0(x)}\right\}^{I(\theta)} \cdot f_0(x)$$

$$= \left\{\frac{\frac{1}{\pi(1+x^2)}}{\frac{1}{\sqrt{2\pi}} e^{-x^2/2}}\right\}^{I(\theta)} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$= g(T(x); \theta) \cdot h(x)$$

where $h(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ and $T(x) = x^2$ or $|x|$

Hence X^2 or $|X|$ is sufficient for the family of distn..

Ex(9). Let $X_1, X_2, \ldots, X_n$ be a r.s. from the PMF's

(i) $P[X=0] = \theta$, $P[X=1] = 2\theta$, $P[X=2] = 1-3\theta$; $0 < \theta < \frac{1}{3}$.

(ii) $P[X=k_1] = \frac{1-\theta}{2}$, $P[X=k_2] = \frac{1}{2}$, $P[X=k_3] = \frac{\theta}{2}$; $0 < \theta < 1$

Ans:- Find a non-trivial sufficient statistic in each case.

(i) Let $T_0(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases}$; $T_1(x) = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{ow} \end{cases}$; $T_2(x) = \begin{cases} 1 & \text{if } x = 2 \\ 0 & \text{ow} \end{cases}$

Then the PMF of X is

$$f(x;\theta) = \theta^{T_0(x)} (2\theta)^{T_1(x)} (1-3\theta)^{T_2(x)}$$

Hence the PMF of $\underset{\sim}{X}$ is

$$\prod_{i=1}^n f(x_i;\theta) = \theta^{\sum_{i=1}^n T_0(x_i)} (2\theta)^{\sum_{i=1}^n T_1(x_i)} (1-3\theta)^{\sum_{i=1}^n T_2(x_i)}$$

$$= \theta^{T_0} (2\theta)^{T_1} (1-3\theta)^{T_2}$$, where, $T_k = \sum_{i=1}^n T_k(x_i)$ represents the frequency of value k, $k = 0, 1, 2, \ldots$

and $T_0 + T_1 + T_2 = n$.

$$\therefore \prod_{i=1}^n f(x_i, \theta) = \theta^{n-T_2} (1-3\theta)^{T_2} \cdot 2^{T_1}$$

$$= g(T_2, \theta) \cdot h(\underset{\sim}{x})$$

Clearly, T_2, the frequency of value 2 in a r.s., is sufficient for θ.

Ex.(10). Let $X_1, X_2, \ldots, X_n$ be a r.s. from the following PDFs. Find the non-trivial sufficient statistic in each case.

(i) $f(x;\theta) = \begin{cases} \theta x^{\theta-1} & ; 0<x<1 \\ 0 & ; \text{OW} \end{cases}$ [ISI]

(ii) $f(x;\mu) = \dfrac{1}{|\mu|\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)}{2\mu^2}}$; $x \in \mathbb{R}$

(iii) $f(x;\alpha,\beta) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\beta(\alpha,\beta)} & , 0<x<1 \\ 0 & , \text{OW} \end{cases}$

(iv) $f(x;\mu,\lambda) = \begin{cases} \dfrac{1}{\lambda} e^{-\frac{(x-\mu)}{\lambda}} & , \text{if } x>\mu \\ 0 & , \text{OW} \end{cases}$

(v) $f(x;\mu,\sigma) = \begin{cases} \dfrac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2} & , \text{if } x>0 \\ 0 & , \text{OW} \end{cases}$

(vi) $f(x;\alpha,\theta) = \begin{cases} \dfrac{\theta\alpha^{\theta}}{x^{\theta+1}} & \text{if } x>\alpha \\ 0 & ; \text{OW} \end{cases}$

(vii) $f(x;\theta) = \begin{cases} \dfrac{2(\theta-x)}{\theta^2} & ; 0<x<\theta \\ 0 & ; \text{OW} \end{cases}$

Ans:- (i) The joint PDF of $X_1, X_2, \ldots, X_n$ is

$$f(\underset{\sim}{x}) = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta-1}$$

$$= g_{\theta}\left\{\prod_{i=1}^{n} x_i\right\} \cdot h(\underset{\sim}{x}) \text{ , where } h(\underset{\sim}{x}) = 1$$

and $T(\underset{\sim}{x}) = \left(\prod_{i=1}^{n} x_i\right)$

∴ By Neyman - Fisher Factorization criterion,

$T = \prod_{i=1}^{n} x_i$ is sufficient for θ.

(ii) $f(x;\mu,\sigma) = \dfrac{1}{|\mu|\sqrt{2\sigma}} \cdot e^{-\frac{(x-\mu)}{2\sigma^2}}$

So, $X \sim N(\mu, \mu^2)$, where $\mu \neq 0$.

By Ex.(3). $T(\underset{\sim}{x}) = \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right)$ is sufficient for μ.

- Note:- If in the range of X_i, there is the parameter of the distribution present, then we have to use the concept of Indicator function ($X_{(1)}$ or $X_{(n)}$) or $\min_i\{x_i\}$ or $\max_i\{x_i\}$.

(iii) $f_\theta(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, if $0<x<1$, $\alpha,\beta>0$

∴ Joint PDF of $x_1, \dots, x_n$ is

$$f(\underset{\sim}{x}) = \left[\frac{1}{B(\alpha,\beta)}\right]^n \left(\prod_{i=1}^n x_i\right)^{\alpha-1} \left(\prod_{i=1}^n (1-x_i)^{\beta-1}\right)$$

$= g(T(\underset{\sim}{x}); \alpha,\beta)\, h(\underset{\sim}{x})$, where $h(\underset{\sim}{x})=1$ and

$T(\underset{\sim}{x}) = \left(\prod_{i=1}^n x_i, \prod_{i=1}^n (1-x_i)\right)$ is jointly sufficient for (α,β)

(iv) $f(\underset{\sim}{x}) = \frac{1}{\theta^n} \cdot e^{-\sum_{i=1}^n \frac{(x_i-\mu)}{\sigma}}$ if $x_i > \mu$

$$= \frac{1}{\sigma^n} \cdot \exp\left\{\frac{-\sum_{i=1}^n x_i - n\mu}{\sigma}\right\} \cdot I(x_{(1)}, \mu), \text{ where}$$

$I(a,b) = 1$ if $a \geq b$
$= 0$ ow

$$= g\left(\sum_{i=1}^n x_i, x_{(1)}; \sigma, \mu\right) \cdot h(\underset{\sim}{x}), \text{ where } h(\underset{\sim}{x}) = 1.$$

Thus, $X_{(1)}$ and $\sum_{i=1}^n x_i$ are jointly sufficient statistic for μ and σ.

