STATISTICAL INFERENCE II

BY

TANUJIT CHAKRABORTY

Indian Statistical Institute

Mail : tanujitisi@gmail.com

STATISTICAL INFERENCE I

OF B.Sc 3RD YEAR

Tanujit Chakraborty

THEORY OF ESTIMATION

STATISTICAL INFERENCE II

Point Estimation (Continuation) :—

- **Measure of Quality of Estimator or Properties of Good Estimator:-**

It is clear that in any given problem of estimation, we may have a large, often infinitely many estimators to choose from. Here, we shall define certain properties or measures of quality of estimators to get a good estimator:

(I) Closeness : Minimum MSE
(II) Consistency
(III) Sufficiency
(IV) Completeness.

(I) **Closeness:-** Clearly, we'd like estimator $T(\underset{\sim}{X}) = T$ to be close to θ and since T is a statistic, the usual measure of closeness $|T-\theta|$ is a R.V.

Example of such measure of closeness are:

(i) $P_\theta[|T-\theta| < \epsilon]$, for some $\epsilon > 0$.

(ii) $E_\theta |T-\theta|^r$, for some $r > 0$

Obviously, we want (i) to be large and (ii) to be small.

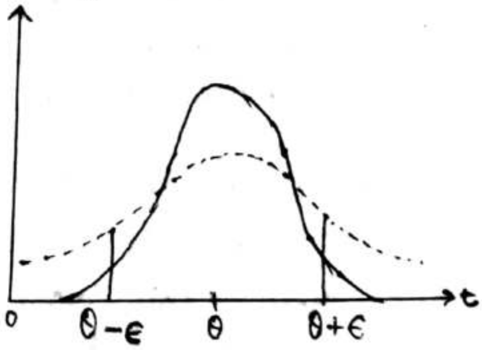
Definition: More concentrated and Most concentrated Estimators:

Let T and T^* be two estimators of θ. Then T^* is called a more concentrated estimator of θ than T iff

$$P_\theta\left[|T^* - \theta| < \epsilon\right] \geq P_\theta\left[|T-\theta| < \epsilon\right],$$

for all $\epsilon > 0$, for each $\theta \in \Omega$.

An estimator T_0 is called most concentrated estimator of θ iff it is more concentrated than any other estimator, that is iff



$$P_\theta\left[|T_0 - \theta| < \epsilon\right] \geq P_\theta\left[|T-\theta| < \epsilon\right]$$

for all T, for all $\epsilon > 0$, for each $\theta \in \Omega$.

Unfortunately, most concentrated estimators seldom exist.

Scanned by CamScanner

Mean Square Error (MSE): — A useful, though perhaps, a crude measure of closeness of an estimator T of θ is $E(T-\theta)^2$ which is obtained from (ii) by putting $r = 2$.

Notation: $MSE_\theta(T) = E\{T-\theta\}^2$

Naturally, we would prefer one with small or smallest MSE. Here, the requirement is to choose T_0 such that $MSE_\theta(T_0) \leq MSE_\theta(T)$, for all T, for each $\theta \in \Omega$.

But such estimators rarely exist.

Note that, $MSE_\theta(T) = Var(T) + \{E(T)-\theta\}^2$

Now, we shall concentrate on the class of all estimators of θ such that $\{E(T)-\theta\}^2 = 0 \Leftrightarrow E(T) = \theta \;\forall\; \theta \in \Omega$.

Now, in the class of unbiased estimators of θ, we shall find an estimator with uniformly minimum variance. This is the concept of unbiasedness and minimum variance.

Definitions:-

(1) An estimator T is said to be <u>unbiased estimator</u> of a parametric function $\psi(\theta)$ iff $E\{T\} = \psi(\theta) \;\forall\; \theta \in \Omega$.

(2) An estimator T_0 is defined to be <u>UMVUE</u> of $\psi(\theta)$ if

i) $E(T_0) = \psi(\theta) \;\forall\; \theta \in \Omega$

ii) $Var(T_0) \leq Var(T)$, for any estimator T such that $E(T) = \psi(\theta) \;\forall\; \theta \in \Omega$.

(3) A parametric function $\psi(\theta)$ is said to be <u>estimable</u> (or, unbiasedly estimable) iff there exists an estimator T such that $E(T) = \psi(\theta) \;\forall\; \theta \in \Omega$.

Unbiasedness alone does not make any sense: —

Justification: — There are situations where unbiasedness ensures poor estimation. Suppose T is an unbiased estimator of θ. Further assume that the sampling distribution of T is extremely positively skewed, i.e. θ lies on the right tail of the sampling distribution. If we regard an observed T that is an estimate to be likely then the estimate should fall close to the mode of the distribution and hence it should not be close to θ. This situation is quite natural since minimisation of MSE ensures the simultaneous minimisation of the bias and variance of the sampling distribution of the statistic.

(II) $\boxed{\text{Consistency :—}}$

Here we shall consider a large sample property of estimators. Define, $T_n = T(X_1, X_2, \ldots, X_n)$, where n indicates the sample size, as an estimator of θ. Actually, we will be considering a sequence of estimators :

$T_1 = T(X_1)$, $T_2 = T(X_1, X_2)$, ...

e.g. $T_n = \frac{1}{n}\sum_{i=1}^{n} T(X_i)$

As the sample size $n \to \infty$, the data $(x_1, x_2, \ldots, x_n)$ are practically the whole population and it is intuitively appealing to desire that a good sequence of estimators $\{T_n\}$ should be one for which values of the estimator tend to concentrate near θ as the sample size increases. If $n \to \infty$, and the values of an estimator are not very close to θ, i.e. the performance of the estimators is not good, then the performance of the estimator will be bad in case the sample size is small. Hence, for $n \to \infty$, if $\{T_n\}$ tends to concentrate near θ, then in small sample the estimator T_n may perform well and we say that the sequence $\{T_n\}$ of estimators is consistent or appropriate for θ.

Defn. :— The sequence $\{T_n\}$ of estimators is defined to be consistent sequence of estimators of θ, if, for every $\epsilon > 0$,

$$P[|T_n - \theta| < \epsilon] \to 1 \text{ as } n \to \infty, \text{ for every } \theta \in \Omega.$$

Remark:- $\{T_n\}$ is consistent for θ iff $P[|T_n - \theta| > \epsilon] \to 0$ as $n \to \infty$

$\Leftrightarrow T_n \xrightarrow{P} \theta$, for every $\theta \in \Omega$.

Ex.(1) Let $X_1, X_2, \ldots, X_n$ be a r.s. from a population with $E|X_1|^k < \infty$. Then show that m_r' is consistent for μ_r' ; $r = 1(1)k$

Solution:- [Khinchinte's WLLN:-

If $\{X_n\}$ is a sequence of iid RV's, then $\bar{X} \xrightarrow{P} \mu$, provided $\mu = E(X_1)$ exists.]

Here $X_1, X_2, \ldots, X_n$ are i.i.d. R.V.'s.

$\Rightarrow X_i^r$'s are i.i.d. RV's with $E|X_i^r| < \infty$

$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i^r = m_r' \xrightarrow{P} E(X_i^r) \;\forall\; r = 1(1)k$, by Khinchinte's WLLN.

$\Rightarrow m_r' \xrightarrow{P} \mu_r'$, $r = 1(1)k$

$\therefore m_r'$ is consistent for μ_r' , $r = 1(1)k$.

Ex.(2). If $X_1, X_2, \ldots, X_n$ be a r.s. from $N(\mu, \sigma^2)$, S.T. $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is consistent for σ^2.

Ans:- Note that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$

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

and $$\mathrm{Var}\left(\frac{(n-1)s^2}{\sigma^2}\right) = 2(n-1)$$

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

and $$\mathrm{Var}(s^2) = \frac{2\sigma^4}{n-1}.$$

For every $\epsilon > 0$,

$$0 \le P\left[|s^2 - \sigma^2| > \epsilon\right] < \frac{V(s^2)}{\epsilon^2} = \frac{2\sigma^4}{(n-1)\epsilon^2} \longrightarrow 0 \text{ as } n \to \infty$$

$$\Rightarrow \lim_{n\to\infty} P\left[|s^2 - \sigma^2| > \epsilon\right] = 0$$

Hence, s^2 is consistent for σ^2.

Remark:- If $\{T_n\}$ is consistent for θ, then

(i) $\{T_n + a_n\}$ is also consistent for θ, provided $a_n \to 0$ as $n \to \infty$

(ii) $\{b_n \cdot T_n\}$ is also consistent for θ, provided $b_n \to 1$ as $n \to \infty$.

For $\epsilon > 0$,

$$P\left[|T_n + a_n - \theta| < \epsilon\right] \simeq P\left[|T_n - \theta| < \epsilon\right], \text{ for sufficiently large } n.$$

$$\longrightarrow 1 \text{ as } n \to \infty \quad \left[\because T_n \xrightarrow{P} \theta\right]$$

Therefore, it is possible to find several consistent estimators of θ, provided there exists a consistent estimator of θ.

(iii) Concept of Consistency of an estimator:-

Consistency is a large property of an estimator. The estimator is said to be consistent if it estimates the population parameter or some other function of the parameter almost correctly even when the sample size is large.

Ex. (3):- Let $X_1, X_2, \dots, X_n$ 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, $P\left[|T_1 - \theta| < \epsilon\right] = P\left[\theta - \epsilon < T_1 < \theta + \epsilon\right]$

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

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

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

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

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

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

$$\longrightarrow 0 \text{ as } n \to \infty$$

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 : $P\left[|X| > \epsilon\right] < \frac{E|X|^r}{\epsilon^r}$, $\epsilon > 0, r > 0$]

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

$$= \frac{2}{n(n+1)} \sum_{i=1}^{n} iE(x_i).$$

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

$$= \mu$$

$Var(T_1) = Var\left\{\frac{2}{n(n+1)} \sum_{i=1}^{n} i.x_i\right\}$

$$= \frac{4}{\{n(n+1)\}^2} \sum_{i=1}^{n} i^2.\sigma^2$$

$$= \frac{4\sigma^2 n(n+1)(2n+1)}{6n^2(n+1)^2}$$

$$= \frac{2\sigma^2(2n+1)}{3n(n+1)}$$

$$\longrightarrow 0 \text{ as } n \to \infty$$

Hence, T_1 is consistent for μ.

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

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

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

$\therefore T_2$ is not consistent for μ.

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

$= \mu$

$Var(T_3) = \frac{6\sigma^2}{n(n+1)(2n+1)} \sum_{i=1}^{n} i^4$

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

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

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

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

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

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

$\therefore T_3$ 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:- $E(\bar{X}) = E(X_1) = \theta + \frac{1}{2}$

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

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

$\Rightarrow V(T_1) = \frac{1}{12n} \longrightarrow 0$ 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:-

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

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

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

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

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

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

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

Remark:- 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 $n \to \infty$.

For example:-

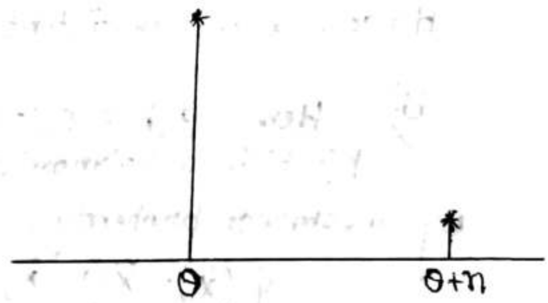
$$T_n = \begin{cases} \theta & \text{with probability } \left(1 - \frac{1}{n}\right) \\ \theta + n & \text{with probability } \frac{1}{n} \end{cases}$$

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 $n \to \infty$.

Invariance Property:– If $\{T_n\}$ is consistent for θ and $\psi(\cdot)$ is a continuous function, then $\{\psi(T_n)\}$ is consistent for $\psi(\theta)$.

Proof:– Here $\psi(\cdot)$ is continuous function. Hence for a given $\epsilon>0$. there exists a $\delta>0$, such that

$$|\psi(T_n)-\psi(\theta)|<\epsilon \text{ whenever } |T_n-\theta|<\delta.$$

Clearly, $\{|T_n-\theta|<\delta\} \subseteq \{|\psi(T_n)-\psi(\theta)|<\epsilon\}$

$$\Rightarrow P\{|T_n-\theta|<\delta\} \le P\{|\psi(T_n)-\psi(\theta)|<\epsilon\}$$

As $\{T_n\}$ is consistent for θ,

$$\therefore 1=\lim_{n\to\infty} P[|T_n-\theta|<\delta] \le \lim_{n\to\infty} P[|\psi(T_n)-\psi(\theta)|<\epsilon] \le 1$$

$$\Rightarrow \lim_{n\to\infty} P[|\psi(T_n)-\psi(\theta)|<\epsilon]=1$$

$\Rightarrow \{\psi(T_n)\}$ is consistent for $\psi(\theta)$.

Ex. 8. If $X_1, X_2, \ldots, X_n$ be a r.s. from Bernoulli distn. with prob. of success p, Show that → (i) $\bar{X}$ is consistent for p,
(ii) $\bar{X}(1-\bar{X})$ is consistent for $p(1-p)$. $= V(X_1)$.

Soln.: i) $\sum X_i \sim Bin(n,p)$

$$E(\bar{X})=E(X_1)=p$$

$$V(\bar{X})=\frac{V(X_1)}{n}=\frac{p(1-p)}{n} \longrightarrow 0 \text{ as } n\to\infty$$

Hence, $\bar{X}$ is consistent for p.

ii) Here $\psi(p)=p(1-p)=V(X_1)$ is a continuous function as $p(1-p)$ is a polynomial in p.

By invariance property,

$$\psi(\bar{X})=\bar{X}(1-\bar{X}) \text{ is consistent for } \psi(p)=p(1-p).$$

Ex. 9. Let $X_1, X_2, \ldots, X_n$ is a r.s. from $Bin(1,p)$. Suggest consistent estimators of (i) e^{-p}, (ii) p^2, (iii) $\sin p$, (iv) $-\ln p$.

Ex.(10). Let $x_1, x_2, \ldots, x_n$ be a r.s. from $N(\mu, \mu)$, $\mu > 0$.
(a) Find a consistent estimator of μ^2. Is it unbiased?
(b) Find out an UE which is consistent?

Soln :- (a) $\bar{X} \sim N\left(\mu, \frac{\mu}{n}\right)$

$$\Rightarrow E(\bar{X}) = \mu$$

$$V(\bar{X}) = \frac{\mu}{n} \to 0 \text{ as } n \to \infty.$$

Hence $\bar{X}$ is consistent for μ.
By invariance property, $\bar{X}^2$ is consistent for μ^2.

But, $E(\bar{X}^2) = V(\bar{X}) + E^2(\bar{X})$

$$= \frac{\mu}{n} + \mu^2 \neq \mu^2 \quad \left[\because x_i \overset{iid}{\sim} N(\mu, \mu) \right]$$

i.e. $\bar{X}^2$ is biased for μ^2.

(b) In a normal sample, $\bar{X}$ and s^2 are independently distributed.

Also, $E(\bar{X}) = \mu$ and $E(s^2) = \mu$.

Hence, $E(\bar{X} \cdot s^2) = E(\bar{X}) \cdot E(s^2)$, due to independence.

$$= \mu^2$$

and
$$\begin{aligned}
Var(\bar{X} \cdot s^2) &= E(\bar{X} \cdot s^2)^2 - E^2(\bar{X} \cdot s^2) \\
&= E(\bar{X}^2 \cdot s^4) - \mu^4 \\
&= E(\bar{X}^2) \cdot E(s^4) - \mu^4 \\
&= \left\{ V(\bar{X}) + E^2(\bar{X}) \right\} \cdot \left\{ V(s^2) + E^2(s^2) \right\} - \mu^4 \\
&= \left\{ \frac{\mu}{n} + \mu^2 \right\} \left\{ \frac{2\mu^2}{n-1} + \mu^2 \right\} - \mu^4 \\
&\longrightarrow 0 \text{ as } n \to \infty
\end{aligned}$$

Hence, $\bar{X} \cdot s^2$ is consistent as well as unbiased for μ^2.

Remark:- In Ex.(10) (the above example)
(a) is an example of a biased consistent estimator.
(b) is an example of an unbiased consistent estimator.

Ex. (11). Give an example of an estimator which is
(i) consistent but not unbiased,
(ii) unbiased but not consistent,
(iii) consistent as well as unbiased.

ANS:- (i) Let $T_1 = \bar{X} + \frac{1}{n}$
Clearly, $T_1 = \bar{X} + \frac{1}{n}$ is consistent but

$$E(T_1) = \mu + \frac{1}{n} \neq \mu$$

So, it is not unbiased.

[If $\{T_n\}$ is consistent for θ, the $\{T_n + a_n\}$ is consistent for θ if $\lim_{n\to\infty} a_n = 0$.]

(ii) Note that, $T = \frac{X_1 + X_n}{2}$ is an unbiased estimator of μ.

$$T \sim N(\mu, \sigma^2/2)$$

Now, $P[|T-\mu| < \epsilon] = P\left[\left|\frac{T-\mu}{\sigma/\sqrt{2}}\right| < \frac{\epsilon\sqrt{2}}{\sigma}\right]$

$$= 2\Phi\left[\frac{\epsilon\sqrt{2}}{\sigma}\right] - 1$$

$$\not\to 1 \text{ as } n \to \infty.$$

Hence, T is unbiased but not consistent for μ.

(iii) Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(\mu, \sigma^2)$
then $\bar{X} \sim N(\mu, \sigma^2/n)$.

$E(\bar{X}) = \mu$, $V(\bar{X}) = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$

$\Rightarrow \bar{X}$ is consistent as well as unbiased.

Ex. (12). Show that for a r.s. from Cauchy distribution with location parameter μ, i.e, $C(\mu, 1)$, the sample mean is not consistent for μ but the sample median is consistent for μ.

ANS:- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $C(\mu, 1)$.
Then $\bar{X} \sim C(\mu, 1)$

Now, $P[|\bar{X} - \mu| < \epsilon] = P[\mu - \epsilon < \bar{X} < \mu + \epsilon]$

$$= \int_{\mu-\epsilon}^{\mu+\epsilon} \frac{d\bar{x}}{\pi\{1 + (\bar{x}-\mu)^2\}}$$

$$= \left[\frac{1}{\pi}\tan^{-1}(\bar{x}-\mu)\right]_{\mu-\epsilon}^{\mu+\epsilon}$$

$$= \frac{2}{\pi}\tan^{-1}\epsilon \not\to 1 \text{ as } n \to \infty$$

Hence $\bar{X}$ is not consistent for μ.

It can be shown that for large samples,

$$\hat{\xi}_p \overset{a}{\sim} N\left(\xi_p\ ,\ \frac{p(1-p)}{n.f^2(\xi_p)}\right),$$

where, $f(\cdot)$ is the PDF of the distribution.

For, $C(\mu,1)$ distribution, $\xi_{1/2} \overset{a}{\sim} N\left(\xi_{1/2}, \frac{1}{4nf^2(\mu)}\right)$

$$\Rightarrow \tilde{x} \overset{a}{\sim} N\left(\mu, \frac{\pi^2}{4n}\right) \left[\because f(\mu) = \frac{1}{\pi}\right]$$

Hence, for large n, $E(\tilde{x}) = \mu$,

$$V(\tilde{x}) = \frac{\pi^2}{4n} \to 0 \text{ as } n \to \infty$$

$\Rightarrow \tilde{x}\,(\xi_{1/2})$ is consistent for μ.