(v) $f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2}$; if $x>0$

The joint PDF of $\underset{\sim}{X}$ is

$$f(\underset{\sim}{x}) = \frac{1}{\left(\prod_{i=1}^n x_i\right)\sigma^n(\sqrt{2\pi})^n} \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (\ln x_i - \mu)^2\right\} \quad \text{, if } x_i>0$$

$$= \frac{1}{\sigma^n(\sqrt{2\pi})^n} \cdot e^{-\left(\frac{\Sigma(\ln x_i)^2}{2\sigma^2} - \mu\frac{\Sigma \ln x_i}{\sigma^2} + \frac{n\mu^2}{\sigma^2}\right)} \cdot \frac{1}{\left(\prod_{i=1}^n x_i\right)}$$

$$= T\left(\sum_{i=1}^n \ln x_i, \sum_{i=1}^n (\ln x_i)^2; \mu, \sigma\right) \cdot h(\underset{\sim}{x}); \text{ where},$$

$h(\underset{\sim}{x}) = \frac{1}{\prod_{i=1}^n x_i}$; $T(\underset{\sim}{X}) = \left(\sum_{i=1}^n \ln x_i, \sum_{i=1}^n (\ln x_i)^2\right)$

is sufficient for μ and σ.

(vi) $$f(\underset{\sim}{x}) = \theta^n \frac{(\alpha^\theta)^n}{\prod_{i=1}(x_i^{\theta+1})} \quad \text{if } x_i > \alpha$$

$$= (\theta\alpha^\theta)^n \cdot \frac{1}{\prod_{i=1}^{n}\{x_i\}^{\theta+1}} I(x_{(1)}, \alpha) \quad \text{if } x_{(1)} > \alpha$$

; where $I(a,b) = 1$ if $a > b$
$= 0$ o/w

$$= g\left(\prod_{i=1} x_i, x_{(1)}; \theta, \alpha\right) \cdot h(x) \text{ ; where,}$$

$h(\underset{\sim}{x}) = 1$ and hence

$T = \left(\prod_{i=1} X_i, X_{(1)}\right)$ is sufficient for θ and α.

(vii) $$f(\underset{\sim}{x}) = \frac{2^n}{\theta^{2n}} \prod_{i=1}^{n}(\theta - x_i) \quad ; 0 < x_i < \theta$$

$$= \left(\frac{2}{\theta^2}\right)^n \cdot (\theta - x_1)(\theta - x_2) \cdots\cdots (\theta - x_n) ; \; 0 < x_i < \theta$$

These can not be expressed in the form of factorization criterion.

So, $(X_1, X_2, \ldots, X_n)$ or $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ are trivially sufficient for θ here. $\therefore$ There is no non-trivial sufficient statistic.

Ex. 11. Let $X_1, \ldots, X_n$ be a r.s. from gamma distn. with pdf

$$f_\theta(x) = \frac{\alpha^p}{\Gamma(p)} \exp[-\alpha x] x^{p-1} \quad \text{if } 0 < x < \infty$$

where, $\alpha > 0, p > 0$

Show that $\sum_i X_i$ and $\prod_i X_i$ are jointly sufficient for (α, p).

Soln: $$f(\underset{\sim}{x}) = \left\{\frac{\alpha^p}{\Gamma p}\right\}^n \cdot \exp\left[-\alpha \sum_i x_i\right] \cdot \left(\prod x_i\right)^{p-1}$$

$$= g(T(\underset{\sim}{x}); \alpha, p) \cdot h(\underset{\sim}{x}) \text{ ; where } h(\underset{\sim}{x}) = 1.$$

$\therefore T(\underset{\sim}{x}) = \left(\sum_{i=1}^{n} X_i, \prod_{i=1}^{n} X_i\right)$ is jointly sufficient for (α, p).

Ex. 12. If $f(x) = \frac{1}{\theta} e^{-x/\theta}$; $0 < x < \theta$. Find a sufficient estimator for θ. [ISI]

Soln: $$f(\underset{\sim}{x}) = \frac{1}{\theta^n} \cdot \exp\left\{-\frac{1}{\theta}\sum_{i=1}^{n} x_i\right\}$$

$$= g\left\{\sum_{i=1}^{n} x_i, \theta\right\} \cdot h(\underset{\sim}{x}) \text{ ; where } h(\underset{\sim}{x}) = 1.$$

$\therefore T = \sum_{i=1}^{n} X_i$ is sufficient statistic for θ.

Ex.(13). If $f_\theta(x) = \frac{1}{2}$; $\theta-1 < x < \theta+1$, then show that $X_{(1)}$ and $X_{(n)}$ are jointly sufficient for θ. ($X_1 \sim U(\theta-1, \theta+1)$).

Soln.→ $f(\underset{\sim}{x}) = \left(\frac{1}{2}\right)^n$

$$= \frac{1}{2^n} \cdot I(\theta-1, x_{(1)})\, I(x_{(n)}, \theta+1) \; ; \; \theta-1 < x_{(1)} < x_{(n)} < \theta+1$$

where $I(a,b) = \begin{cases} 1 & \text{if } a < b \\ 0 & \text{if } a \geq b \end{cases}$

$$= g(T(\underset{\sim}{x}); \theta)\, h(\underset{\sim}{x}) \; ; \text{ where } h(\underset{\sim}{x}) = \frac{1}{2^n}.$$

$\therefore T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$ is jointly sufficient for θ.

Ex.(14). Let $X_1, X_2, \ldots, X_n$ be a r.s. from $C(\theta, 1)$, where θ is the location parameter. S.T. there is no sufficient statistic other than the trivial statistic $(X_1, X_2, \ldots, X_n)$ or $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$.