Remark:- By Khinchinte's WLLN: $\bar{X} \xrightarrow{P} \mu$, provided $E(X_1) = \mu$, the population mean exists. In Cauchy population, the popl^n mean does not exist and μ is not the popl^n mean but it is the popl^n. median. Hence for μ, $\bar{X}$ is not **consistent**, but $\tilde{x}$ is consistent.

Ex. (13). Let $X_1, X_2, \ldots, X_n$ be a r.s. from the popl^n with PDF

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

Show that $X_{(1)}$ is consistent for θ.

ANS:-

$$f_{X_{(1)}}(x) = n\left[1 - \int_\theta^x e^{-(x-\theta)}dx\right]^{n-1} . e^{-(x-\theta)} \;;\; x>\theta$$

$$= n\left[1 + e^{-(x-\theta)} - 1\right]^{n-1} . e^{-(x-\theta)}$$

$$= ne^{-n(x-\theta)} \;;\; x>\theta$$

$$P[|X_{(1)} - \theta| < \epsilon] = P[\theta < X_{(1)} < \theta+\epsilon] = n\int_\theta^{\theta+\epsilon} e^{-n(x-\theta)}dx$$

$$= ne^{n\theta}\left[\frac{e^{-nx}}{-n}\right]_\theta^{\theta+\epsilon}$$

$$= 1 - e^{-n\epsilon}$$

$$\longrightarrow 1 \text{ as } n \to \infty$$

$\therefore X_{(1)}$ is consistent for θ.

Ex. (14). If $X_1, \ldots, X_n$ be a r.s. from $f(x) = \frac{1}{2}(1+\theta x)$; $-1<x<1$, $-1<\theta<1$. Find a consistent estimator of θ.

(ISI)

Solution: — $f(x) = \frac{1}{2}(1+\theta x) \, I_{-1<x<1}$

$$\therefore E(X) = \frac{1}{2}\int_{-1}^{1}(1+\theta x)x \, dx = \frac{\theta}{3}.$$

Now, $E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \theta/3$

$$\Rightarrow E(3\bar{X}) = \theta$$

Now, $E(X^2) = \frac{1}{2}\int_{-1}^{1} x^2(1+\theta x)\, dx = \frac{1}{2}\int_{-1}^{1}(x^2+\theta x^3)\, dx = \frac{1}{3}$

$$\therefore V(X) = E(X^2) - E^2(X)$$

$$\Rightarrow V(X) = \frac{1}{3} - \frac{\theta^2}{9}$$

$$V(\bar{X}) = \frac{1}{n^2}\cdot n\left(\frac{1}{3} - \frac{\theta^2}{9}\right) = \frac{1}{n}\left(\frac{1}{3} - \frac{\theta^2}{9}\right)$$

$$\therefore \lim_{n\to\infty} V(3\bar{X}) = 9 \lim_{n\to\infty} V(\bar{X}) = 9 \lim_{n\to\infty} \frac{1}{n}\left(\frac{1}{3} - \frac{\theta^2}{9}\right) = 0$$

$\therefore 3\bar{X}$ is a consistent estimator of θ.

Ex. (15).

(III) $\boxed{\text{SUFFICIENCY}}$:-

Introduction:- In The problem of statistical inference, the raw data collected from the field of enquiry is too numerous and hence too difficult to deal with and too costly to store. So, a statistician would like to condence the data by determining a function of the sample observation, i.e. by forming a statistic. Here, the condensation should be done in a manner so that there is 'no loss of information' regarding the popln. feature of interest. The statistic which exhaust all the relevant information about the labelling parameter, that contained in the sample are called sufficient statistics and these notion is termed as sufficiency principle. Clearly, sufficiency is an essential criterion of an inferential problem.

Consider the following example :

Let $x_1, x_2, \ldots, x_n$ be a r.s. from $N(\mu, 1)$, μ is unknown.

Apply the orthogonal transformation

$$\underset{\sim}{Y} = A\underset{\sim}{X} \text{ with } \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right) \text{ as the first row of } A.$$

Then $Y_1 = \sqrt{n}\bar{X} \sim N(\sqrt{n}\mu, 1)$

and $Y_i \sim N(0,1)$, $i = 2(1)n$, independently.

To estimate μ, we can use $(X_1, X_2, \ldots, X_n)$ or $Y_1 = \sqrt{n}\bar{X}$, since $Y_2, Y_3, \ldots, Y_n$ provide no information about μ.

Clearly, $Y_1 = \sqrt{n}\bar{X}$ is preferable, since we need not to keep the record of all observations.

Any estimation of the parameter based on $Y_1 = \sqrt{n}\bar{X}$ is just effective as any estimation that could be based on $x_1, x_2, \ldots, x_n$. If we use statistics to extract all the information in the sample about μ then it is sufficient or enough to observe only Y_1.

Let $x_1, \ldots, x_n$ be a random sample from popln. with PDF or PMF $f(x; \theta)$. Following Fisher, we call T a sufficient (or an exhaustive) statistic if it contains all the information about θ that is contained in the sample.

Definition 1. Sufficient statistic

Let $(X_1, X_2, \ldots, X_n)$ be a random sample drawn from F_θ. A statistic $S = S(X_1, X_2, \ldots, X_n)$ is said to be a sufficient statistic of θ iff $P_\theta[\underset{\sim}{X} \in A \mid S = s]$ is independent of θ $\forall\, \theta \in \Omega$ and for all A, i.e. the conditional distribution of $(X_1, X_2, \ldots, X_n)$ given $S = s$ does not depend on θ, for any values s of S.

Remark:- The definition says that a statistic S is sufficient if you know the values of the statistic S, then the sample values themselves are not needed and can tell you nothing more about θ.

1. Illustrative Example :- Let $(X_1, \ldots, X_n)$ be a r.s. from $Bin(1, p)$, show that, using definition, $S = \sum_{i=1}^{n} X_i$ is sufficient for p.

Soln. → [Suppose, we are given a loaded coin and asked to infer about p, the probability of head.

To carry out the inference, the coin is tossed n times and the S-F (success - failure) run has been recorded. Let the records be $x_1, x_2, \ldots, x_n$; where x_i is a realisation on X_i. It is evident that X_i's are independent of each other. To infer about p, it is not necessary to know which trial results in success whereas it is sufficient to know the number of success, i.e. $\sum_{i=1}^{n} X_i$. Now, we show that this goes consistent with the definition.]

Let $X_1, X_2, \ldots, X_n$ be a r.s. from $Bin(1, p)$, where p being the probability of success.

Let us define, $S = \sum_{i=1}^{n} X_i$

Now, we need to show S is sufficient.

Let us consider the conditional distribution of the r.s. given that the distn. of the statistic.

$$P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid S = s]$$

$$= \frac{P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n, S = s]}{P[S = s]}$$

$$= \begin{cases} \dfrac{P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n]}{P[\sum_{i=1}^{n} X_i = s]}, & \text{if } s = \sum_{i=1}^{n} x_i, \\ 0, & \text{ow} \end{cases}$$

$$= \begin{cases} \dfrac{p^{\sum x_i}(1-p)^{n-\sum x_i}}{\binom{n}{s} p^s (1-p)^{n-s}}, & \text{if } s = \sum x_i, \text{ where } x_i = 0 \text{ or } 1\ \forall i = 1(1)n. \\ 0, & \text{ow} \end{cases}$$

$$= \begin{cases} \dfrac{1}{\binom{n}{s}} & \text{if } s = \sum_{i=1}^{n} x_i \\ 0 & \text{ow} \end{cases}$$

Hence, the conditional distribution is independent of p.

$\therefore$ By definition, $S = \sum_{i=1}^{n} X_i$ is sufficient for p.

Note:- The random sample itself $T = (X_1, \ldots, X_n)$ is trivially a sufficient statistic.

Remark:- Definition (1) is not a constructive definition, since it requires that we first guess a statistic T and then check to see whether T is sufficient or not, it does not provide any clue to what the choice of T should be.

Definition 2. Let $X_1, X_2, \ldots, X_n$ be a r.s. from the PMF or PDF $f(x; \theta)$. A statistic S is defined to be a sufficient statistic iff the conditional distribution of T given $S = s$ does not depend on θ, for any statistic T, for any value of s.

This definition in particularly is useful to show that a statistic S is not sufficient.

Definition:- Joint sufficient statistic

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from the density f_θ. The statistics $T_1, T_2, \ldots, T_n$ are defined to be jointly sufficient if the conditional distribution of $X_1, X_2, \ldots, X_n$ given $S_1 = s_1, S_2 = s_2, \ldots, S_n = s_n$ is independent of the unknown parameter θ.

Remark:- If $(X_1, X_2, \ldots, X_n)$ is ordered then the order statistics $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ will also be sufficient, since $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ is nothing but $n!$ permutations of $(X_1, X_2, \ldots, X_n)$. Hence if we consider the conditional distribution of $(X_1, X_2, \ldots, X_n)$ given $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ will be $\frac{1}{n!}$, which is independent of θ.

Another approach of showing $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ as a sufficient statistic is factorization theorem.

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.

Ex. (3). Let $x_1, x_2, \ldots, x_n$ be a r.s. from $P(\lambda)$. S.T. $S = \sum_{i=1}^{n} x_i$ is sufficient for λ.

Ans:-

Ex. (4). Let (X_1, X_2) be a r.s. from $P(\lambda)$, S.T. $T = X_1 + 2X_2$ is not sufficient for λ.

ANS:-

$$P\left[X_1 = 0,\ X_2 = 1 \mid T = 2\right] = \frac{P\left[X_1 = 0,\ X_2 = 1\right]}{P\left[X_1 + 2X_2 = 2\right]}$$

$$= \frac{e^{-\lambda}\left(\lambda e^{-\lambda}\right)}{P\left[X_1 = 0,\ X_2 = 1\right] + P\left[X_1 = 2,\ X_2 = 0\right]}$$

$$= \frac{\lambda e^{-2\lambda}}{\lambda e^{-2\lambda} + \left(\frac{\lambda^2}{2}\right)e^{-2\lambda}}$$

$$= \frac{1}{\left(1 + \frac{\lambda}{2}\right)}, \quad \text{dependent on } \lambda.$$

This depends on λ.

So, T is not sufficient.

Ex. (5). Let $(X_1, \ldots, X_n)$ be a r.s. from Geo (p). Find the conditional distribution of $(X_1, X_2, \ldots, X_n)$ given $\sum_{i=1}^{n} X_i = s$. Hence comment on $\sum X_i$ as an estimator of p.

Solution:- As $X_i \overset{iid}{\sim}$ Geometric (p), $i = 1(1)n$.

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

Now, $P\left[X_1 = x_1, \ldots, X_n = x_n \,\middle|\, \sum_{i=1}^{n} X_i = s\right]$

$$= \frac{P\left[X_1 = x_1, \ldots, X_n = x_n ; \sum_{i=1}^{n} X_i = s\right]}{P\left[\sum_{i=1}^{n} X_i = s\right]}$$

$$= \begin{cases} \dfrac{P[X_1 = x_1, \ldots, X_n = x_n]}{P\left[\sum_{i=1}^{n} X_i = s\right]} & ; \text{ if } s = \sum_{i=1}^{n} x_i \\ 0 & ; \text{ ow} \end{cases}$$

$$= \begin{cases} \dfrac{\prod_{i=1}^{n} \left\{p(1-p)^{x_i}\right\}}{\binom{s+n-1}{s} p^n q^s} & ; \text{ if } s = \sum_{i=1}^{n} x_i \\ 0 & ; \text{ ow} \end{cases}$$

$$= \begin{cases} \dfrac{1}{\binom{s+n-1}{s}} & ; \text{ if } s = \sum_{i=1}^{n} x_i \\ 0 & \text{ ow} \end{cases}$$

, which is independent of p.

Hence, by definition, the statistic $\sum_{i=1}^{n} X_i$ is sufficient for p.

Ex. (6). Let $(X_1, X_2, \ldots, X_n)$ be a r.s. from the p.m.f.

$$P(x; N) = \begin{cases} \dfrac{1}{N} & , x = 1(1)n \\ 0 & , \text{ow} \end{cases}$$

Find the conditional distribution of $(X_1, X_2, \ldots, X_n)$ given $X_{(n)} = s$. Hence comment on $X_{(n)}$ as an estimator of N.

Remark:- Let $f(\underset{\sim}{x};\theta)$ be the PMF or PDF of $\underset{\sim}{X}=(X_1,\dots,X_n)$ and $g(t;\theta)$ be the PMF or PDF of the statistic $T(\underset{\sim}{X})$.

For discrete case, $P[\underset{\sim}{X}=\underset{\sim}{x} \mid T(\underset{\sim}{X})=t]$

$$= \frac{P[\underset{\sim}{X}=\underset{\sim}{x} \,;\, T(\underset{\sim}{X})=t]}{P[T(\underset{\sim}{X})=t]}$$

$$= \begin{cases} \dfrac{P[\underset{\sim}{X}=\underset{\sim}{x}]}{P[T(\underset{\sim}{X})=t]} & \text{if } t=T(\underset{\sim}{x}) \\ 0 & \text{ow} \end{cases}$$

$$= \begin{cases} \dfrac{f(\underset{\sim}{x};\theta)}{g(t;\theta)} & \text{if } t=T(\underset{\sim}{x}) \\ 0 & \text{ow} \end{cases}$$

If $P[\underset{\sim}{X}=\underset{\sim}{x} \mid T(\underset{\sim}{X})=t] = \dfrac{f(\underset{\sim}{x};\theta)}{g(t;\theta)}$ is independent of θ, then $T(\underset{\sim}{X})$ is sufficient for θ.

In general, we have for continuous & discrete distribution, if the ratio $\dfrac{f(\underset{\sim}{x};\theta)}{g(t;\theta)}$ is independent of θ, then $T(\underset{\sim}{X})$ is sufficient for θ.

<u>Ex. (7).</u> Let $X_1, X_2, \dots, X_n$ be a r.s. from $N(\mu,1)$. S.T. using defn., $\overline{X}$ is sufficient for μ.

<u>ANS:-</u> The PDF of $\underset{\sim}{X}=(X_1,X_2,\dots,X_n)$ is

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

and the PDF of $\overline{X}$ is

$$g(\bar{x};\mu)=\left(\frac{1}{\sqrt{\frac{2\pi}{n}}}\right)\cdot e^{-\frac{n}{2}(\bar{x}-\mu)^2} \quad ; \bar{x}\in\mathbb{R} \quad \left[\text{Here } \bar{x}\sim N\left(\mu,\tfrac{1}{n}\right)\right]$$

$\therefore$ The ratio

$$\frac{f(\underset{\sim}{x};\mu)}{g(\bar{x};\mu)}=\frac{\sqrt{n}}{(2\pi)^{\frac{n+1}{2}}}\, e^{-\frac{1}{2}\left\{\sum(x_i-\mu)^2-n(\bar{x}-\mu)^2\right\}}$$

$$=\frac{\sqrt{n}}{(2\pi)^{\frac{n+1}{2}}}\, e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad ; \left[\because \sum(x_i-\mu)^2 = \sum(x_i-\bar{x})^2+n(\bar{x}-\mu)^2\right]$$

which is independent of μ.

Hence, by definition, $\overline{X}$ is sufficient for μ.

Ex.(8). Let $(X_1, \ldots, X_n)$ be a r.s. from $U(0,\theta)$, $\theta > 0$; S.T. $X_{(n)}$ is sufficient for θ.

Soln.:- $X_{(n)}$ is sufficient for θ if the conditional distribution of $\underset{\sim}{X}$ given $X_{(n)} = x_{(n)}$ is independent of θ, i.e. if the ratio $\dfrac{f(\underset{\sim}{x};\theta)}{g(x_{(n)};\theta)}$ is independent of θ.

for $0 < x_i < \theta$, and $0 < X_{(n)} < \theta$;

$$\frac{f(\underset{\sim}{x};\theta)}{g(x_{(n)};\theta)} = \frac{\left(\frac{1}{\theta}\right)^n}{\frac{n\{X_{(n)}\}^{n-1}}{\theta^n}} \quad ; \text{if } 0 < x_{(n)} < \theta$$

$$= \frac{1}{n\{x_{(n)}\}^{n-1}} \quad ; \text{ if } 0 < x_{(n)} < \theta$$

which is independent of θ.

Hence $X_{(n)}$ is sufficient for θ.

— x —

Note:- Definition (I) :- $P[\underset{\sim}{X} = \underset{\sim}{x} \mid S = s]$ is independent of θ.

Definition (II) :- $P[T = t \mid S = s]$ is independent of θ.

Defn. (II) is useful to show that a statistic S is not sufficient since from the idea of sampling distribution, it is known that $P[T = t \mid S = s]$ does not depend on θ.

Factorization Criterion (Due to Fisher) :

The requirement for Factorization theorem: For a given family of distribution if we are to find a sufficient statistic for the labelling parameter, it will be difficult to adopt the definition of sufficiency as a criterion in choosing a sufficient statistic. Because according to the definition of sufficient statistic $P[X \in A | T = t]$ (where, A being a function of t), are not uniquely defined and the question arises whether determinations exist or not for some fixed t. The answer is that is is possible when the sample space is euclidean.

Secondly, the determination of sufficient statistic by means of its definition is inconvenient since it requires, first guessing a statistic T that might be sufficient and then checking whether the conditional distribution of X given $T=t$ is independent of θ or not.

Therefore, we need a simpler criterion which can be adopted as a tool to find a sufficient statistic. Such a criterion is given in terms of factorization theorem due to Fisher and Neyman.

Theorem: Factorization criterion : We now give a criterion for determining sufficient statistics:

Statement:- Let $(X_1, X_2, X_3, \cdots, X_n) = \underset{\sim}{X}$ be a r.s. from PMF or PDF $f(x;\theta)$ $\forall$ $\theta \in \Omega$. Then $T(\underset{\sim}{X})$ is sufficient for θ iff we can factor the PMF or PDF of $\underset{\sim}{X}$ as

$$\prod_{i=1}^{n} f(x_i;\theta) = g(T(\underset{\sim}{x}),\theta)\, h(\underset{\sim}{x}) \quad \text{------------}(*)$$

where, $h(\underset{\sim}{x})$ depends on $\underset{\sim}{x}$ but not on θ and $g(T(\underset{\sim}{x}),\theta)$ depends on θ and on $\underset{\sim}{x}$ only through $T(\underset{\sim}{x})$.

Proof:- [Discrete case only]

Only if (Necessary) Part: — Let, $T(\underset{\sim}{X})$ is sufficient for θ.

Then, $P[\underset{\sim}{X} = \underset{\sim}{x} | T(\underset{\sim}{X}) = t]$ is independent of θ and

$$P_\theta[\underset{\sim}{X} = \underset{\sim}{x}] = P_\theta[\underset{\sim}{X} = \underset{\sim}{x}; T(\underset{\sim}{X}) = t] \quad \text{if} \quad t = T(\underset{\sim}{x})$$

$$= P_\theta[T(\underset{\sim}{X}) = t]\, P[\underset{\sim}{X} = \underset{\sim}{x} | T(\underset{\sim}{X}) = t] \quad \text{if} \quad T(\underset{\sim}{x}) = t$$

For values of $\underset{\sim}{x}$ for which $P_\theta[\underset{\sim}{X} = \underset{\sim}{x}] = 0$ $\forall$ $\theta \in \Omega$.
Let us define, $h(\underset{\sim}{x}) = 0$ and for $\underset{\sim}{x}$ for which $P_\theta[\underset{\sim}{X} = \underset{\sim}{x}] > 0$, for some θ. We define, $h(\underset{\sim}{x}) = P[\underset{\sim}{X} = \underset{\sim}{x} | T(\underset{\sim}{X}) = t]$ and

$$g(T(\underset{\sim}{x});\theta) = P_\theta[T(\underset{\sim}{X}) = t]$$

Thus we see that (*) holds.

If (sufficient) Part :- Let the factorization criterion (*) holds. Then, for fixed t, we have

$$P_\theta[T(\underset{\sim}{X})=t]$$
$$= \sum_{\{\underset{\sim}{x}: T(\underset{\sim}{x})=t\}} P_\theta[\underset{\sim}{X}=\underset{\sim}{x}]$$
$$= \sum_{\{\underset{\sim}{x}: T(\underset{\sim}{x})=t\}} g(T(\underset{\sim}{x});\theta)\cdot h(\underset{\sim}{x})$$
$$= g(t,\theta) \sum_{\{\underset{\sim}{x}: T(\underset{\sim}{x})=t\}} h(\underset{\sim}{x})$$

Suppose that $P_\theta[T(\underset{\sim}{X})=t] > 0$ for some θ.

Then, $P_\theta[\underset{\sim}{X}=\underset{\sim}{x} \mid T(\underset{\sim}{X})=t]$

$$= \frac{P_\theta[\underset{\sim}{X}=\underset{\sim}{x};\ T(\underset{\sim}{X})=t]}{P_\theta[T(\underset{\sim}{X})=t]}$$

$$= \begin{cases} \dfrac{P_\theta[\underset{\sim}{X}=\underset{\sim}{x}]}{P_\theta[T(\underset{\sim}{X})=t]} & \text{if } t=T(\underset{\sim}{x}) \\ 0 & \text{if } t\neq T(\underset{\sim}{x}) \end{cases}$$

$$= \begin{cases} \dfrac{g(T(\underset{\sim}{x}),\theta)\,h(\underset{\sim}{x})}{g(t,\theta)\sum_{\{\underset{\sim}{x}: T(\underset{\sim}{x})=t\}} h(\underset{\sim}{x})} & \text{if } t=T(\underset{\sim}{x}) \\ 0 & \text{ow} \end{cases}$$

$$= \begin{cases} \dfrac{h(\underset{\sim}{x})}{\sum_{\{\underset{\sim}{x}: T(\underset{\sim}{x})=t\}} h(\underset{\sim}{x})} & \text{if } t=T(\underset{\sim}{x}) \\ 0 & \text{ow} \end{cases}$$

, which is independent of θ.

Hence $T(\underset{\sim}{X})$ is sufficient statistic for θ.

Remark :- 1. The factorization criterion can't be used to show that a given statistic T is not sufficient. To do this one would normally have to use the definition of sufficiency.

2. If $T(X)$ is sufficient for $\{F_\theta : \theta \in \Theta\}$, then T is sufficient for $\{F_\theta : \theta \in W\}$, where $W \subseteq \Theta$. This follows trivially from the definition.

Scanned by CamScanner

Result:- 1) If T is sufficient for θ, then any one-to-one function of T is also sufficient for θ, i.e. the bijection of T is also a sufficient statistic for θ.

Proof:- Let $U = \phi(T)$ is a one-to-one function, then $T = \phi^{-1}(U)$ exists.

Now,
$$\prod_{i=1}^{n} f(x_i;\theta) = g(t;\theta)\, h(\underset{\sim}{x})$$
$$= g(\phi^{-1}(u);\theta)\, h(\underset{\sim}{x})$$
$$= g^*(u,\theta)\cdot h(\underset{\sim}{x})$$

By factorization criterion, it is sufficient for θ.

2) If T_1, T_2 be two different sufficient statistics, then they are related.

Proof:-
$$\prod_{i=1}^{n} f(x_i;\theta) = g_1(t_1,\theta)\, h_1(\underset{\sim}{x})$$
$$= g_2(t_2,\theta)\, h_2(\underset{\sim}{x})$$

$\Rightarrow \dfrac{g_1(t_1,\theta)}{g_2(t_2,\theta)} = \dfrac{h_2(\underset{\sim}{x})}{h_1(\underset{\sim}{x})}$, which is independent of θ,

$\Rightarrow \psi(t_1,t_2) = h^*(\underset{\sim}{x})$

$\Rightarrow T_1$ and T_2 are related.

It does not follow that every function of a sufficient statistic is sufficient.

If T_1 is sufficient then $T_2 = f(T_1)$ is sufficient if f is one-to-one; otherwise, T_2 may be <u>or</u> may not be sufficient.

3) For a r.s. $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ from the PMF or PDF $f(x;\theta)$, the entire sample $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ is sufficient for θ. Also the order statistics $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ is sufficient for θ.

Proof:- The PMF <u>or</u> PDF of $\underset{\sim}{X}$ is
$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

Note that,
$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1,x_2,\ldots,x_n;\theta) = n!\, f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta)$$
$$\Rightarrow f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta) = \frac{1}{n!} f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1,\ldots,x_n;\theta)$$
$$= g(T(\underset{\sim}{x}),\theta)\, h(\underset{\sim}{x})$$

where $h(\underset{\sim}{x}) = \dfrac{1}{n!}$ and $T(\underset{\sim}{X}) = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$

By factorization criterion, $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ is sufficient for θ.

Note:- [Concept of sufficiency implies — entire sample's sufficiency = sufficiency of order statistic ; Property of data summarization implies — order statistic is more preferable than entire sample's sufficiency.]

According to the concept of sufficiency as space reduction both $(X_1, X_2, \ldots, X_n)$ and $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ are in the same position and both the statistics are known as trivial sufficient statistics. According to the concept of data summarisation as a property of statistic, the ordered statistics are preferable than the original samples. for, instead of collecting $n!$ original samples, we may collect only the order statistics.

Remark:- Any statistic $T(\underset{\sim}{x})$ defines a form of data reduction or data summary. An experimental who uses only the observed value of the statistic $T(\underset{\sim}{x})$ rather than the entire observed sample $\underset{\sim}{x}$, will treat as $\underset{\sim}{x}$ and $\underset{\sim}{y}$ that satisfy $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$, even though the actual sample values may be different. Data reduction in terms of a particular statistic can be thought of as the partition of the sample-space $\mathscr{X}$. Note that $T(\underset{\sim}{x})$ describes a mapping $T : \mathscr{X} \to \mathcal{T}$, where $\mathcal{T} = \{t : t = T(\underset{\sim}{x}),\ \underset{\sim}{x} \in \mathscr{X}\}$, then $T(\underset{\sim}{x})$ partitions the sample space into sets $A_t : t \in \mathcal{T}$ defined $A_t = \{\underset{\sim}{x} : T(\underset{\sim}{x}) = t\}$ the statistic summarises the data in that rather than reporting all the samples $\underset{\sim}{x}$, it reports only $T(\underset{\sim}{x}) = t$. The sufficiency principle promotes a method of data reduction that does not discard information about θ while achieving some summarization of data.

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

Let $(x_1, x_2, \ldots, x_n)$ be a r.s. from $P(\lambda)$.

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

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

where $h(\underset{\sim}{x}) = \frac{1}{\prod_{i=1}^{n} \lfloor x_i}$ 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) $\underset{\sim}{T_1} = (x_1, x_2, \ldots, x_n)$ is sufficient for λ, as

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

(ii) $\underset{\sim}{T_2} = (x_1, \ldots, x_{n-2}, x_{n-1} + x_n)$ is sufficient for λ, as

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

$\vdots$

(iii) $\underset{\sim}{T_{n-1}} = (x_1, x_2 + x_3 + \cdots + x_n)$ 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 $(x_1, x_2, \ldots, x_n)$ 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

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

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

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

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 :-

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

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

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

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

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

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. (4). 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 $\bar{x}$.

Scanned by CamScanner

Ex.(5). Uniform Distribution :—

Let $X_1, X_2, \ldots, X_n$ be a r.s. from ~~[illegible]~~ $U(0,\theta), \theta>0$. 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

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

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

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

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

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

$X_{(n)} = \left\{ \max_{1 \le i \le n} X_i \right\}$.

$\therefore$ 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

$$\prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} \frac{1}{(\theta_2-\theta_1)^n} & \text{if } \theta_1 \le x_i \le \theta_2 \ \forall\ i=1(1)n \\ 0 & \text{ow} \end{cases}$$

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

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

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

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 $X_1, X_2, \dots, X_n$ 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 $(X_1, \dots, X_n)$ 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 $\{f_i(x);\ i=1,2,\dots\}$.
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,

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

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, $k = 0,1,2,\ldots\ldots$

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

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

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

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

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

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

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

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 $0<x<1$
$\alpha,\beta>0$

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

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

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

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

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

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

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

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

is sufficient for μ and σ.

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

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

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

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

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

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

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

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

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

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

__Soln:__ $f(\underset{\sim}{x}) = \left\{\dfrac{\alpha^p}{\Gamma p}\right\}^n \cdot \exp\left[-\alpha \sum_i x_i\right] \cdot \left(\prod_i x_i\right)^{p-1}$

$= g(T(\underset{\sim}{x}); \alpha, p) \cdot h(\underset{\sim}{x})$; 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:__ $f(\underset{\sim}{x}) = \dfrac{1}{\theta^n} \cdot \exp\left\{-\dfrac{1}{\theta} \sum_{i=1}^{n} x_i\right\}$

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

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

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

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

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

$\therefore T(\underline{X}) = (X_{(1)}, X_{(n)})$ 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 $\prod_{i=1}^{n} \{1+(x_i-\theta)^2\}$

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

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

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

Clearly, $\prod_{i=1}^{n} f(x_i;\theta)$ can not be written as $g(T(\underline{x}), \theta) \cdot h(\underline{x})$ 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 $\{1, 2, \ldots, N\}$, 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 (16). [Theoretical Exercises]

(i) Let $X_1, X_2, \ldots, X_n$ be a r.s. from a discrete distribution. Is the statistic $T = (X_1, \ldots, X_{n-1})$ sufficient?

(ii) Let X_1, X_2 be a RY from $P(\lambda)$. S.T. the statistic $X_1 + \lambda X_2$ $(\lambda > 1)$, λ is an integer, is not sufficient for λ.

(iii) Let $X_1, \ldots, X_n$ be a r.s. from $N(\theta, 1)$. S.T. $\bar{X}$ is sufficient for θ but $\bar{X}^2$ is not. Is $\bar{X}$ sufficient for θ^2?

(iv) Let X be a single observation from $N(0, \sigma^2)$. Is X sufficient for σ? Are $|X|$, X^2, $e^{|X|}$ sufficient for σ?

Ex.(17). Let $X_1, X_2, \ldots, X_n$ be a r.s. from

$$f(x;\mu,\sigma) = \frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}} \; ; x \in \mathbb{R} \; ; \mu \in \mathbb{R}, \sigma > 0.$$

Find a sufficient statistic for
(i) σ when μ is known; (ii) μ when σ is known;
(iii) (μ, σ).

Solution:- $$\prod_{i=1}^{n} f(x_i;\mu,\sigma) = \left(\frac{1}{2\sigma}\right)^n \cdot e^{-\frac{\sum_{i=1}^{n}|x_i-\mu|}{\sigma}} \; ; x_i \in \mathbb{R}$$

(i) μ-known:-

$$\prod_{i=1}^{n} f(x_i;\sigma) = \left(\frac{1}{2\sigma}\right)^n \cdot e^{-\frac{\sum|x_i-\mu|}{\sigma}}$$

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

$$\text{\& } T(\underset{\sim}{x}) = \sum_{i=1}^{n}|x_i-\mu|$$

$\therefore \sum_{i=1}^{n}|x_i-\mu|$ is sufficient for σ.

(ii) σ-known:-

$$\prod_{i=1}^{n} f(x_i;\mu) = \left(\frac{1}{2\sigma}\right)^n \cdot e^{-\frac{\sum_{i=1}^{n}|x_{(i)}-\mu|}{\sigma}}$$

Note that, $\sum_{i=1}^{n}|x_i-\mu| = |x_1-\mu| + |x_2-\mu| + \cdots + |x_n-\mu|$
can't be simplified as μ is not known.
So, $(X_1, \ldots, X_n)$ or $(X_{(1)}, \ldots, X_{(n)})$ is sufficient but there is no other sufficient statistic.

(iii)

Ex. (18).

(a) Let $X_1, \ldots, X_n$ be independently distributed RV's with densities

$$f(x_i;\theta) = \begin{cases} e^{i\theta - x_i} & , \text{if } x_i \geq i\theta \\ 0 & , \text{ow} \end{cases}$$

(Here X_i's are not random samples)

Find a one-dimensional sufficient statistic for θ. [ISI]

(b) Let $X_1, \ldots, X_n$ be independently distributed RV's with PDFs

$$f(x_i;\theta) = \begin{cases} \dfrac{1}{2i\theta} & ; -i(\theta-1) \leq x_i \leq i(\theta+1) \\ 0 & ; \text{ow} \end{cases}$$

Find a two-dimensional sufficient statistic for θ. Also, find a one-dimensional sufficient statistic, if exists.

Solution:-

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

$$\prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} e^{\theta\sum_{i=1}^{n} i - \sum_{i=1}^{n} x_i} & ; \text{if } x_i \geq i\theta \ \forall i = 1(1)n \\ 0 & ; \text{ow} \end{cases}$$

$$= \begin{cases} e^{\frac{n(n+1)\theta}{2} - \sum_{i=1}^{n} x_i} & ; \text{if } \frac{x_i}{i} \geq \theta \ \forall i = 1(1)n \\ 0 & ; \text{ow} \end{cases}$$

$$= \begin{cases} e^{\frac{n(n+1)\theta}{2} - \sum_{i=1}^{n} x_i} & , \text{if } \min_i \left\{\frac{x_i}{i}\right\} \geq \theta \\ 0 & , \text{ow} \end{cases}$$

$$= e^{\frac{n(n+1)\theta}{2} - \sum_{i=1}^{n} x_i} \cdot I\left(\theta, \min\left\{\frac{x_i}{i}\right\}\right) \text{ ; where}$$

$$= e^{\frac{n(n+1)\theta}{2}} \cdot I\left(\theta, \min\left\{\frac{x_i}{i}\right\}\right) \cdot e^{-\sum_{i=1}^{n} x_i} ; \quad I(a,b) = \begin{cases} 1 & , a \leq b \\ 0 & \text{ow} \end{cases}$$

$$= g(T(\underset{\sim}{x});\theta) \cdot h(\underset{\sim}{x}) \text{ ; where } h(\underset{\sim}{x}) = e^{-\sum_{i=1}^{n} x_i};$$

and $T(\underset{\sim}{x}) = \min_i \left\{\frac{x_i}{i}\right\}$ is sufficient for θ, by factorization criterion.

(ii) Hints:-

$$(\theta - 1) \leq \frac{x_i}{i} \leq (\theta + 1)$$

$$\& \ Y_i = \frac{X_i}{i} \sim U(-\theta+1, \theta+1)$$

$$Y_i - 1 \sim U(-\theta, \theta).$$

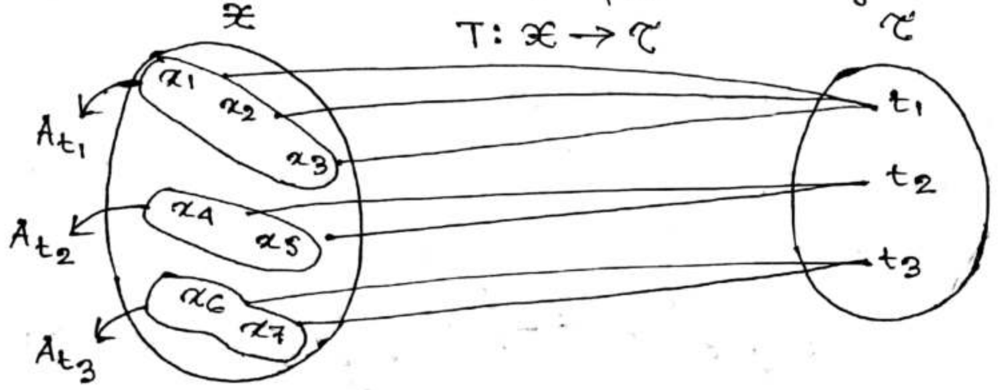
$$T_1 = \left(\min_i \left\{\frac{x_i}{i}\right\}, \max_i \left\{\frac{x_i}{i}\right\}\right)$$

$$T_2 = \max_i \left\{ \left| \frac{x_i}{i} - 1 \right| \right\}.$$

Remark:- <u>Data summarization And Sufficiency :</u>

Any statistic $T(\underset{\sim}{X})$ defines a form of data reduction or data summary. An experimental who uses only the observed value of the statistic rather than the observed sample. We will treat as equal to two sample $\underset{\sim}{x}$ and $\underset{\sim}{y}$ that satisfy $T(\underset{\sim}{X})=T(\underset{\sim}{Y})$, even though the actual samples may be different. The data reduction in terms of a particular statistic can be thought of as the partition of the sample space $\mathscr{X}$.

Note that $T(\underset{\sim}{X})$ describes a mapping $T: \mathscr{X} \rightarrow \mathcal{T}$, where $\mathcal{T} = \{t : t = T(\underset{\sim}{x}),\ \underset{\sim}{x} \in \mathscr{X}\}$ and $T(\underset{\sim}{X})$ partitions the sample space $\mathscr{X}$ into the set $A_t = \{\underset{\sim}{x} : T(\underset{\sim}{x}) = t\}$.



The statistic summarises the data, it reports only $T(\underset{\sim}{x}) = t$ rather than reporting all the samples $\underset{\sim}{x}_i$'s for which $T(\underset{\sim}{x}_i) = t$.

<u>The sufficiency principle promotes a method of data summarization that does not discard any information about θ (the parameter) while achiving some summarization of the data.</u>

['<u>Sufficiency</u>' implies —
(<u>Data summarization + 100% information carries out</u>, i.e. no loss of information)

Whenever '<u>statistic</u>' just <u>summarises the data, there may be some loss of information.</u>]

Note that, $T_1 = (X_1, \ldots, X_n)$ are $T_2 = (X_{(1)}, \ldots, X_{(n)})$ are both sufficient statistics. But instead of collecting $n!$ original samples we can collect only order statistics. According to the concept of data summarization, the order statistics are more preferable than the original samples.

Minimal Sufficient Statistic: Since the objective for looking for a sufficient statistic is to condense the data without loosing any information. One should always be on the look out for that sufficient statistic which results in the greatest reduction of the data and such a statistic is called minimal sufficient statistic.

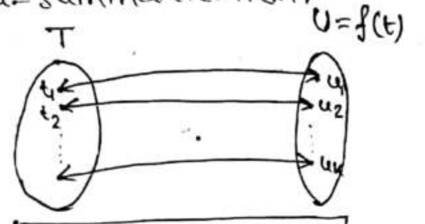
Definition:- A statistic T is called a minimal sufficient statistic for θ, provided

(i) T is sufficient for θ.

(ii) T is a function of every sufficient statistic.