If a random sample of size $n \geq 2$ from a Cauchy distn with p.d.f.

$$f_\theta(x) = \frac{1}{\pi[1+(x-\theta)^2]}, \text{ where } -\infty < \theta < \infty, \text{ is considered.}$$

Then can you have a single sufficient statistic for θ?

Soln.→ The PDF of $(X_1, \ldots, X_n)$ is

$$\prod_{i=1}^{n} f(x_i; \theta) = \frac{1}{\pi^n \left\{\prod_{i=1}^{n} [1+(x_i-\theta)^2]\right\}}$$

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

$= \{1+(x_1-\theta)^2\}\{1+(x_2-\theta)^2\} \cdots \cdots \{1+(x_n-\theta)^2\}$

$= 1 +$ term involving one x_i + term involving two x_i's + ……
+ term involving all x_i's.

$$= 1 + \sum_i (x_i-\theta)^2 + \sum_{i \neq j}\sum (x_i-\theta)^2(x_j-\theta)^2 + \cdots\cdots + \prod_{i=1}^{n}(x_i-\theta)^2$$

Clearly, $\prod_{i=1}^{n} f(x_i;\theta)$ can not be written as $g(T(\underset{\sim}{x}), \theta) \cdot h(\underset{\sim}{x})$ for a statistic other than the trivial choices

$(X_1, \ldots, X_n)$ or $(X_{(1)}, \ldots, X_{(n)})$.

Hence there is no non-trivial sufficient statistic

Therefore, in this case, no reduction in the space is possible.

⇒ The whole set $(X_1, \ldots, X_n)$ is jointly sufficient for θ.

Ex.(15). Let X_1 and X_2 be iid RVs having the discrete uniform distribution on $\{1,2,\cdots,N\}$, where N is unknown. Obtain the conditional distribution of X_1, X_2, given $(T=\max(X_1,X_2))$. Hence show that T is sufficient for N but X_1+X_2 is not.

Ans:- (i) $P(T=t) = P[\max(X_1,X_2)=t]$

$$= P[X_1<t, X_2=t] + P[X_1=t, X_2<t] + P[X_1=t, X_2=t]$$

$$= P[X_1<t]P[X_2=t] + P[X_1=t]P[X_2<t] + P[X_1=t]P[X_2=t]$$

Now, $P[X_1<t] = P[X_1=1]+P[X_1=2]+\cdots+P[X_1=t-1]$

$$= \underbrace{\frac{1}{N}+\frac{1}{N}+\cdots+\frac{1}{N}}_{(t-1)\text{ times}}$$

$$= \frac{t-1}{N}.$$

& $P[X_1=t] = P[X_2=t] = \frac{1}{N}$

$$\therefore P[T=t] = \frac{1}{N}\cdot\frac{t-1}{N} + \frac{t-1}{N}\cdot\frac{1}{N} + \frac{1}{N}\cdot\frac{1}{N}$$

$$= \frac{2(t-1)+1}{N^2}$$

$$\therefore P[X_1=x_1, X_2=x_2 \mid T=t] = \begin{cases} \dfrac{P[X_1=x_1, X_2=x_2]}{P[T=t]} & \text{if } \max(x_1,x_2)=t, \\ 0 & \text{, ow} \end{cases}$$

$$= \frac{\frac{1}{N}\cdot\frac{1}{N}}{\frac{2(t-1)+1}{N^2}} = \frac{1}{2(t-1)+1},$$

which is independent of N.

(ii) $T=X_1+X_2$. Then,

for $2\le t\le N+1$; $P[T=t] = P[X_1=1, X_2=t-1] + P[X_1=2, X_2=t-2] + \cdots + P[X_1=t-1, X_2=1]$

$$= \frac{t-1}{N^2}.$$

for $N+2\le t\le 2N$; $P[T=t] = P[X_1=t-N, X_2=N] + P[X_1=t-N+1, X_2=N-1] + \cdots + P[X_1=N, X_2=t-N]$

$$= \frac{2N-t+1}{N^2}$$

$$\therefore P[X_1=x_1; X_2=x_2 \mid T=t] = \frac{P[X_1=x_1; X_2=x_2]}{P[X_1+X_2=t]}$$

$$= \begin{cases} \dfrac{1/N^2}{\frac{t-1}{N^2}} = \dfrac{1}{t-1} & \text{if } x_1+x_2=t \\ \dfrac{1/N^2}{\frac{2N-t+1}{N^2}} = \dfrac{1}{2N-t+1} & \text{if } x_1+x_2=t \end{cases}$$

which depends on N, so for the 2nd case (X_1+X_2) is not sufficient.

Ex.(2). Example of a statistic that is not sufficient :—

Let (X_1, X_2, X_3) be a r.s. from $Bin(1, p)$. Is $T = X_1 + 2X_2 + X_3$ sufficient for p? Is $X_1 X_2 + X_3$ is sufficient for p?

Ans:-

(i) Here T takes the values $0, 1, 2, 3, 4$.

$$P[X_1 = 1, X_2 = 0, X_3 = 1 \mid T = 2]$$

$$= \frac{P[X_1 = 1, X_2 = 0, X_3 = 1; T = 2]}{P[T = 2]}$$

$$= \frac{P[X_1 = 1, X_2 = 0, X_3 = 1]}{P[X_1 = 1, X_2 = 0, X_3 = 1] + P[X_1 = 0, X_2 = 1, X_3 = 0]}$$

$$= \frac{p^2(1-p)}{p^2(1-p) + p(1-p)^2} = \frac{p}{p + 1 - p} = p \text{, which depends on } p.$$

Hence T is not sufficient for p.

(ii) Here, $X_1 X_2 + X_3 = T$

Let us consider a specific case, $x_1 = 1, x_2 = 1, x_3 = 0$ and $T = 1$.

Here $x_1 x_2 + x_3 = 1$ for,

$$\{(x_1 = 1, x_2 = 1, x_3 = 0), (x_1 = 1, x_2 = 0, x_3 = 1), (x_1 = 0, x_2 = 1, x_3 = 1), (x_1 = 0, x_2 = 0, x_3 = 1)\}$$

$$\therefore P[(X_1 = 1, X_2 = 1, X_3 = 0) \mid T = 1]$$

$$= \begin{cases} \dfrac{P[X_1 = 1, X_2 = 1, X_3 = 0]}{P[T = 1]} & \text{, if } T = 1 \\ 0 & \text{, ow} \end{cases}$$

$$= \begin{cases} \dfrac{p^2(1-p)}{3p^2(1-p) + (1-p)^2 p} & \text{, if } T = 1 \\ 0 & \text{, ow} \end{cases}$$

$$= \begin{cases} \dfrac{p}{2p + 1} & \text{, if } T = 1 \\ 0 & \text{, ow} \end{cases}$$

i.e. T is not sufficient for p.

Efficiency

Let t_1 and t_2 two unbiased estimates of a parameter θ. Then t_1 is said to be more efficient than t_2 if V (t_1) is less than V (t_2). The ratio $\frac{V(t_2)}{V(t_1)}$ is called the relative efficiency of t_1 with respect to t_2 and it may be used to compare the efficiencies of the estimates.

Example 1:

Let $X_1, X_2, \ldots, X_n$ be a sample from a population N(μ, σ). Define $t_1 = X_1$, $t_2 = \frac{X_1+X_2}{2}$, …, $t_n = \frac{X_1+X_2+\cdots+X_n}{n}$ are proposed estimates of μ.

Now $E(t_1) = E(X_1) = \mu$, $E(t_2) = \frac{E(X_1)+E(X_2)}{2} = \frac{\mu+\mu}{2} = \mu$, Similarly we can prove that $E(t_3) = E(t_4) = \cdots = E(t_n) = \mu$. So all the estimates t_1 , t_2 , t_3, …,t_n are unbiased.

$$V(t_1) = V(X_1) = \sigma^2$$
$$V(t_2) = \frac{V(X_1) + V(X_2)}{4} = \frac{\sigma^2 + \sigma^2}{4} = \frac{\sigma^2}{2}$$
$$V(t_n) = \frac{V(X_1) + V(X_2) + \cdots + V(X_n)}{n^2} = \frac{\sigma^2}{n}$$

Which indicate that t_2 is more efficient than t_1 , t_3 is more efficient than t_1 and t_2, So on t_n is more efficient than t_1, t_2, t_3, …, t_{n-1}, as $\sigma^2 > \frac{\sigma^2}{2} > \frac{\sigma^2}{3} > \cdots > \frac{\sigma^2}{n}$.

Example 2:

Let X_1, X_2, X_3 be a sample from a population N(μ, σ). Define $t_1 = X_1 + X_2 - X_3$ and $t_2 = 2X_1 + 3X_2 - 4X_3$. Compare the efficiency of t_1 and t_2.

$$E(t_1) = E(X_1) + E(X_2) - E(X_3) = \mu + \mu - \mu = \mu$$

and

$$E(t_2) = 2E(X_1) + 3E(X_2) - 4E(X_3) = 2\mu + 3\mu - 4\mu = \mu$$

So both t_1 and t_2 are unbiased estimate of μ. Now

$$V(t_1) = V(X_1) + V(X_2) + V(X_3) = \sigma^2 + \sigma^2 + \sigma^2 = 3\sigma^2.$$

and

$$V(t_2) = 4V(X_1) + 9V(X_2) + 16V(X_3) = 4\sigma^2 + 9\sigma^2 + 16\sigma^2 = 29\sigma^2.$$

$\therefore V(t_1) < V(t_2)$. Hence t_1 more efficient than t_2.

The relative efficiency of t_1 with respect t_2 is $\frac{V(t_2)}{V(t_1)} = \frac{3\sigma^2}{29\sigma^2} = \frac{3}{29}$.

Cramer-Rao inequality, condition for its attainment and the method of minimum variance

Cramer-Rao inequality, is defined as follows
The Let f(x, θ) be the p.d.f of a random variable X with only one parameter θ. Let x_1 , x_2 , …x_n, be a random sample taken from the population and let t(x_1 , x_2 , …x_n,) be an unbiased estimator of θ ie, E(t) = θ. If,

1. The range of variation of X is independent of θ and
2. Differentiation under the integral sign or summation sign (according as the variable is continuous or discrete) is valid for f(x, θ),

$$V(t) \geq \frac{1}{\left[E\left(\frac{\partial logL}{\partial\theta}\right)^2\right]}$$

Where V(t) is the sampling variance of t and L, the likelihood function of the sample.

Note: If t is not an unbiased estimate of θ, but E(t) = $\varphi(\theta)$

Where $\varphi(\theta)$ is a function of θ, the inequality becomes,

$$V(t) \geq \frac{[\varphi']^2}{\left[E\left(\frac{\partial logL}{\partial\theta}\right)^2\right]} \quad \text{where } \varphi'(\theta) = \left(\frac{\partial\varphi(\theta)}{\partial\theta}\right)$$

It can be shown that $E\left(\frac{\partial logL}{\partial\theta}\right)^2 = -\ E\left(\frac{\partial^2 logL}{\partial\theta^2}\right)$

Note that the inequality becomes an equality when $\frac{\partial \log L}{\partial\theta}$ = A(t – θ).

Where A is independent of the observations but may be a function of θ. If this condition is satisfied we see that t is an unbiased minimum variance estimate of θ. If such an estimate exists it is the most efficient estiate and hence the best estimate that we can get. In this case it can be shown that 1/A is the minimum value of the variance of t. This leads us to a method of estimation known as the method of minimum variance.

The Method of Minimum Variance

Let f(x, θ) be the p.d.f of the population with one parameter θ and (x_1 , x_2 , …x_n,) a random sample. Let L(x_1 , x_2 , …x_n: , θ) be the likelihood function. If $\frac{\partial \log L}{\partial\theta}$ can be put in the form k(t – θ) where k is either a constant or a function of θ and t a function of the observations only, then t is the minimum variance unbiased estimator of θ.