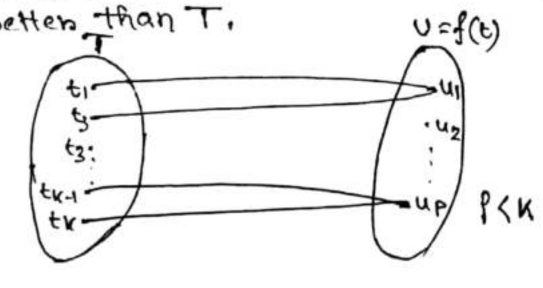
Remark:- If T and U are two sufficient statistics and $U = f(T)$. Which one is better?

$\Rightarrow$ If $f(\cdot)$ is one-to-one then T and $U = f(T)$ are equivalent with respect to data-summarization.



$\boxed{f(\cdot) \text{ is one-to-one}}$

If $f(\cdot)$ is not one-to-one, then U reduces the space most than T and so U is better than T.



$\boxed{f(\cdot) \text{ is not one-to-one}}$

Theorem:- For two points $\underset{\sim}{x}$ and $\underset{\sim}{y}$ in the sample space, the ratio $\dfrac{f(\underset{\sim}{x};\theta)}{f(\underset{\sim}{y};\theta)}$ is independent of θ if $T(\underset{\sim}{x}) = T(\underset{\sim}{y})$, then $T(\underset{\sim}{x})$ is minimal sufficient for θ.

Proof:- Here $T(\underset{\sim}{x})$ is sufficient statistic for θ.

$$f(\underset{\sim}{x};\theta) = g(T(\underset{\sim}{x});\theta)\, h(\underset{\sim}{x}) \quad [\text{By factorization criterion}]$$

To show $T(\underset{\sim}{x})$ is minimal, let $T'(\underset{\sim}{x})$ be any other sufficient statistic. By the factorization theorem, there exist fuction g' and h' $\ni$ $f(\underset{\sim}{x};\theta) = g'(T(\underset{\sim}{x});\theta)\cdot h'(\underset{\sim}{x})$. Let, $T'(\underset{\sim}{x}) = T'(\underset{\sim}{y})$, then,

$$\frac{f(x;\theta)}{f(\underset{\sim}{y};\theta)} = \frac{g'(T'(\underset{\sim}{x});\theta)\, h'(\underset{\sim}{x})}{g'(T'(\underset{\sim}{y});\theta)\, h'(\underset{\sim}{y})} = \frac{h'(\underset{\sim}{x})}{h'(\underset{\sim}{y})}$$

since the ratio does not depend on θ, so $T(\underset{\sim}{x})$ is minimal sufficient for θ.

Ex.(1). Let $X_1, X_2, \ldots, X_n$ be a r.s. from $Bin(1, p)$. S.T. $\sum_{i=1}^{n} X_i$ is a minimal sufficient statistic for p.

Soln.:→

$$\frac{f(\underset{\sim}{x}; p)}{f(\underset{\sim}{y}; p)} = \frac{p^{\sum_{i=1}^{n} x_i} (1-p)^{n-\sum_{i=1}^{n} x_i}}{p^{\sum_{i=1}^{n} y_i} (1-p)^{n-\sum_{i=1}^{n} y_i}}$$

$$= \left(\frac{p}{1-p}\right)^{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i}; \text{ is independent of } p$$

iff $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$.

Hence $T = \sum_{i=1}^{n} x_i$ is minimal sufficient for p.

Ex.(2). Let $X_1, \ldots, X_n$ be a r.s. from $N(\mu, \sigma^2)$. Then S.T. $(\bar{X}, S^2)$ is a minimal sufficient statistic for (μ, σ^2).

Soln.→ (Normal minimal sufficient statistic)

$$\frac{f(\underset{\sim}{x}; \mu, \sigma^2)}{f(\underset{\sim}{y}; \mu, \sigma^2)} = \frac{(2\pi\sigma^2)^{-n/2} \exp\left(-\left[n(\bar{x}-\mu)^2 + (n-1)s_x^2\right]/2\sigma^2\right)}{(2\pi\sigma^2)^{-n/2} \exp\left(-\left[n(\bar{y}-\mu)^2 + (n-1)s_y^2\right]/2\sigma^2\right)}$$

$$= \exp\left[\left\{-n(\bar{x}^2 - \bar{y}^2) + 2n\mu(\bar{x}-\bar{y}) - (n-1)(s_x^2 - s_y^2)\right\}/2\sigma^2\right]$$

This ratio will be a constant as a function of μ and σ^2 iff $\bar{x} = \bar{y}$ and $s_x^2 = s_y^2$. Then by the theorem, $(\bar{X}, S^2)$ is a minimal sufficient statistic for (μ, σ^2).

Ex.(3). Let $X_1, \ldots, X_n$ be a random sample from $U(\theta, \theta+1)$, $-\infty < \theta < \infty$. S.T. $(X_{(1)}, X_{(n)})$ is a minimal sufficient statistic.

Soln.→ The PDF can be written in the form:

$$f(\underset{\sim}{x}; \theta) = \begin{cases} 1 & \text{if } \max_i x_i - 1 < \theta < \min_i x_i \\ 0 & \text{OW} \end{cases}$$

Letting $X_{(1)} = \min_i X_i$ and $X_{(n)} = \max_i x_i$, then we have $T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$ is a minimal sufficient statistic.

This is a case where the dimension of a minimal sufficient statistic does not match with the dimension of the parameter.

Remark:- A minimal sufficient statistic is not unique. Any one-to-one function of a minimal sufficient statistic is also a minimal sufficient statistic. Example:-

i) $T'(\underset{\sim}{X}) = \left(X_{(n)} - X_{(1)}, (X_{(n)} + X_{(1)})/2\right)$ is also a minimal statistic in Ex.(3). (For uniform distn.)

ii) $T'(\underset{\sim}{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is also a minimal sufficient statistic in Ex.(2). (For normal distn.).

(IV) COMPLETENESS :—

Let $(X_1, \ldots, X_n)$ be a r.s. from the distn. with PMF/PDF $f(x;\theta)$, $\theta \in \Omega$. Let $\{g(t;\theta) : \theta \in \Omega\}$ be the family of distn. of a statistic T.

<u>Definition</u>:– The family of distn. $\{g(t;\theta) ; \theta \in \Omega\}$ of a statistic T defined to be complete iff $E\{h(T)\} = 0 \;\forall\; \theta \in \Omega$ implies $P[h(T) = 0] = 1 \;\forall\; \theta \in \Omega$.

Also, the statistic T is said to be complete iff its family of distns $\{g(t;\theta) : \theta \in \Omega\}$ is complete.

<u>Ex.(1).</u> Let $X_1, \ldots, X_n$ be a r.s. from Bin(1, p). S.T. $(X_1 - X_2)$ is not complete but $T = \sum_{i=1}^{n} X_i$ is complete for the population distn..

<u>soln.</u> :→ Note that, $E(X_1 - X_2) = p - p = 0 \;\forall\; p \in (0,1)$

but
$$P[(X_1 - X_2) = 0] = P[X_1 = 0, X_2 = 0] + P[X_1 = 1, X_2 = 1]$$
$$= (1-p)^2 + p^2$$
$$\neq 1$$

Hence, $(X_1 - X_2)$ is not complete.

[T is not complete ⇒ there exists some $h(T) \neq 0 \ni E[h(T)] = 0$]

Now, note that, $T = \sum_{i=1}^{n} X_i \sim Bin(n, p)$

Now, $E(h(T)) = 0 \;\forall\; p \in (0,1)$

$$\Rightarrow \sum_{t=0}^{n} h(T) \binom{n}{t} p^t (1-p)^{n-t} = 0 \;\forall\; p \in (0,1)$$

$$\Rightarrow \sum_{t=0}^{n} h(T) \binom{n}{t} \left(\frac{p}{1-p}\right)^t = 0$$

$$\Rightarrow \sum_{t=0}^{n} h(T) \binom{n}{t} u^t = 0 \quad \forall \quad u = \frac{p}{1-p} ; \; u \in (0, \infty)$$

Equating the coefficients of u^t on both sides, we get

$$h(T) \binom{n}{t} = 0 \;\forall\; t = 1(1)n$$

$$\Rightarrow h(T) = 0 \;, \; t = 0(1)n, \text{ as } \binom{n}{t} > 0$$

i.e. $P[h(T) = 0] = 1 \;\forall\; p \in (0,1)$.

Hence, $T = \sum_{i=1}^{n} X_i$ is <u>complete and sufficient statistic</u>.

Ex.(2) Let X be an observation from $P(\lambda)$ distn.. S.T. X is complete, i.e. the family of distn. $\{P(\lambda) : \lambda > 0\}$ is complete.

Soln. →

$$\sum h\left(\frac{e^{-\lambda} \cdot \lambda^x}{x!}\right) = 0$$

Ex.(3). Let $X_1, \ldots, X_n$ be a r.s. from $U(0, \theta)$; $\theta > 0$. S.T. $X_{(n)}$ is complete.

solution:- The family of distn. of $T = X_{(n)}$ is $\{g(t;\theta) : \theta > 0\}$

where $g(t;\theta) = \begin{cases} \frac{nt^{n-1}}{\theta^n} & \text{if } 0 < t < \theta \\ 0 & \text{ow} \end{cases}$

Now, $E_\theta\{h(t)\} = 0 \quad \forall \theta > 0$

$$\Rightarrow \int_0^\theta h(t) \cdot \frac{nt^{n-1}}{\theta^n} dt = 0 \quad \forall \theta > 0$$

$$\Rightarrow \int_0^\theta h(t) \cdot t^{n-1} dt = 0 \quad \forall \theta > 0$$

Differentiating w.r.t. θ, we get

$h(\theta) \cdot \theta^{n-1} = 0 \quad \forall \theta > 0$

$\Rightarrow h(\theta) = 0 \quad \forall \theta > 0$

$\Rightarrow h(T) = 0 \quad \forall t > 0$

$\therefore P[h(t) = 0] = 1 \,;\, \theta > 0$

Hence, $T = X_{(n)}$ is complete for the popln. distn. $U(0, \theta)$, $\theta > 0$.

[**Leibnitz Rule:-**

(a) $$\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} f(x) dx = f(b(\theta)) \cdot b'(\theta) - f(a(\theta)) \cdot a'(\theta).$$

(b) $$\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} f(x;\theta) dx = \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta} f(x;\theta) dx + f(b(\theta)) \cdot b'(\theta) - f(a(\theta)) a'(\theta)$$]

Ex.(4). Example of sufficient statistic that is not complete:

Let $x_1, x_2, \ldots, x_n$ be a r.s. from $N(\theta, \theta^2)$. Then

$$\prod_{i=1}^{n} f(x_i;\theta) = \frac{1}{(2\pi.\theta^2)^{n/2}} \cdot \exp\left\{-\frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\} \quad ; \theta \neq 0$$

$$= \frac{1}{(2\pi\theta^2)^{n/2}} \cdot \exp\left\{-\frac{1}{2}\left[\frac{\sum x_i^2}{\theta^2} - \frac{2\sum x_i}{\theta} + 1\right]\right\}$$

$$= g\left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2 ; \theta\right) \cdot h(\underset{\sim}{x}), \text{ where } h(\underset{\sim}{x}) = 1.$$

$\Rightarrow T = \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right)$ is sufficient for θ. (This is minimal sufficient statistic)

Note that, $E\left(\sum_{i=1}^{n} x_i^2\right) = \sum_{i=1}^{n}\left\{V(x_i) + E^2(x_i)\right\}$

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

and $E\left(\sum_{i=1}^{n} x_i\right)^2 = E(n\bar{x})^2 = n^2 E(\bar{x})^2$

$$= n^2\left\{V(\bar{x}) + E^2(\bar{x})\right\}$$

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

$$= n(n+1)\theta^2$$

Hence, $E\left\{\frac{\sum_{i=1}^{n} x_i^2}{2n} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n(n+1)}\right\} = 0 \quad \forall\ \theta \neq 0$

$$\Rightarrow E\left\{(n+1)\sum_{i=1}^{n} x_i^2 - 2\left(\sum_{i=1}^{n} x_i\right)^2\right\} = 0 \ \forall\ \theta \neq 0$$

$\Rightarrow E(h(T)) = 0$, where $h(T) = (n+1)\sum_{i=1}^{n} x_i^2 - 2\left(\sum_{i=1}^{n} x_i\right)^2$

is not identically zero.

Hence $T = \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right)$ is not complete but sufficient.

Ex.(5). Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(\alpha\sigma, \sigma^2)$; α known.

S.T. $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is sufficient but not complete.

Ex. (6). Let $X_1, \ldots, X_n$ be a r.s. from $U(\theta, \theta+1)$. S.T. $(X_{(1)}, X_{(n)})$ is sufficient but not complete.

Solution:- Let $R = X_{(n)} - X_{(1)}$ is independent of location parameter θ (as dispersion is indep. of location).

The p.d.f. is $f_R(r) = n(n-1) r^{n-2}(1-r)$

$$E(R) = \frac{n-1}{n+1}$$

$$\Rightarrow E\left(X_{(n)} - X_{(1)} - \frac{n-1}{n+1}\right) = 0 \ \forall \theta$$

$$\Rightarrow P\left[X_{(n)} - X_{(1)} - \frac{n-1}{n+1} = 0\right] \neq 1$$

Hence $T = (X_{(1)}, X_{(n)})$ is sufficient but not complete.

Ex. (7). Let $X_1, \ldots, X_n$ be a r.s. from the PMF

$$p(x; N) = \begin{cases} \frac{1}{N} & , x = 1, 2, \ldots, N \\ 0 & \text{ow} \end{cases}$$

where, N is a positive integer.

Show that the family of distn. $X_{(n)}$ is complete.

Soln.:→ Let $T = X_{(n)}$, the CDF of T is given by,

$$\therefore \quad F_T(t) = P[X_{(n)} \le t]$$

$$= \prod_{x=1}^{n} P[X_1 \le t]$$

$$= \left(\frac{t}{N}\right)^n \ ; \ x = 1, \ldots, N.$$

$$P[T = t] = F_T(t) - F_T(t-1)$$

$$= \begin{cases} \frac{t^n - (t-1)^n}{N^n} & ; t = 1(1)N \\ 0 & ; \text{ow} \end{cases}$$

The family of distn. of $T = X_{(n)}$ is $\{g(t; N) : N = 1, 2, 3, \ldots\}$

where $g(t; N) = \begin{cases} \frac{t^n - (t-1)^n}{N^n} & , t = 1, 2, \ldots, N \\ 0 & , \text{ow} \end{cases}$

Now, let $E\{h(T)\} = 0 \ \forall N \geq 1$

$$\Rightarrow \sum_{t=1}^{N} h(t) \left\{\frac{t^n - (t-1)^n}{N^n}\right\} = 0 \ \forall \ N \geq 1$$

$$\Rightarrow \sum_{t=1}^{N} h(t) \cdot \{t^n - (t-1)^n\} = 0 \ \ \forall N \geq 1$$

For $N=1$, $h(1)\{1^n - 0^n\} = 0 \Rightarrow h(1) = 0$

For $N=2$, $h(1)\{1^n - 0^n\} + h(2)\{2^n - 1^n\} = 0$

$\Rightarrow h(2)\{2^n - 1^n\} = 0$ as $h(1) = 0$

$\Rightarrow h(2) = 0$

and so on.

Using an inductive argument, we have

$$h(1) = h(2) = h(3) = \cdots\cdots = h(N) = 0$$

$$\Rightarrow P[h(T) = 0] = 1 \quad \forall N = 1, 2, \ldots\ldots$$

Hence, $T = X_{(n)}$ is complete.

Remark on Completeness:

(1) Another way of stating that a statistic T is complete is the following :- T is complete iff the only unbiased estimator of zero, i.e. a function of T is the statistic that is identically zero.

(2) If T is complete statistic, then an unbiased estimator on θ based on T is unique.

Proof:— If possible, let $h_1(T)$ and $h_2(T)$ be two UEs of θ.

then $E(h_1(T)) = \theta = E(h_2(T)) \quad \forall \theta$

$\Rightarrow E(h_1(T) - h_2(T)) = 0 \quad \forall \theta$

$\Rightarrow h_1(T) - h_2(T) = 0$, with prob. 1, $\forall \theta$

$\Rightarrow h_1(T) = h_2(T)$, with prob 1, $\forall \theta$

Hence, an UE of θ based on T is unique.

(3) Concept of completeness:— If T is complete, then by definition, $E\{h(T)\} = 0 \ \forall \theta \Rightarrow h(T) = 0$ with prob. 1 $\forall \theta$. In other words, if $h(T) \neq 0$ then $E\{h(T)\} \neq 0$ and is a function of θ, that is, every non-null function of T possesses some information about θ.

If T is not complete, then there exists some non-null function of T, say $h(T)$, for which $E\{h(T)\} = 0$, that is, there exists some non-null function of T ($h(T)$), which don't contain any information about θ, or, $\exists$ some non-null functions of T which forget to carry any information about θ.

But if T is complete, then every non-null function of T carries some information about θ. This is the concept of completeness.

Ex.(8). Let $X_1, X_2, \ldots, X_n$ be a r.s. from Geometric distn with parameter p. S.T. $\sum_{i=1}^{n} X_i$ is complete for the family.

Solution:→ Let $T = \sum_{i=1}^{n} X_i$ then $T \sim NB(n, p)$.

$$E\{h(T)\} = 0$$

$$\Rightarrow \sum_{t=0}^{n} h(T) \binom{t+n-1}{t} p^n q^t = 0 \ \forall \ p \in (0,1) \text{ and } p+q=1.$$

$$\Rightarrow \sum_{t=0}^{n} h(T) \binom{-n}{t} q^t = 0$$

Equating the coefficient of q^t on both sides, we get,

$$h(T) \binom{-n}{t} = 0 \quad , \text{where } t = 1, 2, \ldots$$

$$\Rightarrow h(T) = 0$$

i.e. $P[h(T) = 0] = 1 \ \forall \ p \in (0,1)$

Hence, T is complete.

Exponential Family of Distributions:

A. One parameter Exponential Family of Distributions: (OPEF)

A one-parameter family of distributions $\{f(x;\theta): \theta \in \Omega\}$ that can be expressed as

$f(x;\theta) = \exp[u(\theta)\cdot T(x) + v(\theta) + w(x)]$, where the following regularity conditions hold:

C_1: The support $S = \{x; f(x;\theta) > 0\}$ does not depend on $\theta \,\forall\, \theta \in \Omega$

C_2: The parameter space Ω is an open interval of $\mathbb{R}$, that is, $\underline{\theta} < \theta < \overline{\theta}$.

C_3: $\{1, T(x)\}$ or $\{1, u(\theta)\}$ are linearly independent, that is, $T(x)$ or, $u(\theta)$ are non-constant functions; is defined to be a one-parameter exponential family (OPEF) of distn.s.

Ex. (1). Let $X \sim P(\lambda)$, $\lambda (>0)$ is unknown. Show that the family of distns $\{P(\lambda): \lambda > 0\}$ of X is an OPEF.

Solution:- The PMF of X is

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

$$= \exp[-\lambda + x\ln\lambda - \ln \lfloor x]$$

$$= \exp[u(\lambda)T(x) + v(\lambda) + w(x)]$$

where, $u(\lambda) = \ln\lambda$, $T(x) = x$, $v(\lambda) = -\lambda$, $w(x) = -\ln \lfloor x$.