Example:

Let $X_1, X_2, \ldots, X_n$ be a sample from a population N(μ, σ) where σ is assumed to be known. Then the likelihood of the sample L can be written as

$L(x_1, x_2, \ldots x_n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \quad \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_2-\mu)^2}{2\sigma^2}} \quad \ldots\ldots\ldots \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}$

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

So taking logarithm we get

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

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

if we take

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

Then t is a minimum variance unbiased estimate of μ with variance $\frac{\sigma^2}{n}$.

Problem:- Let X be a single observation from $P(\lambda)$. Is $\frac{1}{\lambda}$ unbiasedly estimable based on X ?

ANS:- X be a single observation from $P(\lambda)$. PMF of X is given by, $f(x) = \frac{e^{-\lambda}.\lambda^x}{x!}$, $x = 0, 1, 2,$ $\lambda >$

Now, $E(X) = \sum_{x=0}^{\infty} x . \frac{e^{-\lambda}.\lambda^x}{x!}$

$= \lambda \sum_{x-1=0}^{\infty} \frac{e^{-\lambda}.\lambda^{x-1}}{(x-1)!}$

$= \lambda . e^{-\lambda}. e^{\lambda}$

$= \lambda$.

$\therefore \frac{1}{\lambda}$ is not unbiasedly estimable based on where $X \sim P(\lambda)$.

Remark 3. There may exist infinitively many unbiased estimator.

example: Let us consider $X_1, X_2, \ldots\ldots, X_n \sim$ i.i.d. $P(\lambda)$

Then $E_\lambda(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E_\lambda(X_i) = \frac{1}{n}\cdot n\lambda = \lambda$

and $E_\lambda\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = (n-1)\lambda$

i.e. $E_\lambda(s^2) = \lambda$, where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Let us define, $T_\alpha(\underset{\sim}{X}) = \alpha\bar{X} + (1-\alpha)s^2, \quad 0 \le \alpha \le 1$

$$E_\lambda\{T_\alpha(\underset{\sim}{X})\} = \alpha E(\bar{X}) + (1-\alpha)E(s^2)$$
$$= \alpha\lambda + \lambda(1-\alpha)$$
$$= \lambda$$

$\therefore$ For each $\alpha \in [0,1]$, $T_\alpha(\underset{\sim}{X})$ is unbiased for λ.

Hence this completes the proof.

is $\gamma(\theta)$,

Remark 1. Unbiased estimators does not always exist.

example:

i) Let us consider a random variable, $X \sim bin(1, p)$. Suppose we want to estimate the parametric function $\gamma(p) = p^2$. Now, for a statistic $T(X)$ to be unbiased for $\gamma(p)$, one must require,

$E_p(T(X)) = p^2, \quad 0<p<1$

i.e. $p^2 = pT(1) + (1-p)T(0)$

$\Rightarrow p^2 + p[T(0) - T(1)] - T(0) = 0$

But the LHS of the above expression is a power series (with at least one co-efficient non-zero), which vanishes $\forall\, p \in (0,1)$, which is impossible. Therefore, we can't have an unbiased estimator for p^2.

ii) Suppose $X \sim Bin(n, p)$, where, n is specified. Here, no unbiased estimator of $\frac{1}{p}$ exists based on X. If possible let,

$T(x)$ is unbiased for $\frac{1}{p}$.

$\therefore E_p[T(X)] = \frac{1}{p} \;\forall\; p \in (0,1)$

$\Rightarrow \sum_j T(j)\binom{n}{j} p^j (1-p)^{n-j} = \frac{1}{p}$

Note that, LHS $\le \sum_j |T(j)|\binom{n}{j} p^j (1-p)^{n-j}$ [a finite quantity]

But RHS $\to \infty$ as $p \to 0$, i.e. a contradiction occurs.

iii) Catch-recatch Problem : — Let there be θ fishes in a tank of which M are caught, tagged and released. thereafter n fishes are caught again of which x are found to be tagged then there does not exist any unbiased estimator of θ based on x.

Note that,
$$P_\theta[X=x] = \frac{\binom{M}{x}\binom{\theta-M}{n-x}}{\binom{\theta}{n}}$$

Given the sample the parameter space is
$$\theta \in \{(M+n-x), (M+n-x+1), \ldots\ldots\}$$
i.e. the parameter space is not bounded above. If possible let, $T(X)$ be unbiased for θ,

Define, $a = \min\{T(0), T(1), \ldots, T(n)\}$
$b = \max\{T(0), T(1), \ldots, T(n)\}$

evidently,
$$a \le E_\theta\{T(X)\} \le b$$
$$\Rightarrow a \le \theta \le b$$

Hence, the contradiction, since the parameter space is not bounded.

Remark 2. Unbiased estimator may sometimes be absurd.

example:

i) Let us consider the random variable, $X \sim P(\lambda)$.
Let us define a statistic $T(X) = (-2)^X$ for estimating the parametric function $\gamma(\lambda) = e^{-3\lambda}$

$$\therefore E_\lambda[T(X)] = \sum_{x=0}^{\infty} (-2)^x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=0}^{\infty} e^{-\lambda}.\frac{(-2\lambda)^x}{x!}$$
$$= e^{-\lambda}.e^{-2\lambda}$$
$$= e^{-3\lambda}$$

$\therefore T(X) = (-2)^X$ is an unbiased estimator for $\gamma(\lambda) = e^{-3\lambda}$

but $(-2)^X = \begin{cases} +ve & \text{if } X \text{ is even} \\ -ve & \text{if } X \text{ is odd} \end{cases}$

i.e. if x is odd, the $(-2)^X$ is negative., and it is absurd to have a negative estimator of a positive parametric function.

ii) Let $X_1, \ldots, X_n$ be a random sample drawn from a $N(\mu, 1)$ population. We know $\bar{X} \sim N(\mu, \frac{1}{n})$. Here $\bar{X}^2 - \frac{1}{n}$ unbiasedly estimate μ^2 which is positive for $\mu \neq 0$, where as an unbiased estimate may occasionally be negative.