C_1: The support $S = \{x: f(x,\lambda) > 0\} = \{0,1,2,3,\ldots\}$ is independent of λ.

C_2: The parameter space $\Omega = \{\lambda: 0 < \lambda < \infty\}$ is an open interval of R.

C_3: Here $T(x) = x$ or $u(\lambda) = \ln\lambda$ are non-constant functions.

Hence, the family of distribution $\{P(\lambda): \lambda > 0\}$ is an OPEF.

Ex. (2). Consider a family of distn. with PMF given by

$$f(x;\theta) = \begin{cases} \dfrac{a_x\theta^x}{g(\theta)} & , x = 0,1,2,\ldots \\ 0 & , \text{ow} \end{cases}$$

where, $0 < \theta < \rho$, $a_x \geq 0$ and $g(\theta) = \sum_{x=0}^{\infty} a_x\theta^x$.

ST $\{f(x;\theta): 0 < \theta < \rho\}$ is an OPEF of distns.

Solution:- Here, $f(x;\theta) = \exp[x\ln\theta - \ln g(\theta) + \ln a_x]$, $x = 0,1,2,3,\ldots$

$= \exp[u(\theta)\cdot T(x) + v(\theta) + w(x)]$, $x = 0,1,2,3,\ldots$

where, $T(x) = x$, $u(\theta) = \ln\theta$, etc.

C_1:- The support $S = \{0,1,2,\ldots\}$ is independent of θ.

C_2:- The parameter space $\Omega = \{\theta: 0 < \theta < \rho\}$ is an open interval of R.

C_3:- $T(x) = x$ and $u(\theta) = \ln\theta$ are non-constant functions.

Hence, the family of distn. is OPEF.

Remark:→

(1). As Power Series distn. are in OPEF, the distributions: Binomial, Poisson, Negative Binomial, etc. are in OPEF.

(2). We should verify that the families $\{N(\mu,1): \mu \in \mathbb{R}\}$, $\{Exp(\lambda); \lambda > 0\}$ are of OPEFs.

(3) As examples of families of PDFs, which are not of OPEF's are:

(i) $\{U(0,\theta): \theta > 0\}$ as the support $S=(0,\theta)$ depends on θ. → (one parameter case)

(ii) $\{$ Hypergeometric (N,m,n) : $N \in \{1,2,\ldots\}$, $m \in \{0,1,\ldots,N\}$, $n \in \{1,2,\ldots,N\}\}$

as the support $S \in \{\max(0, n+m-N), \ldots, \min(m,n)\}$ depend on the parameters. → (3 parameter case)

(iii) $\{f(x;\theta): \theta \in R\}$ where, $f(x;\theta) = \frac{1}{2} e^{-|x-\theta|}$; $x \in R$, or,

$f(x;\theta) = \frac{1}{\pi\{1+(x-\theta)^2\}}$; $x \in R$ as $f(x;\theta)$ can't be expressed in the form

$\exp[u(\theta) \cdot T(x) + v(\theta) + w(x)]$ but here c_1, c_2 holds but c_3 does not hold.

:→ This is an another example of one-parameter families of distn. which are not of One parameter exponential family of distns.

(iv) $\{f(x;\theta): \theta \in R\}$ where $f(x;\theta) = \begin{cases} e^{-(x-\theta)}, & x > \theta \\ 0, & ow \end{cases}$

is not in OPEF as the support $S = (\theta, \infty)$ depends on θ.

- **Theorem:**– Let $(X_1, X_2, \ldots, X_n)$ be a r.s. from an OPEF $\{f(x;\theta): \theta \in \Omega\}$, where,

  $f(x;\theta) = \exp[u(\theta)T(x) + v(\theta) + w(x)]$, then

  (a) $\sum_{i=1}^{n} T(X_i)$ is sufficient for θ.

  (b) $\sum_{i=1}^{n} T(X_i)$ is a complete sufficient statistic.

Solution:– (a) The PDF/PMF of $(X_1, \ldots, X_n)$ is

$$\prod_{i=1}^{n} f(x_i;\theta) = \exp\left[u(\theta) \cdot \sum_{i=1}^{n} T(x_i) + n v(\theta) + \sum_{i=1}^{n} w(x_i)\right]$$

$$= \exp\left[u(\theta) \cdot \left(\sum_{i=1}^{n} T(x_i)\right) + n v(\theta)\right] \times \exp\left[\sum_{i=1}^{n} w(x_i)\right]$$

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

By Neyman-fisher factorization criterion, $\sum_{i=1}^{n} T(X_i)$ is sufficient for θ.

Ex.(3):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from ~~an OPEF~~ the PDF

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1} & ; 0<x<1 \\ 0 & ; \text{ow} \end{cases}$$

Find a complete sufficient statistic for the family of distn.

Solution:- Note that,

$$f(x;\theta) = \exp\left[(\theta-1)\ln x + \ln\theta\right] \quad , 0<x<1$$

$$= \exp\left[\theta \ln x + \ln\theta - \ln x\right]$$

$$= \exp\left[u(\theta)\cdot T(x) + v(\theta) + w(x)\right], \text{ where,}$$

$$T(x) = \ln x, \quad u(\theta) = \theta, \text{ etc.}$$

C1: The support $S = \{x : 0<x<1\}$ is independent of θ.

C2: The parameter space $\Omega = \{\theta : 0<\theta<\infty\}$ is an open interval of $\mathbb{R}$.

C3: $T(x) = \ln x$, or, $u(\theta) = \theta$ are non-constant functions.

Hence, the family $\{f(x;\theta) : \theta \in \Omega\}$ of distn. is an OPEF.

Hence, by the above theorem, $\sum_{i=1}^{n} T(X_i) = \sum_{i=1}^{n} \ln X_i$ is a complete sufficient statistic.

Ex.(4). Let $X_1, \ldots, X_n$ be a r.s. from $f(x;\sigma) = \frac{1}{2\sigma} e^{-\frac{|x|}{\sigma}}$; $x \in \mathbb{R}$, $\sigma > 0$

Find the complete sufficient statistic for the family.

Ex.(5). Let $X_1, \ldots, X_n$ be a r.s. from $f(x;\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}$; $x \in \mathbb{R}, \mu \in \mathbb{R}$

Find the complete sufficient statistic.

Soln:→

[B]. K-parameter Exponential Family of Distribution :—

A K-parameter family of PDFs or PMFs $\{f(x;\underset{\sim}{\theta}) : \theta \in \Omega \subseteq R^k\}$ that can be expressed as

$$f(x;\underset{\sim}{\theta}) = \exp\left[\sum_{i=1}^{k} u_i(\underset{\sim}{\theta}) T_i(x) + v(\underset{\sim}{\theta}) + w(x)\right]$$

with the regular conditions:

<u>C_1</u>:- The support $S = \{x : f(x;\underset{\sim}{\theta}) > 0\}$ does not depend on $\underset{\sim}{\theta}$.

<u>C_2</u>:- The parameter space Ω is an open region of R^K that is, $\underline{\theta_i} < \theta_i < \overline{\theta_i}$, $i = 1(1)K$, containing K-dimensional rectangle.

<u>C_3</u>:- $\{1, T_1(x), T_2(x), \ldots, T_K(x)\}$ or $\{1, u_1(\underset{\sim}{\theta}), \ldots, u_K(\underset{\sim}{\theta})\}$ are linearly independent ; is called a K-parameter exponential family.

<u>Remark</u>:-

⟨1⟩ If $\{1, T_1(x), T_2(x), \ldots, T_K(x)\}$ or $\{1, u_1(\underset{\sim}{\theta}), \ldots, u_K(\underset{\sim}{\theta})\}$ is LD, then the no. of terms in the exponent can be reduced and K need not be the dimension of Ω. Hence, WLG, we shall assume that the representation is minimal in the sense that neither T_i's nor u_i's satisfy a linear constraint.

*⟨2⟩ Let $X_1, X_2, \ldots, X_n$ be a r.s. from the family $\{f(x;\underset{\sim}{\theta}) : \underset{\sim}{\theta} \in \Omega \subseteq R^K\}$ of distributions, where,

$$f(x;\underset{\sim}{\theta}) = \exp\left[\sum_{i=1}^{K} u_i(\theta) T_i(x) + v(\underset{\sim}{\theta}) + w(\underset{\sim}{x})\right], \text{ then}$$

$$T(\underset{\sim}{X}) = \left(\sum_{i=1}^{n} T_1(X_i), \sum_{i=1}^{n} T_2(X_i), \ldots, \sum_{i=1}^{n} T_K(X_i)\right)$$

is a complete sufficient statistic for the family.

<u>Ex.(1)</u>:- Consider the family $\{N(\mu, \sigma^2) : \mu \in R, \sigma^2 > 0\}$ of distn.s. Show that the family of distn.s is a two parameter exponential family. Hence, obtain a complete sufficient statistic based on a r.s. $(X_1, X_2, \ldots, X_n)$.

<u>Solution</u>:- Here $\underset{\sim}{\theta} = (\mu, \sigma)$, $\Omega = \{(\mu, \sigma) : \mu \in \mathbb{R}, 0 < \sigma < \infty\}$ the family of distn. is

$\{f(x;\underset{\sim}{\theta}) : \underset{\sim}{\theta} \in \Omega\}$, where,

$$f(\underset{\sim}{x};\underset{\sim}{\theta}) = \exp\left[-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2} - \frac{1}{2}\left\{\frac{\mu^2}{\sigma^2} + \ln(2\pi\sigma^2)\right\}\right]$$

$$= \exp\left[u_1(\underset{\sim}{\theta})\cdot T_1(x) + u_2(\underset{\sim}{\theta})\cdot T_2(x) + v(\theta) + w(x)\right]$$

where, $u_1(\theta) = -\frac{1}{2\sigma^2}$, $u_2(\theta) = \frac{\mu}{\sigma^2}$, $T_1(x) = x^2$, $T_2(x) = x$, etc.

C_1:- The support $S = R$ is independent of $\underset{\sim}{\theta}$.

C_2:- The parameter space Ω is an open subset of R^2.

C_3:- $\{1, T_1(x), T_2(x)\} = \{1, x, x^2\}$ or $\{1, u_1(\theta), u_2(\theta)\}$

$= \{1, -\frac{1}{2\sigma^2}, \frac{\mu}{2\sigma^2}\}$ are LIN.

Hence the family of distributions is two-parameter exponential family.

By Remark (2):- $T(\underset{\sim}{X}) = \left(\sum_{i=1}^{n} T_1(X_i), \sum_{i=1}^{n} T_2(X_i)\right) = \left(\sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} X_i\right)$ is a complete sufficient statistic for the family.

Ex. (2):- Is the family $\{N(\theta, \theta^2) : \theta \neq 0\}$ a two-parameter exponential family or OPEF? - Justify your answer.

Solution:- The family of distributions is given by $\{f(x;\theta) : \theta \neq 0\}$, where,

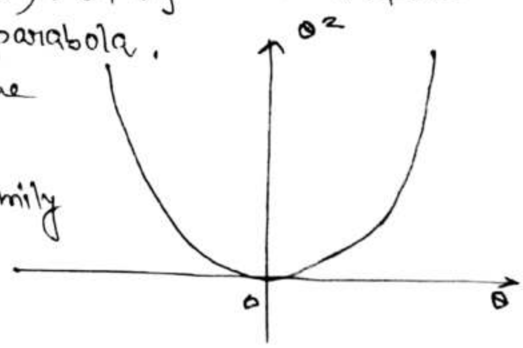
$$f(x;\theta) = \begin{cases} \frac{1}{\sqrt{2\pi}\,\theta} e^{-\frac{(x-\theta)^2}{2\theta^2}} & ; x \in R \\ 0 & ; \text{OW} \end{cases}$$

$$= \exp\left[-\frac{x^2}{2\theta^2} + \frac{x}{\theta} - \frac{1}{2}\left\{1 + \ln(2\pi\theta^2)\right\}\right]$$

$$= \exp\left[u_1(\theta)\cdot T_1(x) + u_2(\theta) T_2(x) + v(\theta) + w(x)\right]$$

where $u_1(\theta) = -\frac{1}{2\theta^2}$, $u_2(\theta) = \frac{1}{\theta}$, $T_1(x) = x^2$, $T_2(x) = x$, etc.

But the parameter space $\Omega = \{(\theta, \theta^2) : \theta \neq 0\}$ is not an open rectangle in R^2, infact, it is a parabola.



Hence, C_2 does not hold that is, the family is not a two-parameter exponential family. This type of family is known as two-parameter curved exponential family.

The PDF $f(x;\theta)$ does not ensure the form of the OPEF and Ω is not an open interval in R. Hence, it is not an OPEF.

Also note that $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is not complete but sufficient.

Ex. (3):- Consider the families of distn.s

(i) $\{Gamma(a, p) : a > 0, p > 0\}$

(ii) $\{Beta(\alpha, \beta) : \alpha > 0, \beta > 0\}$

Show that the families are two-parameter exponential family. Suggest a complete sufficient statistic for each case, based on a r.s. $(X_1, \cdots, X_n)$.

Ex. (4):- Consider the two parameter families of distn.s:

(i) $\{U(\theta_1, \theta_2) : \theta_1 < \theta_2\}$,

(ii) $\left\{ f(x : \alpha, \theta) = \dfrac{\theta x^{\theta - 1}}{\alpha \theta} ; \alpha \in R, \theta > 0, x > \alpha \right\}$

(iii) $\left\{ f(x ; \theta, \alpha) = \frac{1}{\theta} e^{-\left(\frac{x - \alpha}{\theta}\right)} ; x > \alpha, \alpha \in R, \theta > 0 \right\}$

Show that they are not two-parameter exponential families.

—: UMVUE and Method of finding UMVUE : —

Uniformly Minimum Variance Unbiased Estimator (UMVUE) :—

Let T_1 and T_2 be two different UEs of θ. Then $\exists$ an infinitely many UEs of θ of the form:

$$T_\alpha = \alpha T_1 + (1-\alpha) T_2 \; ; \; 0 \le \alpha \le 1$$

Which of these should we choose?
Here comes the concept of UMVUE.

Definition: UMVUE

(a) An estimator T^* is defined to be UMVUE of θ iff
(i) $E(T^*) = \theta \; \forall \; \theta \in \Omega$
(ii) $Var_\theta(T^*) \le Var_\theta(T) \; \forall \; \theta \in \Omega$
for any estimator T which satisfies $E(T) = \theta \; \forall \; \theta \in \Omega$.

(b) An UE is said to be UMVUE of θ if it has minimum variance among all UEs of θ.

Ex. (1):— Let $X_1, X_2, \ldots, X_n$ be a r.s. from $U(0, \theta)$. Find two UEs of θ, one based on $\bar{X}$ and other based on $X_{(n)}$. Which one is better?

Solution:— $E(\bar{X}) = E(X_1) = \dfrac{\theta}{2}$

$$\Rightarrow E(2\bar{X}) = \theta$$

Hence $T_1 = 2\bar{X}$ is an UE of θ.

$$E(X_{(n)}) = \int_0^\theta x \cdot \frac{n x^{n-1}}{\theta^n} dx \quad \left[\because f_{X_{(n)}}(x) = \begin{cases} \dfrac{n x^{n-1}}{\theta^n} & ; 0 < x < \theta \\ 0 & ow \end{cases} \right]$$

$$= \frac{n}{\theta^n} \int_0^\theta x^n dx = \frac{n\theta}{n+1}.$$

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

Hence, $T_2 = \dfrac{n+1}{n} X_{(n)}$ is an UE of θ.

Now, $Var(T_1) = 4 \cdot V(\bar{X}) = 4 \cdot \dfrac{V(X_1)}{n} = \dfrac{4 \cdot \theta^2}{12 n} = \dfrac{\theta^2}{3n}$.

and
$$Var(T_2) = \left(\frac{n+1}{n}\right)^2 E\left(X_{(n)}^2\right) - E^2\left(\frac{n+1}{n} X_{(n)}\right)$$

$$= \left(\frac{n+1}{n}\right)^2 \int_0^\theta \frac{x^2 \cdot n x^{n-1}}{\theta^n} dx - \theta^2$$

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

Note that, $\dfrac{V(T_1)}{V(T_2)} = \dfrac{n+2}{3} \ge 1 \; \forall \; n \in \mathbb{N}$

For $n > 1$, $V(T_1) > V(T_2)$ and T_2 has smaller variance than T_1. Hence, $T_2 = \dfrac{(n+1)}{n} X_{(n)}$ is better estimator in finding θ.

Theorem:- The UMVUE of parameter, if exists, is unique.

Proof:- If possible, let T_1 and T_2 be two UMVUEs of θ.

Then $V(T_1) = V(T_2) = \gamma$, say.

Clearly, $Var(T) \geq \gamma$ where $T = \frac{T_1+T_2}{2}$ is an UE of θ.

$$\Rightarrow Var\left(\frac{T_1+T_2}{2}\right) \geq \gamma$$

$$\Rightarrow \frac{1}{4}\left[V(T_1)+V(T_2)+2Cov(T_1,T_2)\right] \geq \gamma$$

$$\Rightarrow \frac{1}{4}\left[\gamma+\gamma+2\rho\gamma\right] \geq \gamma \quad \left[\because Cov(T_1,T_2) = \rho\sqrt{V(T_1)V(T_2)} = \rho\gamma\right]$$

$\Rightarrow \rho \geq 1$ but we know $|\rho| \leq 1$

Hence $\rho = 1 \Rightarrow T_1 = a + bT_2$ with prob. 1, where $b>0$

Now, $E(T_1) = a + bE(T_2)$

$\Rightarrow \theta = a + b\theta \;\; \forall \theta$

$\Rightarrow a = 0, b = 1$, equating the coefficient of constant term and θ.

$\left[\begin{array}{l} V(T_1) = b^2 V(T_2) \Rightarrow b^2 = 1,\ b>0, \Rightarrow b = 1, \text{ and} \\ E(T_1) = a + bE(T_2) \Rightarrow \theta = a + 1.\theta \Rightarrow a = 0 \end{array} \right]$

Hence $T_1 = T_2$ with prob. 1.

i.e. UMVUE, if exists, is unique.

Ex.(2). Let T_1 and T_2 be two UEs with common variance $\alpha\sigma^2$, where σ^2 is the variance of the UMVUE. Show that,

$$\rho_{T_1,T_2} \geq \frac{2-\alpha}{\alpha}.$$

Solution:-

Note that, $T = \frac{T_1+T_2}{2}$ is an UE of the parameter.