- Method of moments: EXAMPLE : $\longrightarrow$

Example 1. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ. As there is only one parameter, hence only one equation, which is

$$M_1' = \mu_1' = \mu_1'(\lambda) = \lambda.$$

Hence the method-of-moments estimator of λ is $M_1' = \overline{X}$, which says estimate the population mean λ with the sample mean $\overline{x}$.

Example 2. Let $X_1, X_2, \ldots, X_n$ be a random sample from the negative exponential density $f(x;\theta) = \theta e^{-\theta x} I_{(0,\infty)}(x)$. To estimate θ, The method-of-moments equation is

$$M_1' = \mu_1' = \mu_1'(\theta) = \frac{1}{\theta};$$

Hence the method-of-moments estimator of θ is $1/M_1' = \frac{1}{\overline{X}}$.

Example 3. Let $X_1, X_2, \ldots, X_n$ be a random sample from a uniform distribution on $(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)$. Here the unknown parameters are two, namely μ and σ, which are the population mean and standard deviation. The method-of-moments equations are

$$M_1' = \mu_1' = \mu_1'(\mu, \sigma) = \mu$$

and

$$M_2' = \mu_2' = \mu_2'(\mu, \sigma) = \sigma^2 + \mu^2;$$

Hence the method-of-moments estimators are $\overline{X}$ for μ and

$$\sqrt{\frac{1}{n}\sum X_i^2 - \overline{X}^2} = \sqrt{\frac{1}{n}\sum (X_i - \overline{X})^2} \text{ for } \sigma.$$

Remark:— Method-of-moments estimators are not uniquely defined.

FURTHER PROBLEMS:—

Ex. 1. Estimating p^2 for Bernoulli distribution

(a) Let $X_1, X_2, \ldots, X_n$ be a r.s. from $B(1,p)$, $0<p<1$, $n \geq 2$. Can we estimate p^2 unbiasedly based on $X_1, \ldots, X_n$? If so, how?

(b) Let X be a single observation from $B(1,p)$. Can you estimate p^2 unbiasedly based on X?

Solution:-

(a) Let $T = \sum_{i=1}^{n} X_i$. Then T denotes the no. of successes in n independent Bernoulli trials.

Hence, $T \sim Bin(n, p)$.

$\left[\because E\left[(T)_r\right] = (n)_r \cdot p^r,\ r \leq n\right]$

We have, $E\{T(T-1)\} = n(n-1)p^2$

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

Hence $h(T) = \frac{T(T-1)}{n(n-1)}$ is an UE of p^2.

(b) If possible, let $T(x)$ be an UE of p^2.

Then by definition,

$$E(T(X)) = p^2 \quad \forall\ p \in (0,1)$$

$$\Rightarrow \sum_{x=0}^{1} T(x)\, P[X=x] = p^2$$

$$\Rightarrow T(0)\cdot(1-p) + T(1)p = p^2$$

$$\Rightarrow p^2 + \{T(0) - T(1)\}p - T_0 = 0 \quad \forall\ p \in (0,1) \quad \text{—— (i)}$$

Clearly, (i) is an identity in p.

Equating the coefficients of p^2, p and constant term, we get,

$1 = 0 \rightarrow$ absurd

and $T(0) - T(1) = 0$

Hence, there exists no $T(X)$ which will satisfy " $E[T(X)] = p^2$ " $\forall\ p \in (0,1)$.

Hence, there is no UE of p^2 based on a single observation X from $Bin(1,p)$.

Ex. (2). Let X be a single observation from $P(\lambda)$. Does there exist an UE of $\frac{1}{\lambda}$?

Solution:— If possible, let $T(X)$ be an UE of $\frac{1}{\lambda}$.

Then $E(T(X)) = \frac{1}{\lambda}\ \forall\ \lambda > 0$

$$\Rightarrow \sum_{x=0}^{\infty} T(x)\, e^{-\lambda} \frac{\lambda^x}{x!} = \frac{1}{\lambda} \quad \forall\ \lambda > 0$$

$$\Rightarrow \sum_{x=0}^{\infty} T(x) \cdot \frac{\lambda^{x+1}}{x!} = e^{\lambda}$$

$$\Rightarrow \sum_{x=0}^{\infty} T(x) \cdot \frac{\lambda^{x+1}}{x!} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}\ ,\ \lambda > 0$$

$$\Rightarrow 1 + \left\{\frac{1}{1!} - \frac{T(0)}{0!}\right\}\lambda + \left\{\frac{1}{2!} + \frac{T(1)}{1!}\right\}\lambda^2 + \cdots\cdots = 0\ \forall\ \lambda > 0$$

By uniqueness of power series, we have

$$1 = 0 \text{ (absurd)}$$

$$\frac{1}{1!} - \frac{T(0)}{0!} = 0,\quad \frac{1}{2!} - \frac{T(1)}{1!} = 0,\ \cdots\cdots$$

Hence, there exists no UE of $\frac{1}{\lambda}$ based on X.

Ex. 3.

(a) Starting from the equation $\sigma^2 = E(X^2) - \mu^2$, we get $\mu^2 = E(X^2 - \sigma^2)$ and $(X^2 - \sigma^2)$ is an UE of μ^2, what is its principal defects?

Solution:—

Hints:— (a) If σ is unknown, then $(X^2 - \sigma^2)$ is not a statistic and not measurable or observable. Then, $(X^2 - \sigma^2)$ can not be used as an estimator of μ^2.

(b) Show that if $\hat{\theta}$ is an UE of θ and $Var(\hat{\theta}) \neq 0$, $\hat{\theta}^2$ is not an UE of θ^2.

Hints:—

$$0 < Var(\hat{\theta}) = E(\hat{\theta}^2) - E^2(\hat{\theta})$$
$$= E(\hat{\theta}^2) - \theta^2$$
$$\Rightarrow E(\hat{\theta}^2) > \theta^2.$$

Ex. 4. Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(0, \sigma^2)$ distn. Suggest an UE of σ based on $\sum_{i=1}^{n} |X_i|$ and also an alternative UE based on $\sum_{i=1}^{n} X_i^2$.