Clearly, $V(T) \geq \sigma^2$

$$\Rightarrow V\left(\frac{T_1+T_2}{2}\right) \geq \sigma^2$$

$$\Rightarrow \frac{1}{4}\left[V(T_1)+V(T_2)+2Cov(T_1,T_2)\right] \geq \sigma^2$$

$$\Rightarrow \frac{1}{4}\left[2\alpha\sigma^2 + 2\rho_{T_1,T_2}\cdot\alpha\sigma^2\right] \geq \sigma^2$$

$$\Rightarrow \frac{\alpha}{2}\left\{1+\rho_{T_1,T_2}\right\} \geq 1$$

$$\Rightarrow \rho_{T_1,T_2} \geq \frac{2}{\alpha} - 1 = \frac{2-\alpha}{\alpha}.$$

UMVUE (Continued) : —

*Theorem (1) :- A necessary and sufficient condition for UMVUE :

Let X have distn. given by $\{f(x;\theta): \theta \in \Omega\}$. Let us define
$U_\psi = \{T(X): E(T(X)) = \psi(\theta),\ V(T(X)) < \infty\ \forall\ \theta \in \Omega\}$ and
$U_0 = \{u(X): E(u(X)) = 0,\ V(u(X)) < \infty\ \forall\ \theta \in \Omega\}$
Then $T^* \in U_\psi$ is UMVUE of $\psi(\theta)$ iff $\text{Cov}(u, T^*) = 0\ \forall\ u \in U_0$ and for all $\theta \in \Omega$. ———— $(*)$

Proof :- Necessary Part (Only if) :—
Suppose that T^* is UMVUE of $\psi(\theta)$.
If possible, let $(*)$ does not hold. Then $\exists$ a $u_0 \in U_0$ and a $\theta_0 \in \Omega$ such that

$$\text{Cov}(u_0, T^*) \neq 0 \text{ at } \theta = \theta_0$$

Now, for any real λ,

$$E(T^* + \lambda u_0) = \psi(\theta) + \lambda \cdot 0 = \psi(\theta)$$

$\Rightarrow T^* + \lambda u_0 \in U_\psi$.

and
$$V(T^* + \lambda u_0) = V(T^*) + \lambda^2 V(u_0) + 2\lambda \text{cov}(u_0, T^*)$$
$$= V(u_0)\left\{\lambda^2 + 2\lambda \frac{\text{cov}(u_0, T^*)}{V(u_0)}\right\} + V(T^*)$$
$$= V(u_0)\left\{\lambda + \frac{\text{cov}(u_0, T^*)}{V(u_0)}\right\}^2 + V(T^*) - \frac{\text{cov}^2(u_0, T^*)}{V(u_0)}$$

Set, $\lambda = -\dfrac{\text{cov}(u_0, T^*)}{V(u_0)}$, then at $\theta = \theta_0$,

$$V(T^* + \lambda u_0) = V(T^*) - \frac{\text{Cov}^2(u_0, T^*)}{V(u_0)} < V(T^*)$$

Since, $\text{Cov}(u_0, T^*) \neq 0$ at $\theta = \theta_0$,
which contradicts the fact that T^* is UMVUE.
Hence, we must have $\text{cov}(u, T^*) = 0\ \forall\ \theta \in \Omega\ \forall\ u \in U_0$.

Sufficient Part (If part) :— Suppose that $\text{Cov}(u, T^*) = 0\ \forall\ \theta \in \Omega_0\ \forall\ u \in U_0$
Consider any $T \in U_\psi$, then, as $T^* \in U_\psi$,
we have $E(T - T^*) = \psi(\theta) - \psi(\theta) = 0\ \forall\ \theta \in \Omega$

$\Rightarrow T - T^* \in U_0$

Hence, $\text{cov}(T - T^*, T^*) = 0\ \forall\ \theta \in \Omega$

$\Rightarrow \text{cov}(T, T^*) = V(T^*)\ \forall\ \theta \in \Omega$ .... $(*)$

Now,
$$0 \le \text{Var}(T - T^*) = V(T) + V(T^*) - 2\text{cov}(T, T^*)$$
$$\le V(T) - V(T^*) \quad [\text{By } (*)]$$
$$\Rightarrow V(T^*) \le V(T)\ \forall\ \theta \in \Omega$$

Hence, T^* is UMVUE of $\psi(\theta)$.

Theorem (2):- Let $T_1 \in U_\psi$ be UMVUE of $\psi(\theta)$. Then T_1 is necessarily unique.

Proof:- If possible, let T_2 be also UMVUE of $\psi(\theta)$.

Then $E(T_1 - T_2) = \psi(\theta) - \psi(\theta) = 0 \;\forall\, \theta \in \Omega$

$\Rightarrow T_1 - T_2 \in U_0$.

$\therefore$ By theorem (1),

$\text{cov}(T_1 - T_2, T_1) = 0 = \text{cov}(T_1 - T_2, T_2)$

$\Rightarrow V(T_1) = \text{cov}(T_1, T_2) = V(T_2) \;\forall\, \theta \in \Omega$

Now, $V(T_1 - T_2) = V(T_1) + V(T_2) - 2\text{cov}(T_1, T_2) = 0 \;\forall\, \theta \in \Omega$

$\Rightarrow E(T_1 - T_2)^2 = 0 \;\forall\, \theta \in \Omega$ as $E(T_1 - T_2) = 0$.

$\Rightarrow T_1 - T_2 = 0$ with prob. 1. $\forall\, \theta \in \Omega$

Hence, UMVUE of a parametric function is unique, if it exists.

Theorem (3):- Let T_i be UMVUE of $\psi_i(\theta)$, $i = 1(1)K$, then $T = \sum_{i=1}^{K} a_i T_i$ is UMVUE for $\psi(\theta) = \sum_{i=1}^{K} a_i \psi_i(\theta)$.

Hints:- $\text{cov}(u, T_i) = 0 \;\forall\, \theta \in \Omega \;\forall\; u \in U_0, \forall\, i = 1(1)K$

$\text{cov}(u, T) = \sum_{i=1}^{K} a_i \,\text{cov}(u, T_i) = 0 \;\forall\, \theta \in \Omega,\; u \in U_0$

* **Theorem (4):-** (Rao - Blackwell) [CU] '10 (5)

Let X have the distn from $\{f(x;\theta) : \theta \in \Omega\}$ and 'h' be any statistic in $U_\psi = \{h(X) : E(h(X)) = \psi(\theta),\; V(h(X)) < \infty,\; \forall\, \theta \in \Omega\}$. Let T be a sufficient statistic for θ. Then the conditional expectation $E[h/T]$ is an UE of $\psi(\theta)$.

Moreover, $\text{Var}\{E(h/T)\} \le \text{Var}\{h\} \;\forall\, \theta \in \Omega$ .... $(*)$

The equality in $(*)$ holds iff $h = E[h/T]$, with prob. 1, $\forall\, \theta \in \Omega$

Proof:- As T is sufficient for θ, the conditional distn of 'h' given T is independent of θ and $E\{h/T\}$ is independent of θ. Hence $E\{h/T\}$ is a function of T and is a statistic.

Note that $E\{E(h/T)\} = E(h) = \psi(\theta) \;\forall\, \theta \in \Omega$

$\Rightarrow E(h/T)$ is an UE of $\psi(\theta)$.

Now, $Var(h) = Var\{E(h/T)\} + E\{Var(h/T)\}$

$$= Var\{E(h/T)\} + E\{E[(h-E(h/T))^2/T]\}$$

$$= Var(E(h/T)) + E\{h - E(h/T)\}^2$$

Clearly, $Var(h) \geq Var\{E(h/T)\}$, since $E\{h-E(h/T)\}^2 \geq 0$

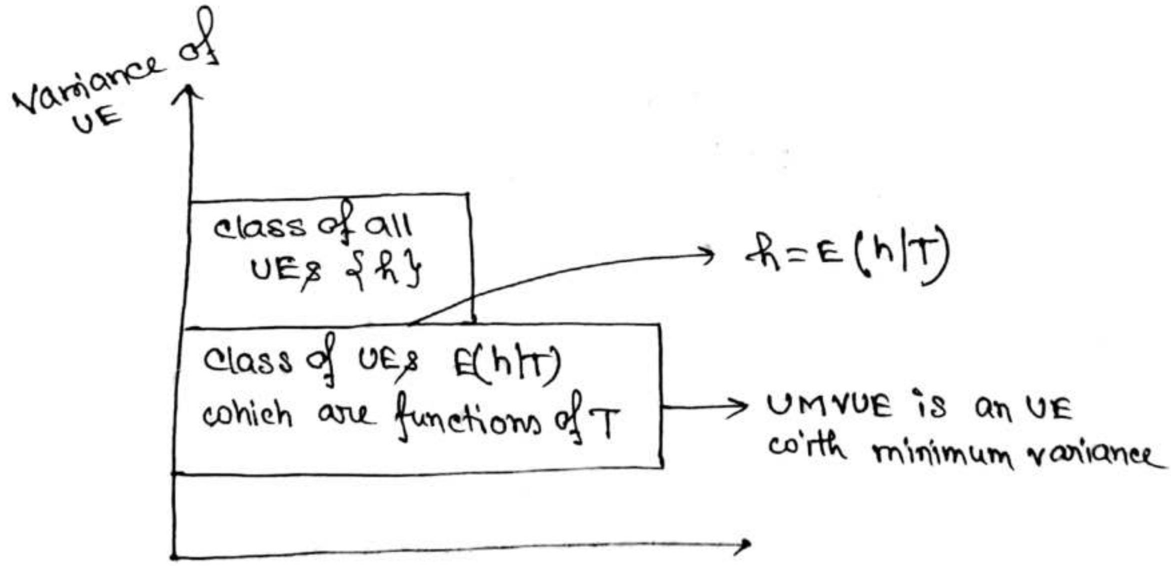
'=' holds iff $E\{h-E(h/T)\}^2 = 0$

iff $h = E(h/T)$ with probability 1, $\forall \theta \in \Omega$.

[CU'10] (2)

Implication of Rao-Blackwell Theorem:— If we start with an arbitrary unbiased estimator $h(x)$ of $\psi(\theta)$. Then we can include the estimator or we can get a better estimator than $h(x)$ by considering $E[h/T]$ where T is sufficient for θ, in the sense of having minimum MSE. Hence, Rao-Blackwell theorem says that to find UMVUE, we can concentrate only on those unbiased estimators which are functions of T, i.e. the UMVUE in the estimator which has minimum variance among all unbiased estimators which are functions of T.

Hence UMVUE is necessarily a functions of a sufficient statistic.



* **Theorem(s).** [Lehmann - Scheffe]

Let X has distribution from $\{f(x;\theta) : \theta \in \Omega\}$ and let T be a complete sufficient statistic. Again, if $E\{h(T)\} = \psi(\theta)$, then the UMVUE of $\psi(\theta)$ is the unique UE $h(T)$ [which is given by $E[h^*(X) | T]$ where $h^*(X)$ is an UE of $\psi(\theta)$].

Proof:- Let $h_1(T)$ and $h_2(T)$ be two UEs of $\psi(\theta)$.
then $E\, h_1(T) = \psi(\theta) = E\, h_2(T)$, $\forall\, \theta \in \Omega$

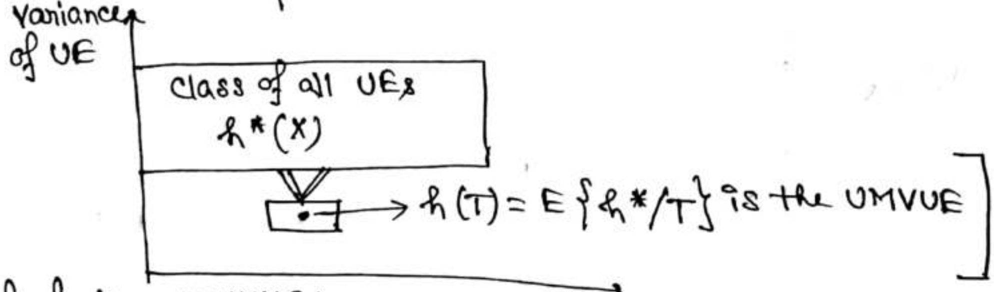
$$\Rightarrow E\{h_1(T) - h_2(T)\} = 0 \quad \forall\, \theta \in \Omega$$

$$\Rightarrow h_1(T) - h_2(T) = 0, \text{ with prob. } 1 \ \forall\, \theta \in \Omega$$

Hence, UE $h(T)$, based on T, of $\psi(\theta)$ is unique.

By Rao - Blackwell theorem, finding UMVUE in the class of all UEs is equivalent to finding UMVUE in the class of UEs based on T. But there is only one UE based on complete sufficient statistic T, say $h(T)$. Hence, $h(T)$ is the UMVUE of $\psi(\theta)$.

[Again, from Rao-Blackwell theorem, $E\{h^*(X|T)\}$ is an UE of $\psi(\theta)$ for any UE $h^*(X)$ and it is a function of T. As UEs based on T is unique, hence $h(T)$ must be $E\{h^*(X) | T\}$



]

Method of finding UMVUE:- Two systematic methods are available for deriving UMVUE through the Lehmann-Scheffe theorem.

(I) Method one:- Sometimes, we happened to know an UE $h(T)$ of $\psi(\theta)$, where T is a complete sufficient statistic, then the Lehmann - Scheffe theorem states that $h(T)$ is UMVUE of $\psi(\theta)$.

(II) Method two:- Conditioning method: If h is any UE of $\psi(\theta)$. It follows from Lehmann-Scheffe theorem that the UMVUE can be obtained as $E(h/T)$. For this derivation, it does not matter which UE h is being conditioned; one can choose h so that $E(h/T)$ is easily obtainable.

Ex.(1):- Let $(X_1, X_2, \ldots, X_n)$ be a r.s. from $Bin(1,p)$. Find that UMVUE of (i) $p = E(X_1)$ (ii) $Var(X_1) = p(1-p)$, (iii) p^s ; $s \leq n$

Solution:- The PMF of the family $\{B(1,p) : 0<p<1\}$ is

$$f(x;p) = p^x (1-p)^{1-x} \; ; \; x = 0,1$$

$$= \left(\frac{p}{1-p}\right)^x (1-p)$$

$$= \exp\left[\ln\left(\frac{p}{1-p}\right)x + \ln(1-p)\right]$$

$$= \exp\left[u(p) \cdot T(x) + v(p) + w(x)\right];$$

where, $T(x) = x$

It can be shown that $\{B(1,p) : 0<p<1\}$ is an OPEF.

Hence, $T = \sum_{i=1}^{n} T(x_i) = \sum_{i=1}^{n} X_i$ is a complete sufficient statistic.

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

and $E\left[(T)_s\right] = (n)_s \cdot p^s$, $s \leq n$

(i) $E(T) = np$

$$\Rightarrow E\left(\frac{T}{n}\right) = p$$

By Lehmann-Scheffe theorem,

$h_1(T) = \frac{T}{n} = \bar{X}$ is the UMVUE of p.

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

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

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

By Lehmann-Scheffe theorem,

$h_2(T) = \frac{T(n-T)}{n(n-1)}$ is the UMVUE of $p(1-p)$.

(iii) $E\left\{\frac{(T)_s}{(n)_s}\right\} = p^s$; $s \leq n$

By L-S theorem, $h_3(T) = \frac{(T)_s}{(n)_s}$ is the UMVUE of p^s.

Ex.(2):- Let $(X_1, \ldots, X_n)$ be a r.s. from $\{P(\lambda): \lambda > 0\}$. Find the UMVUE of (i) λ (ii) λ^s (iii) $\sum_{s=0}^{\infty} a_s \lambda^s$, (iv) $P[X_1 = k]$, (v) $P[X_1 = 0 \text{ or } 1]$.

Hints:- It can be shown that $T = \sum_{i=1}^{n} X_i$ is complete sufficient.

Then $T = \sum_{i=1}^{n} X_i \sim P(n\lambda)$ and $E[(T)_s] = (n\lambda)^s$; $s \in \mathbb{N}$

(i) $\frac{T}{n} = \bar{X}$ is the UMVUE of λ.

(ii) $\frac{(T)_s}{n^s}$ is the UMVUE of λ^s; $s \in \mathbb{N}$

(iii) By theorem (3),

$\sum a_s \cdot \frac{(T)_s}{n^s}$ is the UMVUE of $\sum_{s=0}^{\infty} a_s \lambda^s$.

(iv) Here $\psi(\lambda) = P[X_1 = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$

Define, $h(\underset{\sim}{X}) = \begin{cases} 1, & X_1 = k \\ 0, & \text{ow} \end{cases}$

$$E(h(\underset{\sim}{X})) = 1 \cdot P[X_1 = k] + 0 \cdot P[X_1 \neq k]$$
$$= P[X_1 = k] = \psi(\lambda).$$

Hence, $h(\underset{\sim}{X})$ is an UE of $\psi(\lambda)$.

By L-S theorem $E[h(\underset{\sim}{X}) | T]$ is the UMVUE of $\psi(\lambda) = P[X_1 = k]$

Now,
$$E[h(\underset{\sim}{X}) | T = t] = 1 \cdot P[X_1 = k / T = t] + 0$$
$$= \frac{P[X_1 = k; T = t]}{P[T = t]}$$
$$= \frac{P[X_1 = k; \sum_{i=1}^{n} X_i = t]}{P[\sum_{i=1}^{n} X_i = t]}$$
$$= \frac{P[X_1 = k]\, P[\sum_{i=2}^{n} X_i = t - k]}{P[\sum_{i=1}^{n} X_i = t]}, \text{ due to independence of } X_i\text{'s.}$$
$$= \frac{\left(e^{-\lambda} \cdot \frac{\lambda^k}{k!}\right)\left(e^{-(n-1)\lambda} \cdot \frac{\{\overline{n-1}\,\lambda\}^{t-k}}{\lfloor t-k}\right)}{e^{-n\lambda} \cdot \frac{(n\lambda)^t}{t!}}$$
$$= \frac{t!}{k!\,(t-k)!} \cdot \frac{(n-1)^{t-k}}{n^t}, \quad k = 0(1)t$$
$$= \binom{t}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{t-k}$$

Hence $E[h(\underset{\sim}{X})/T] = \binom{T}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{T-k}$

with $T = \sum_{i=1}^{n} X_i$, is the UMVUE of $\psi(\lambda) = P[X_1 = K]$

(v) Here $\psi(\lambda) = P[X_1 = 0 \text{ or } 1]$

$= P[X_1 = 0] + P[X_1 = 1]$

Note that $\binom{T}{0}\left(\frac{1}{n}\right)^0\left(1-\frac{1}{n}\right)^{T-0} = \frac{(n-1)^T}{n^T}$ and

$\binom{T}{1}\left(\frac{1}{n}\right)^1\left(1-\frac{1}{n}\right)^{T-1} = \frac{T(n-1)^{T-1}}{n^T}$ are the

UMVUE of $P[X_1=0]$ and $P[X_1=1]$ respectively.

By Theorem (3), $\frac{(n-1)^T + T(n-1)^{T-1}}{n^T}$ is the UMVUE of

$P[X_1 = 0 \text{ or } 1] = \psi(\lambda)$.

Direct Derivation:— Define, $h = \begin{cases} 1, & \text{if } X_1 = 0 \text{ or } 1 \\ 0, & \text{ow} \end{cases}$

as an UE of $P[X_1 = 0 \text{ or } 1] = \psi(\lambda)$

By L-S Theorem, $E\{h/T\}$ is the UMVUE of $P[X_1 = 0 \text{ or } 1] = \psi(\lambda)$

Now, $E[h/T=t] = 1.\, P[X_1 = 0 \text{ or } 1 \mid \sum_{i=1}^{n} X_i = t]$ $[\because (A\cup B)\cap C = (A\cap C)\cup(B\cap C)]$

$$= \frac{P[X_1 = 0 \text{ or } 1;\ \sum_{i=1}^{n} X_i = t]}{P[\sum_{i=1}^{n} X_i = t]}$$

$$= \frac{P[X_1=0;\ \sum_{i=1}^{n} X_i = t] + P[X_1=1;\ \sum_{i=1}^{n} X_i = t]}{P[\sum_{i=1}^{n} X_i = t]}$$

$$= \frac{P[X_1=0;\ \sum_{i=2}^{n} X_i = t] + P[X_1=1;\ \sum_{i=2}^{n} X_i = t-1]}{P[\sum_{i=1}^{n} X_i = t]}$$

$$= \frac{e^{-\lambda}.\frac{\lambda^0}{\lfloor 0}.e^{-(n-1)\lambda}.\frac{(\overline{n-1}\,\lambda)^t}{\lfloor t}}{e^{-n\lambda}.\frac{(n\lambda)^t}{\lfloor t}} + \frac{e^{-\lambda}.\frac{\lambda^1}{\lfloor 1}.e^{-(n-1)\lambda}.\frac{(\overline{n-1}\,\lambda)^{t-1}}{\lfloor t-1}}{e^{-n\lambda}.\frac{(n\lambda)^t}{t!}}$$

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

Hence, $\frac{(n-1)^T + T(n-1)^{T-1}}{n^T}$ is the UMVUE of $P[X_1 = 0 \text{ or } 1]$

$= \psi(\lambda)$.

Ex. (3):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $Bin(1, p)$.

(a) S.T. $T_1 = \begin{cases} 1 & \text{if } X_1 = 1, X_2 = 0 \\ 0 & \text{ow} \end{cases}$

is an UE of variance $X_1 = p(1-p)$

Hence find UMVUE of $V(X_1)$.

(b) Find an UE of p^r based on $X_1, X_2, \ldots, X_r$ only. Hence find UMVUE of p^r.

Solution:- (a)

(b) $\psi(p) = p^r = P[X_1 = 1, \ldots, X_r = 1]$

Hence $h(X_1, \ldots, X_r) = \begin{cases} 1, & \text{if } X_1 = 1, \ldots, X_r = 1 \\ 0, & \text{ow} \end{cases}$

is an UE of $p^r = \psi(p)$ based on $X_1, \ldots, X_r$.

By L-S theorem,

$$E\left[h(X_1, \ldots, X_r) \middle| T\right] \text{ is the UMVUE of } p^r.$$

$$E\left[h(X_1, \ldots, X_r) \middle| T = t\right]$$

$$= 1. \frac{P\left[X_1 = 1, \ldots, X_r = 1; \sum_{i=r+1}^{n} X_i = t - r\right]}{P\left[\sum_{i=1}^{n} X_i = t\right]}$$

$$= \frac{p^r \binom{n-r}{t-r} p^{t-r} (1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}} = \frac{\binom{n-r}{t-r}}{\binom{n}{t}} = \frac{(t)_r}{(n)_r}$$

Ex.(4):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $P(\lambda)$. Find an UE of $e^{-2\lambda}$ based on only X_1 and X_2. Hence find UMVUE of $e^{-2\lambda}$.

Ex.(5):- Let $X_1, \ldots, X_n$ be a r.s. from $N(\theta, 1)$. Find the UMVUE of (i) θ, (ii) θ^2.

Solution:- (i) The family $\{N(\theta,1) : \theta \in \mathbb{R}\}$ of distn is an OPEF and $T = \sum_{i=1}^{n} X_i$ is a complete sufficient statistic.

Now, $\bar{X} \sim N\left(\theta, \frac{1}{n}\right)$

$\Rightarrow E(\bar{X}) = \theta$

$\Rightarrow h(T) = \bar{X} = \frac{T}{n}$ is the UMVUE of θ.

(ii) By L-S Theorem, and $Var(\bar{X}) = \frac{1}{n}$

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

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

$\Rightarrow h_1(T) = \bar{X}^2 - \frac{1}{n} = \left\{\left(\frac{T}{n}\right)^2 - \frac{1}{n}\right\}$ is the UMVUE of θ^2.

For a given sample, $\left(\bar{x}^2 - \frac{1}{n}\right)$ may give negative value in estimating a positive parameter θ^2.

Hence, the UMVUE is not <u>sensitive</u> (<u>or</u> is <u>absurd</u>).

*Ex. (6) :- Let $X_1, \ldots, X_n$ be a r.s. from $U(0,\theta)$, $\theta > 0$. Find the UMVUE of $E(X_1)$ and $Var(X_1)$.

Solution:- $E(X_1) = \theta/2$ and $Var(X_1) = \frac{\theta^2}{12}$.

It has already been shown that $T = X_{(n)}$ is complete sufficient.

Now, $$E(T^r) = \int_0^\theta t^r . \frac{n t^{n-1}}{\theta^n} dt$$

$$= \frac{n}{\theta^n} \int_0^\theta t^{r+n-1} dt$$

$$= \frac{n}{\theta^n} . \frac{\theta^{n+r}}{n+r}$$

$$= \frac{n}{n+r} . \theta^r$$

& $E(T) = \frac{n}{n+1} . \theta$

$$\Rightarrow E\left\{ \frac{n+1}{2n} . T \right\} = \frac{\theta}{2} = E(X_1)$$

and $E(T^2) = \frac{n}{n+2} . \theta^2$

$$\Rightarrow E\left\{ \frac{n+2}{12.n} . T^2 \right\} = \frac{\theta^2}{12} = V(X_1)$$

By L-S Theorem, $h_1(T) = \frac{n+1}{2n} . T$ and $h_2(T) = \frac{n+2}{12n} . T^2$ are the UMVUEs of $E(X_1)$ and $Var(X_1)$.

*Ex. (7) :- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $U(-\theta, \theta)$, $\theta > 0$. Find UMVUE of θ.

Hints:-

$$X_i \overset{iid}{\sim} U(-\theta, \theta)$$

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

Ex.(8):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $f(x;\sigma) = \frac{1}{2\sigma} e^{-\frac{|x|}{\sigma}}, x \in \mathbb{R}$ where $\sigma > 0$. Find UMVUE of σ^r.

Hints:- $\{f(x;\sigma) : \sigma > 0\}$ is an OPEF,

and $T = \sum_{i=1}^{n} |X_i|$ is completely sufficient.

Here $\frac{2T}{\sigma} \sim \chi^2_{2n}$

$$\therefore E\left\{\frac{2T}{\sigma}\right\}^r = E(\chi^2_{2n})^r = \frac{2^{n+r}\,\Gamma(n+r)}{2^n\,\Gamma(n)}$$

$$\Rightarrow E\left\{\frac{\Gamma(n)}{\Gamma(n+r)} \cdot T^r\right\} = \sigma^r \quad \text{if } r > n$$

*Ex.(9):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $f(x;p) = \begin{cases} p(1-p)^x, & x = 0,1,2,\ldots \\ 0, & \text{ow} \end{cases}$

S.T. UE of p based on $T = \sum_{i=1}^{n} X_i$ is unique. Hence or otherwise find the UMVUE of p.

Soln.:- $T = \sum_{i=1}^{n} X_i \sim NB(n, p)$

To solve for $h(T)$ such that

$$E\{h(T)\} = p \;\forall\; p \in (0,1)$$

$$\Rightarrow \sum_{t=0}^{\infty} h(t)\binom{t+n-1}{n-1} p^n q^t = p \quad \forall p$$

$$\Rightarrow \sum_{t=0}^{\infty} h(t) \cdot \binom{t+n-1}{n-1} q^t = p^{-(n-1)} = (1-q)^{n-1}$$

$$\Rightarrow \sum_{t=0}^{\infty} h(t)\binom{t+n-1}{n-1} q^t = \sum_{t=0}^{\infty} \binom{\overline{n-1}+t-1}{t} q^t \quad, \text{as } 0<q<1$$

By uniqueness property of Power series, we get

$$h(t)\binom{t+n-1}{n-1} = \binom{n+t-2}{t}, \; t = 0,1,2,\ldots$$

Hence $h(T) = \frac{n-1}{t+n-1}$ is the only solution of "$E\{h(T)\} = p$, $\forall p$".

$\Rightarrow h(T)$ is the only UE of p based on T.

It can be shown that $T = \sum_{i=1}^{n} x_i$ is sufficient.

By Rao-Blackwell theorem, UMVUE ~~of~~ is a function of T.

As there is only one UE of p based on T, then UE $h(T)$ is the UMVUE of p.

<u>Alternative:-</u> Define $h = \begin{cases} 1 & , \text{ if } X_1=0 \\ 0 & , \text{ ow} \end{cases}$

is an UE of $p = P[X_1=0]$.

Here $T=\sum_{i=1}^{n} X_i$ is complete sufficient.

By LS theorem, $E(h/T)$ is the UMVUE of p.

Now, $E\{h/T=t\} = 1.P[X_1=0/\sum_{i=1}^{n}X_i=t]$

$$= \frac{P[X_1=0 ; \sum_{i=2}^{n}X_i=t]}{P[\sum_{i=1}^{n}X_i=t]}$$

$$= \frac{p.\binom{t+n-2}{n-2}p^{n-1}q^t}{\binom{t+n-1}{n-1}p^n q^t}$$

$$= \frac{\binom{t+n-2}{n-2}}{\binom{t+n-1}{n-1}} = \frac{n-1}{t+n-1}.$$

Hence, $h(T) = \frac{n-1}{t+n-1}$ is the UMVUE of p.

***

<u>Ex.(10):-</u> Let $X_1, X_2, \dots, X_n$ be a r.s. from

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

Show that $T=X_{(1)}$ is a complete sufficient statistic. Hence find the UMVUE of θ.

<u>Solution:-</u>

The PDF of $\underset{\sim}{X} = (X_1, X_2, \dots, X_n)$ is

$$\prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} e^{-\sum_{i=1}^{n}(x_i-\theta)} & , \text{ if } x_i>\theta \,.\, \forall\, i=1(1)n \\ 0 & , \text{ ow} \end{cases}$$

$$= \begin{cases} e^{-\sum_{i=1}^{n}(x_i-\theta)} & , \text{ if } x_{(1)}>\theta \\ 0 & , \text{ ow} \end{cases}$$

$$= e^{-\sum_{i=1}^{n}(x_i-\theta)} . I(x_{(1)}, \theta) \text{, where } I(a,b) = \begin{cases} 1 & \text{if } a>b \\ 0 & \text{if } a<b \end{cases}$$

$$= e^{\theta}.I(x_{(1)},\theta) . e^{-\sum_{i=1}^{n}x_i}$$

$$= g(T(\underset{\sim}{x}), \theta) . h(\underset{\sim}{x}) \text{ with } T(\underset{\sim}{x}) = x_{(1)}.$$

By factorization criterion $T = X_{(1)}$ is sufficient.

Let $E\{h(T)\} = 0 \quad \forall \theta$

$\Rightarrow \int_{-\infty}^{\infty} h(t) \cdot f_T(t)\, dt = 0 \quad \forall \theta$

$\Rightarrow \int_{\theta}^{\infty} h(t) \cdot n e^{-n(t-\theta)}\, dt = 0 \;\forall \theta$

$\Rightarrow \int_{\theta}^{\infty} h(t) \cdot e^{-nt}\, dt = 0 \;\forall \theta$

$$\left[\begin{aligned} F_T(t) &= 1 - P[T > t] \\ &= 1 - P[X_{(1)} > t] \\ &= 1 - \{P[X_1 > t]\}^n \\ &= 1 - \left\{\int_t^{\infty} e^{-(x_1-\theta)} dx_1\right\}^n \quad \text{if } t > \theta \\ &= 1 - e^{-n(t-\theta)}, \text{ if } t > \theta \\ \therefore f_T(t) &= n e^{-n(t-\theta)}, \text{if } t > \theta \\ &= 0 \qquad\qquad , \text{ow} \end{aligned}\right]$$

Differentiating w.r.t. θ,

$0 - h(\theta) \cdot e^{-n\theta} = 0 \;\forall \theta$

$\Rightarrow h(\theta) = 0 \;\forall \theta$ as $e^{-n\theta} > 0$.

Hence, $h(T) = 0$, with prob. 1, $\forall \theta$.

$\Rightarrow T$ is complete.

Now, $E(T-\theta) = \int_{-\infty}^{\infty} (t-\theta) f_T(t)\, dt$

$= \int_{\theta}^{\infty} (t-\theta) n e^{-n(t-\theta)}\, dt$

$= \frac{1}{n} \int_0^{\infty} u e^{-u}\, du$, where $u = n(t-\theta)$

$= \frac{1}{n} \cdot \Gamma(2)$

$= \frac{1}{n}$

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

By LS Theorem, $h(T) = T - \frac{1}{n} = X_{(1)} - \frac{1}{n}$ is the UMVUE of θ.

*Ex.(11):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from

$$p(x;N) = \begin{cases} \frac{1}{N}, & x = 1, 2, \ldots, N \\ 0, & \text{ow} \end{cases}$$

Find UMVUE of N.

Solution:- It has been shown that $T = X_{(n)}$ is a complete sufficient statistic for this distn.

Method I:- $P[T=t] = \begin{cases} \dfrac{t^n - (t-1)^n}{N^n}, & t = 1(1)N \\ 0, & \text{ow} \end{cases}$

Consider the function $h(T) = \dfrac{T^{n+1} - (T-1)^{n+1}}{T^n - (T-1)^n}$

Now, $E\{h(T)\} = \sum_{t=1}^{N} h(t) \cdot P[T=t]$

$$= \sum_{t=1}^{N} \frac{t^{n+1} - (t-1)^{n+1}}{t^n - (t-1)^n} \times \frac{t^n - (t-1)^n}{N^n}$$

$$= \frac{1}{N^n} \cdot \sum_{t=1}^{N} \left\{ t^{n+1} - (t-1)^{n+1} \right\}$$

$$= \frac{N^{n+1} - 0^{n+1}}{N^n} = N$$

By L.S. theorem, $h(T)$ is the UMVUE of N.

Method II:- Conditioning Method:-

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

$\Rightarrow E(2X_1 - 1) = N$

Here $h = (2X_1 - 1)$ is an UE of N.

Now, By L.S theorem, $E(h/T)$ is the UMVUE of N.

Now, $E[h/T=t]$

$$= \sum_{x_1=1}^{t} (2x_1 - 1) \cdot P[X_1 = x_1 / T = t]$$

$$= \sum_{x_1=1}^{t} (2x_1 - 1) \cdot \frac{P[X_1 = x_1; X_{(n)} = t]}{P[X_{(n)} = t]}$$

[For $x_1 = 1(1)t-1$,

$$P[X_1 = x_1;\ X_{(n)} = t] = P\left[X_1 = x_1;\ \max_{i=2(1)n}\{X_i\} = t\right]$$

$$= P[X_1 = x_1] \cdot P\left[\max_{i=2(1)n}\{X_i\} = t\right]$$

$$= \frac{1}{N} \cdot \frac{t^{n-1} - (t-1)^{n-1}}{N^{n-1}}$$

For $x_1 = t$, $P[X_1 = x_1;\ X_{(n)} = t]$

$$= P\left[X_1 = t;\ \max_{i=2(1)n}\{x_i\} \le t\right]$$

$$= P[X_1 = t]\ P\left[\max_{i=2(1)n}\{x_i\} \le t\right]$$

$$= \frac{1}{N} \cdot \left(\frac{t}{N}\right)^{n-1}$$]

Hence, $E(h/T=t)$

$$= \sum_{x_1=1}^{t-1} (2x_1 - 1) \cdot \frac{t^{n-1} - (t-1)^{n-1}}{t^n - (t-1)^n} + (2t-1)\frac{t^{n-1}}{t^n - (t-1)^n}$$

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

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

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

Hence, $E(h/T) = \dfrac{T^{n+1} - (T-1)^{n+1}}{T^n - (T-1)^n}$ is the UMVUE of N.

Ex. (12):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from the PDF

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

Find the UMVUE of (i) $\frac{1}{\theta}$, (ii) θ, (iii) $P[X_1 > K] = 1 - F_{X_1}(K)$.

Solution:-

Note that the family $\{f(x;\theta) : \theta > 0\}$ is an OPEF with

$$f(x;\theta) = \exp[-\theta x + \ln\theta]$$
$$= \exp[u(\theta) \cdot T(x) + v(\theta) + w(x)] \text{ with } T(x) = x.$$

$\therefore T = \sum_{i=1}^{n} T(X_i) = \sum_{i=1}^{n} X_i$ is complete sufficient.

(i) $E(T) = E\left(\sum_{i=1}^{n} X_i\right) = \frac{n}{\theta}$

$\Rightarrow E\left(\frac{T}{n}\right) = \frac{1}{\theta}$

$\therefore$ By L-S Theorem, $h_1(T) = \frac{T}{n} = \bar{X}$ is the UMVUE of $\frac{1}{\theta}$.

(ii) To find an UE of θ, we should try with the statistic $\frac{1}{T}$.

$$E\left(\frac{1}{T}\right) = \int_0^\infty \frac{1}{t} \cdot \frac{\theta^n}{\Gamma n} \cdot e^{-\theta t} \cdot t^{n-1} dt \quad \left[\because T = \sum_{i=1}^{n} X_i \sim \text{Gamma}(\theta, n)\right]$$

$$= \frac{\theta^n}{\Gamma n} \int_0^\infty e^{-\theta t}\, t^{(n-1)-1} dt$$

$$= \frac{\theta^n}{\Gamma n} \cdot \frac{\Gamma(n-1)}{\theta^{n-1}} \quad \text{if } (n-1) > 0$$

$$= \frac{\theta}{n-1} \quad \text{if } n > 1.$$

$\Rightarrow E\left(\frac{n-1}{T}\right) = \theta$.

$\therefore$ By L-S theorem, $h_2(T) = \frac{n-1}{T} = \frac{n-1}{n\bar{X}}$ is the UMVUE of θ.

(iii) Here $\psi(\theta) = P[X_1 > K] = 1 - F_{X_1}(K)$
$= e^{-\theta K}$, $K > 0$

Define, $h = \begin{cases} 1 & , \text{ if } X_1 > K \\ 0 & , \text{ ow} \end{cases}$

is an UE of $\psi(\theta)$.

By L-S theorem, $E[h/T]$ is the UMVUE of $\psi(\theta) = P[X_1 > K]$

Now, $E[h/T=t] = 1 . P[X_1 > k/T=t]$

$$= \int_k^{\infty} f_{X_1/T}(x_1/t)\,dx_1$$

Now, note that $f_{x_1/T}(x_1/t) = \dfrac{f_{x_1,T}(x_1,t)}{f_T(t)}$

$$= \frac{f_{X_1, T-X_1}(x_1, t-x_1)}{f_T(t)}$$

[Using the transformation $(X_1, T) \longrightarrow (X_1, T-X_1)$]

$$= \frac{f_{X_1}(x_1)\, f_{T-X_1}(t-x_1)}{f_T(t)}$$

$$= \frac{\theta e^{-\theta x_1} \cdot \frac{\theta^{n-1}}{\Gamma(n-1)} \cdot e^{-\theta(t-x_1)} \cdot (t-x_1)^{n-2}}{\frac{\theta^n}{\Gamma n} \cdot e^{-\theta t} \cdot t^{n-1}} \text{ if } (t-x_1)>0$$

$$= \frac{(n-1)(t-x_1)^{n-2}}{t^{n-1}} \text{ if } x_1 < t$$

Now, $E[h/T=t] = \displaystyle\int_k^{\infty} \frac{(n-1)(t-x_1)^{n-2}}{t^{n-1}}\,dx_1$ if $t > k$

$$= \frac{n-1}{t^{n-1}} \left[-\frac{(t-x_1)^{n-1}}{n-1} \right]_k^t$$

$$= \frac{(t-k)^{n-1}}{t^{n-1}}, \; t > k$$

$$= \left(1 - \frac{k}{t}\right)^{n-1}, \; t > k$$

Hence the UMVUE of $\psi(\theta) = P[X_1 > k]$ is

$$E[h/T] = \begin{cases} \left(1 - \frac{k}{t}\right)^{n-1}, & \text{if } t > k \\ 0, & \text{if } t \le k \end{cases}$$

Scanned by CamScanner

*Ex. (13):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(\theta,1)$.
Let $p = \Phi(k-\theta)$. Find the UMVUE of p.