Solution: – Note that, $E\left(\sum_{i=1}^{n} |X_i|\right) = \sum_{i=1}^{n} E|X_i| = \sum_{i=1}^{n} \sigma \sqrt{\frac{2}{\pi}}$

$$= \sigma \cdot n \cdot \sqrt{\frac{2}{\pi}}$$

$$\Rightarrow E\left\{\sqrt{\frac{\pi}{2}} \cdot \frac{1}{n} \sum_{i=1}^{n} |X_i|\right\} = \sigma$$

$$\Rightarrow T_1 = \sqrt{\frac{\pi}{2}} \cdot \left(\frac{1}{n} \sum_{i=1}^{n} |X_i|\right) \text{ is an UE of } \sigma^2.$$

Now, $\chi^2 = \dfrac{\sum_{i=1}^{n} X_i^2}{\sigma^2} \sim \chi_n^2$

$$\left[\; E(\chi^2) = n \Rightarrow E\left(\frac{1}{n} \sum_{i=1}^{n} X_i^2\right) = \sigma^2 \right.$$

$$\left. \Rightarrow \frac{1}{n} \sum_{i=1}^{n} X_i^2 \text{ is an UE of } \sigma^2 \right]$$

Now, $E\left[\sqrt{\chi^2}\right] = \displaystyle\int_0^\infty \sqrt{x} \cdot \frac{1}{2^{n/2} \Gamma(n/2)} \cdot e^{-x/2} \, x^{\frac{n}{2}-1} \, dx$

$$= \frac{2^{\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right)}{2^{n/2} \Gamma(n/2)} = \frac{\sqrt{2}\, \Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = c_n, \text{ say}$$

$$\Rightarrow E\left(\frac{\sum_{i=1}^{n} X_i^2}{\sigma^2}\right)^{1/2} = c_n \Rightarrow E\left(\frac{1}{c_n} \cdot \sqrt{\sum_{i=1}^{n} X_i^2}\right) = \sigma.$$

$$\Rightarrow T_2 = \frac{1}{c_n} \cdot \sqrt{\sum_{i=1}^{n} X_i^2} \text{ is an UE of } \sigma.$$

Ex. 5. Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(\mu, 1)$. Find an UE of μ^2.

Solution: – $V(\bar{X}) = \frac{1}{n}$

$$\Rightarrow E(\bar{X}^2) - E^2(\bar{X}) = \frac{1}{n}$$

$$\Rightarrow E\left(\bar{X}^2 - \frac{1}{n}\right) = \mu^2.$$

Note that, the estimator $\left(\bar{X}^2 - \frac{1}{n}\right)$ can take negative values in estimating a positive parameter μ^2 and $\left(\bar{X}^2 - \frac{1}{n}\right)$ is not so sensitive.

Ex.6. Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(\mu, \mu)$, $\mu > 0$. Find an UE of μ^2 based on both $\bar{X}$ and s^2.

Solution:-

Here $\bar{X}$ is an UE of population mean $E(X_1) = \mu$ and s^2 is UE of popl^n. variance $V(X_1) = \mu$.

Hence, $E(\bar{X} \cdot s^2) = E(\bar{X}) \cdot E(s^2) = \mu^2$.

[For a normal sample, $\bar{X}$ and s^2 are independently distributed]

N.T. $\alpha\bar{X} + (1-\alpha)s^2$ is an UE of μ, $0 \le \alpha \le 1$.

Ex.7. Let $X_1, X_2, \ldots, X_n$ be a r.s. from the PDF

$$f(x) = \begin{cases} \theta x^{\theta-1} & , 0<x<1 \\ 0 & , \text{OW} \end{cases}, \text{ where } \theta > 0.$$

Find an UE of (i) $\frac{1}{\theta}$, (ii) θ.

Solution:- Let $Z_i = -2\theta \ln X_i$, then $X_i = e^{-\frac{Z_i}{2\theta}}$

The PDF of Z_i is,

$$f_{Z_i}(z_i) = \begin{cases} \theta\left(e^{-\frac{z_i}{2\theta}}\right)^{\theta-1} \left|\frac{d}{dz_i}\left(e^{-\frac{z_i}{2\theta}}\right)\right| & , \text{if } 0<z_i<\infty \\ 0 & , \text{OW} \end{cases}$$

$$= \begin{cases} \frac{1}{2} e^{-z_i/2} & , 0<z_i<\infty \\ 0 & , \text{OW} \end{cases}$$

$$\Rightarrow Z_i \overset{iid}{\sim} \chi^2_2 \quad \forall\, i=1(1)n.$$

$$\Rightarrow \sum_{i=1}^{n} Z_i \sim \chi^2_{2n}$$

i.e. $Y_1 = \sum_{i=1}^{n}(-2\theta \ln X_i) \sim \chi^2_{2n}$

Now, $E\left(\sum_{i=1}^{n} -2\theta \ln X_i\right) = 2n$

$$\Rightarrow E\left(-\frac{1}{n}\sum_{i=1}^{n} \ln X_i\right) = \frac{1}{\theta}$$

$$\Rightarrow T_1 = \frac{1}{n}\sum_{i=1}^{n} -\ln X_i \text{ is an UE of } \frac{1}{\theta}.$$

ii) Now, $E\left(\frac{1}{Y}\right) = E\left(\frac{1}{\chi^2_{2n}}\right) = 2^{-1}\frac{\Gamma\left(\frac{2n}{2}-1\right)}{\Gamma\left(\frac{2n}{2}\right)}$ if $n>1$

$$= \frac{1}{2} \cdot \frac{\Gamma(n-1)}{\Gamma(n)} = \frac{1}{2(n-1)}, \; n>1.$$

$$\Rightarrow E\left(\frac{1}{\sum_{i=1}^{n} -2\theta \ln X_i}\right) = \frac{1}{2(n-1)}, \; n>1$$

$$\Rightarrow E\left(\frac{n-1}{\sum_{i=1}^{n} -\ln X_i}\right) = \theta, \; n>1.$$

$$\Rightarrow T_2 = \frac{n-1}{\sum_{i=1}^{n} -\ln X_i} \text{ is an UE of } \theta.$$

Ex. 8. Unbiased estimator may sometimes be absurd.

Give an example of Absurd Unbiased estimator.

Solution:- Let X be a single observation of $P(\lambda)$. If possible, let, $T(X)$ be an UE of $e^{-3\lambda}$.

Then $E[T(X)] = e^{-3\lambda}$, $\forall \lambda > 0$

$$\Rightarrow \sum_{x=0}^{\infty} T(x) \cdot e^{-\lambda} \cdot \frac{\lambda^x}{x!} = e^{-3\lambda}$$

$$\Rightarrow \sum_{x=0}^{\infty} T(x) \cdot \frac{\lambda^x}{x!} = e^{-2\lambda} = \sum_{x=0}^{\infty} \frac{(-2\lambda)^x}{x!}, \lambda > 0$$

By uniqueness of Power series, we have

$$\frac{T(x)}{x!} = \frac{(-2)^x}{x!} \quad \forall x = 0, 1, 2, \ldots$$

$\Rightarrow T(x) = (-2)^x \; \forall x = 0, 1, 2, \ldots$

Hence, $T(x) = (-2)^x$ is the unique UE of $e^{-3\lambda}$.

N.T. $T(x) = (-2)^x = \begin{cases} 2^x, & x = 0, 2, 4, \ldots \\ -2^x, & x = 1, 3, 5, \ldots \end{cases}$

Hence, $T(X)$ is UE but it takes negative values in estimating a positive parameter $e^{-3\lambda}$. This is an example of absurd UE.

Remark:- (1) Here $T(X) = (-2)^X$ is the only on unique UE of $e^{-3\lambda}$. Hence, $T(X) = (-2)^X$ is the UMVUE of $e^{-3\lambda}$.

(2) For $X \sim P(\lambda)$, $P_X(t) = e^{\lambda(t-1)}$, $t \in \mathbb{R}$

$\Rightarrow E[t^X] = e^{\lambda(t-1)}$, $t \in \mathbb{R}$

Put, $t = -2$,

$$E[(-2)^X] = e^{-3\lambda}.$$

Ex. 9. If $X \sim Bin(n, p)$, then show that only polynomial in p of degree $\leq n$ are unbiasedly estimable.

Solution:- [A parametric function $\psi(\theta)$ is unbiasedly estimable if $E\{T(X)\} = \psi(\theta)$, for some $T(X)$, $\forall \theta \in \Omega$.]

Let $\psi(p)$ be an unbiasedly estimable parametric function.

Then $\exists$ a statistic $T(X) \ni$

$$\psi(p) = E(T(X)) \quad \forall p \in (0,1)$$

$$= \sum_{x=0}^{n} T(x) \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} T(x) \cdot \binom{n}{x} p^x \left\{ \sum_{k=0}^{n-x} \binom{n-x}{k} (-p)^k \right\}$$

$$= \sum_{x=0}^{n} \sum_{k=0}^{n-x} (-1)^k T(x) \binom{n}{x} \binom{n-x}{k} p^{x+k},$$ which is a polynomial in p of degree $\leq n$.

Remark:- N.T. (i) $\sqrt{p}$, (ii) $\frac{1}{p}$, (iii) e^p, (iv) $\log p$ are not polynomials and hence not unbiasedly estimable. If $X \sim B(1, p)$, then only linear function in p are unbiasedly estimable. Hence, p^2, a 2nd degree polynomial is not unbiasedly estimable.

Best Linear Unbiased Estimator (BLUE) :—

Let $X_1, X_2, \ldots, X_n$ be a r.s. from a population with mean μ and variance $\sigma^2 (<\infty)$. Then an estimator $T = \sum_{i=1}^{n} a_i X_i$ is called a linear estimator. A linear estimator $T = \sum_{i=1}^{n} a_i X_i$ is unbiased for μ

iff $E(T) = \mu \ \forall \mu$

iff $\left(\sum_{i=1}^{n} a_i\right)\mu = \mu \ \forall \mu$

iff $\sum_{i=1}^{n} a_i = 1$.

[The estimator , $T = \sum_{i=1}^{n} a_i e^{X_i}$ is not linear estimator also, $T_3 = \bar{X}^2, T_4 = s^2$ are linear estimators.]

Definition:— A linear unbiased estimator $T = \sum_{i=1}^{n} a_i X_i$ with $\sum_{i=1}^{n} a_i = 1$ of μ that has the minimum variance among all linear unbiased estimators of μ, is called the BLUE of μ.

Theorem:— If $X_1, X_2, \ldots, X_n$ be a r.s. from a population with mean μ and variance σ^2, show that the sample mean $\bar{X}$ is the BLUE of μ. [WBSU'11]

Proof:— BLUE of μ is the estimator which has the minimum variance in the class $\mathcal{C} = \left\{ T : T = \sum_{i=1}^{n} a_i X_i , \sum_{i=1}^{n} a_i = 1 \right\}$ of all linear UEs of μ.

Note that, $Var(T) = \left(\sum_{i=1}^{n} a_i^2\right)\sigma^2$, as X_i's are iid and $\sum_{i=1}^{n} a_i = 1$.

To minimize $Var(T) = \sigma^2 \left(\sum_{i=1}^{n} a_i^2\right)$ subject to $\sum_{i=1}^{n} a_i = 1$,

By C-S inequality,

$$\left(\sum_{i=1}^{n} a_i . 1\right)^2 \le \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} 1^2\right)$$

$$\Rightarrow \sum_{i=1}^{n} a_i^2 \geq \frac{1}{n} \quad \text{as} \quad \sum_{i=1}^{n} a_i = 1.$$

N.T. with $\sum_{i=1}^{n} a_i = 1$, $\sum_{i=1}^{n} a_i^2$ attains its minimum

iff '=' holds in C-S inequality.

iff $a_i \propto 1 \ \forall i = 1(1)n$

iff $a_i = k \ \forall i = 1(1)n$

iff $a_i = \frac{1}{n} \ \forall i$ as $1 = \sum_{i=1}^{n} a_i = nk$

Hence, $T = \frac{1}{n}\sum_{i=1}^{n} X_i$ has the minimum variance among all linear UEs of μ.

$\Leftrightarrow T = \bar{X}$ is the BLUE of μ.