Solution:-

The family $\{N(\theta,1) : \theta \in \mathbb{R}\}$ is an OPEF with

$$f(x;\theta) = \exp\left[-\frac{x^2}{2} + \theta x - \frac{1}{2}\{\theta^2 + \ln(2\pi)\}\right]$$

$$= \exp\left[u(\theta) \cdot T(x) + v(\theta) + w(x)\right] \text{ with } T(x) = x,$$

and $T = \sum_{i=1}^{n} X_i$ or $\overline{X}$ is complete sufficient.

Here $p = \Phi(k-\theta) = P[X_1 \le k]$

$\left[\begin{array}{l} P[X_1 - \theta \le k-\theta] \\ = \Phi(k-\theta), \\ \text{since } (X_1-\theta) \overset{a}{\sim} N(0,1) \end{array} \right]$

Here $h = \begin{cases} 1 & \text{if } X_1 \le k \\ 0 & \text{, ow} \end{cases}$

is an UE of $p = P[X_1 \le k]$

By LS theorem, $E(h/\overline{X})$ is the UMVUE of $p = \Phi(k-\theta)$.

Now, $E[h/\overline{X} = \bar{x}]$

$$= 1 \cdot P[X_1 \le k / \overline{X} = \bar{x}]$$

$$= P[X_1 - \overline{X} \le k - \bar{x} / \overline{X} = \bar{x}]$$

$\Big[$ $X_i \sim N(\theta,1) \Rightarrow \overline{X} \sim N\left(\theta, \frac{1}{n}\right)$

Now, $(X_1 - \overline{X}, \overline{X}) \sim BN\left(0, \theta, 1-\frac{1}{n}, \frac{1}{n}, \rho = 0\right)$

Here $Var(X_1 - \overline{X}) = V(X_1) + V(\overline{X}) - 2\,cov(X_1, \overline{X})$

$$= 1 + \frac{1}{n} - 2\,Cov\left(X_1, \frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$

$$= 1 + \frac{1}{n} - \frac{2}{n} V(X_1)$$

$$= 1 + \frac{1}{n} - \frac{2}{n}$$

$$= 1 - \frac{1}{n}$$

and $Cov(X_1 - \overline{X}, \overline{X}) = Cov(X_1, \overline{X}) - V(\overline{X})$

$$= \frac{1}{n} - \frac{1}{n} = 0$$

Here, $X_1 - \overline{X}$ and $\overline{X}$ are independently distributed,
and $X_1 - \overline{X} \sim N\left(0, 1-\frac{1}{n}\right)$. $\Big]$

$$\therefore\ E\left[h/\bar{X}=\bar{x}\right] = P\left[X_1-\bar{X} \le k-\bar{x}\right]$$

$$= P\left[\frac{(X_1-\bar{X})-0}{\sqrt{1-\frac{1}{n}}} \le (k-\bar{x})\sqrt{\frac{n}{n-1}}\right]$$

$$= \Phi\left[(k-\bar{x})\sqrt{\frac{n}{n-1}}\right]$$

Hence, $\Phi\left[\sqrt{\frac{n}{n-1}}\,(k-\bar{x})\right]$ is the UMVUE of $p=\Phi(k-\theta)$.

Ex.(14):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(\mu,\sigma^2)$. Find UMVUE of (i) μ, σ (ii) σ^r, (iii) $\frac{\mu}{\sigma}$, (iv) the p^{th} quantile of X_1 $[\xi_p]$

Solution:- The family $\{N(\mu,\sigma^2) : \mu\in\mathbb{R}, \sigma>0\}$ is a two-parameter exponential family of distn.s and $T=(\bar{X}, s^2)$ is a complete sufficient.

(i) $E(\bar{X})=\mu$ and $E(s^2)=\sigma^2$

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

Hence $\bar{X}$ and s^2 are the UMVUE of μ and σ^2 respectively.

(ii) Note that $\frac{(n-1)s^2}{\sigma^2}\sim\chi^2_{n-1}$.

$$\text{Now, } E\left\{\frac{(n-1)s^2}{\sigma^2}\right\}^{r/2} = E\left(\chi^2_{n-1}\right)^{r/2}$$

$$= \frac{2^{r/2}\,\Gamma\left(\frac{n-1}{2}+\frac{r}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}, \text{ if } r>-(n-1)$$

$$\therefore\ E\left\{\frac{\Gamma\left(\frac{n-1}{2}\right)(n-1)^{r/2}}{2^{r/2}\cdot\Gamma\left(\frac{n-1+r}{2}\right)}\cdot s^r\right\} = \sigma^r$$

$$\Rightarrow E\left(k_{n-1,r}\cdot s^r\right) = \sigma^r \quad \text{if } r>-(n-1).$$

Hence, $\left(k_{n-1,r}\cdot s^r\right)$ is the UMVUE of σ^r; $r>-(n-1)$.

(iii) Note that $E(\bar{X}) = \mu$ and

$$E\left(K_{n-1,-1} \cdot S^{-1}\right) = \sigma^{-1}$$

Hence, $E\left(\bar{X} \cdot K_{n-1,-1} \cdot S^{-1}\right) = E(\bar{X}) \cdot E\left(K_{n-1,-1} \cdot S^{-1}\right)$ due to independence of $\bar{X}$ and S^2.

$$= \frac{\mu}{\sigma}.$$

Hence, $K_{n-1,-1}\left(\frac{\bar{X}}{S}\right)$ is the UMVUE of $\frac{\mu}{\sigma}$.

(iv) By definition,

$$p = P\left[X_1 \le \xi_p\right] = P\left[\frac{X_1 - \mu}{\sigma} \le \frac{\xi_p - \mu}{\sigma}\right]$$

$$= \Phi\left(\frac{\xi_p - \mu}{\sigma}\right)$$

$\Rightarrow \Phi\left(\frac{\xi_p - \mu}{\sigma}\right) = p = \Phi\left(Z_p\right)$, where Z_p is the pth quantile of $N(0,1)$.

$$\Rightarrow \frac{\xi_p - \mu}{\sigma} = z_p$$

$$\Rightarrow \xi_p = \mu + \sigma z_p.$$

Note that, $E\left(\bar{X} + Z_p \cdot K_{n-1,1} \cdot S\right)$

$$= E(\bar{X}) + Z_p \cdot E\left(K_{n-1,1} \cdot S\right)$$

$$= \mu + Z_p \cdot \sigma$$

$$= \xi_p.$$

By L-S theorem, $\left\{\bar{X} + Z_p\left(K_{n-1,1}\right) \cdot S\right\}$ is the UMVUE of ξ_p.

Ex. (15):- Let $X_1, \ldots, X_n$ be a r.s. from

$$f(x;\theta) = \begin{cases} \frac{\theta}{x^{\theta+1}}, & \text{if } x > 1 \\ 0, & \text{o/w}; \text{ where } \theta > 0 \end{cases}$$

Find the UMVUE of (i) θ, (ii) $\frac{1}{\theta^r}$.

(Do yourself)

* **Ex.(16):-** Let $X_1, X_2, \ldots, X_n$ be a r.s. from $U(\theta_1, \theta_2)$. Find the UMVUE of $\frac{\theta_1+\theta_2}{2}$ and $\frac{\theta_1-\theta_2}{2}$.

Solution:- Here $T = (X_{(1)}, X_{(n)})$ is sufficient for the family.

Let, $E\{h(T)\} = 0 \quad \forall \ \theta_1 < \theta_2$

$$E\{h(X_{(1)}, X_{(n)})\} = 0 \ \forall \ \theta_1 < \theta_2$$

$$\Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(y_1, y_2) f_{X_{(1)}, X_{(n)}}(y_1, y_2)\, dy_1\, dy_2 = 0 \quad \forall \ \theta_1 < \theta_2$$

$$\Rightarrow \int_{\theta_1}^{\theta_2}\left\{\int_{\theta_1}^{y_2} h(y_1, y_2) \cdot \frac{n(n-1)(y_2-y_1)^{n-2}}{(\theta_2-\theta_1)^n}\, dy_1\right\} dy_2 = 0 \quad \forall \ \theta_1 < \theta_2$$

$$\Rightarrow \int_{\theta_1}^{\theta_2}\left\{\int_{\theta_1}^{y_2} h(y_1, y_2)(y_2-y_1)^{n-2}\, dy_1\right\} dy_2 = 0 \ \forall \ \theta_1 < \theta_2$$

Differentiating w.r.t. θ_2, we get,

$$\int_{\theta_1}^{\theta_2} h(y_1, \theta_2)(\theta_2-y_1)^{n-2}\, dy_1 = 0 \ \forall \ \theta_1$$

Differentiating w.r.t. θ_1, we get,

$$0 - h(\theta_1, \theta_2) \cdot (\theta_2-\theta_1)^{n-2} = 0 \quad \forall \ \theta_1 < \theta_2$$

$$\Rightarrow h(\theta_1, \theta_2) = 0 \ \forall \ \theta_1 < \theta_2$$

i.e. $h(y_1, y_2) = 0 \ \forall \ y_1 < y_2$

Hence, $T = (X_{(1)}, X_{(2)})$ is complete.

Now, $E(X_{(1)}) = \theta_1 + \frac{\theta_2-\theta_1}{n+1}$

$$E(X_{(n)}) = \theta_2 - \frac{\theta_2-\theta_1}{n+1}$$

$$\Rightarrow E\left(\frac{X_{(1)}+X_{(n)}}{2}\right) = \frac{\theta_1+\theta_2}{2}$$

and $E\left\{\frac{n+1}{2(n-1)}\left(X_{(1)}-X_{(n)}\right)\right\} = \frac{\theta_2-\theta_1}{2}$

Hence, $\frac{X_{(1)}+X_{(n)}}{2}$ and $\frac{n+1}{n-1} \cdot \frac{(X_{(1)}-X_{(n)})}{2}$ are UMVUE of $\frac{\theta_1+\theta_2}{2}$ and $\frac{\theta_2-\theta_1}{2}$, respectively.

Cramer-Rao Lower Bound for the variance of an Unbiased Estimator:

Let X be a RV with PDF or PMF belonging to the family $\{f(x;\theta): \theta \in \Omega\}$ satisfies the following regularity conditions:

(i) The parameter space Ω is an open interval.

(ii) The support $S = \{x : f(x,\theta) > 0\}$ does not depend on θ.

(iii) For $x \in S$, $\theta \in \Omega$, the derivative $\frac{\partial}{\partial\theta} f(x;\theta)$ exists and is finite.

(iv) The identity $\int_S f(x,\theta)\,dx = 1$ $\left[\text{or}, \sum_{x \in S} f(x,\theta) = 1\right]$

can be differentiated under integral [or, summation] sign.

(v) $T(x) \in U_\psi = \{T(x) : E(T(X)) = \psi(\theta),\ Var(T(X)) < \infty, \forall \theta \in \Omega\}$ is any statistic for which the derivative w.r.t. θ of $\psi(\theta) = E\{T(X)\}$ exists and can be evaluated by differentiating under the integral (or, summation) sign in $E\{T(X)\}$.

Then $Var(T(X)) \geq \dfrac{\{\psi'(\theta)\}^2}{I(\theta)}$; where,

$$I(\theta) = E\left[\frac{\partial}{\partial\theta} \ln f(X,\theta)\right]^2 > 0.$$

Proof:- Let X be a continuous RV with PDF $f(x;\theta)$, $\theta \in \Omega$. Differentiating the identity $\int_S f(x,\theta)\,dx = 1$, w.r.t. θ, we get,

$$\frac{\partial}{\partial\theta}\int_S f(x,\theta)\,dx = \frac{d}{d\theta}(1).$$

$$\Rightarrow \int_S \frac{\partial}{\partial\theta} f(x,\theta)\,dx = 0 \quad , \text{by (ii)}$$

$$\Rightarrow \int_S \frac{d}{d\theta}\{\ln f(x,\theta)\}.f(x,\theta)\,dx = 0$$

$$\Rightarrow E\left(\frac{\partial}{\partial\theta} \ln f(x,\theta)\right) = 0.$$

Differentiating $\psi(\theta) = E\{T(X)\}$ w.r.t. θ, we have, —

$$\psi'(\theta) = \frac{\partial}{\partial\theta} E\{T(X)\}$$
$$= \frac{\partial}{\partial\theta} \cdot \int_S T(X) \cdot f(x,\theta)\,dx$$
$$= \int_S T(X) \cdot \frac{\partial}{\partial\theta} f(x,\theta)\,dx, \text{ by (v).}$$
$$= \int_S T(X) \cdot \frac{\partial}{\partial\theta} \{\ln f(x,\theta)\} \cdot f(x,\theta)\,dx$$
$$= E\left[T(X) \cdot \frac{\partial}{\partial\theta} \ln f(x,\theta)\right]$$
$$= Cov\left[T(X), \frac{\partial}{\partial\theta} \ln f(x,\theta)\right] \text{ as } E\left[\frac{\partial}{\partial\theta} \ln f(x,\theta)\right] = 0$$

Now, $\{\psi'(\theta)\}^2 = Cov^2\left[T(X), \frac{\partial}{\partial\theta} \ln f(x,\theta)\right]$
$$\le Var(T(X)) \cdot Var\left[\frac{\partial}{\partial\theta} \ln f(x,\theta)\right], \text{ by C-S inequality.}$$

$\Rightarrow Var(T(X)) \geq \dfrac{\{\psi'(\theta)\}^2}{I(\theta)}$; where

$$I(\theta) = Var\left[\frac{\partial}{\partial\theta} \ln f(x;\theta)\right]$$
$$= E\left[\frac{\partial}{\partial\theta} \ln f(x;\theta)\right]^2$$

Remarks:-

(1) The inequality " $Var(T) \geq \dfrac{\{\psi'(\theta)\}^2}{I(\theta)}$ with the regularity conditions (i) - (v) " is called the Cramer-Rao inequality, and then the RHS $= \dfrac{\{\psi'(\theta)\}^2}{I(\theta)}$ is called the Cramer-Rao Lower Bound for the variance of an UE of $\psi(\theta)$.

(2) Cramer-Rao inequality can also be expressed as
$$Var(T(X)) \geq \frac{\left\{\frac{\partial}{\partial\theta} E[T(X)]\right\}^2}{I(\theta)},$$ where $T(X)$ is any statistic with $Var(T(X)) < \infty$, which provides the lower bound of the variance of an UE of $E\{T(X)\}$.
Here $T(X)$ is not necessarily unbiased for $\psi(\theta)$.
Let $ET(X) = \psi(\theta) + b(\theta)$, then
$$Var(T(X)) \geq \frac{\{\psi'(\theta) + b'(\theta)\}^2}{I(\theta)}.$$

(3) Let $X_1, \ldots, X_n$ be a r.s. from $f(x;\theta)$, $\theta \in \Omega$. Then the PDF of $\underset{\sim}{X} = (X_1, \ldots, X_n)$ is

$$L(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

Then $I_{\underset{\sim}{X}}(\theta) = n \cdot I_{X_1}(\theta)$

Proof:-

(4) If, in addition to the regularity condition (i) to (v). The 2nd derivative w.r.t. θ of $\ln f(x;\theta)$ exists and 2nd derivative w.r.t. θ of $\int_S f(x;\theta)dx = 1$ $\left(\text{or}, \sum_{x \in S} f(x;\theta) = 1\right)$ can be obtained by differentiating twice under the integral or summation sign, $I(\theta) = E\left[-\frac{\partial^2}{\partial\theta^2} \ln f(x;\theta)\right]$.

Proof:- We have $E\left(\frac{\partial}{\partial\theta} \ln f(x,\theta)\right) = 0$

$$\Rightarrow \int_S \frac{\partial}{\partial\theta} \ln f(x,\theta) \cdot f(x,\theta)\,dx = 0$$

Differentiating w.r.t. θ, we get, —

$$\int_S \frac{\partial^2}{\partial\theta^2} \ln f(x,\theta) \cdot f(x,\theta)\,dx + \int_S \left\{\frac{\partial}{\partial\theta} \ln f(x,\theta)\right\}^2 f(x,\theta)\,dx = 0$$

$$\Rightarrow I(\theta) = E\left(\frac{\partial}{\partial\theta} \ln f(x,\theta)\right)^2 = E\left(-\frac{\partial^2}{\partial\theta^2} \ln f(x,\theta)\right)$$

$$\left[\because \frac{\partial}{\partial\theta} f(x,\theta) = \frac{\partial}{\partial\theta} \ln f(x,\theta) \cdot f(x,\theta)\right]$$

(5) Fisher's Information :

The Fisher's Information about θ in a RV X from a PDF or PMF $f(x,\theta)$, $\theta \in \Omega$, is given by $I_X(\theta) = E\left(\frac{\partial}{\partial\theta} \ln f(x,\theta)\right)^2$

Justification :-

Note that $\frac{\partial}{\partial\theta} \ln f(x;\theta)$ is the rate of change of log-likelihood of the values x w.r.t. θ. The function $\frac{\partial}{\partial\theta} \ln f(x;\theta)$ viewed as a function of x for fixed θ, is called the score function and for each θ.

$\frac{\partial}{\partial\theta} \ln f(x;\theta)$ is a R.V., with PDF or PMF $f(x;\theta)$.

We want a measure of average rate of change of the log-likelihood w.r.t. θ, but $E\left(\frac{\partial}{\partial\theta} \ln f(x;\theta)\right) = 0$, can't be used as a measure. Then, ignoring the sign, $E\left(\frac{\partial}{\partial\theta} \ln f(x;\theta)\right)^2$ can be used as a measure of sensitivity of the log likelihood w.r.t. θ or the amount of information about θ in X.

In this sense, $I(\theta)$ gives a measure of information about θ contains in X.

Exercise :- The Fisher information about θ in a statistic T is always less than or equal to that in the original sample. Again, there is no loss of information iff T is sufficient.

Solution :-

Equality in CR inequality :—

Suppose that the family of distns $\{f(x;\theta) : \theta \in \Omega\}$ satisfies all the regularity conditions (i) – (v), then $T \in U_\psi$ attains CRLB iff

$$Var(T) = \frac{\{\psi'(\theta)\}^2}{I(\theta)}$$

$$\text{iff } Cov\left(T, \frac{\partial}{\partial\theta} \ln f(x;\theta)\right)^2 = V(T) \cdot I(\theta)$$

$$\text{iff } \left\{ \frac{Cov\left[T, \frac{\partial}{\partial\theta} \ln f(x;\theta)\right]}{\sqrt{V(T)}\sqrt{V\left(\frac{\partial}{\partial\theta} \ln f(x;\theta)\right)}} \right\}^2 = 1$$

iff the correlation coefficient between T and $\frac{\partial}{\partial\theta} \ln f(x;\theta)$ is ± 1.

$$\text{iff } \frac{\frac{\partial}{\partial\theta} \ln f(x;\theta) - E\left(\frac{\partial}{\partial\theta} \ln f(x;\theta)\right)}{\sqrt{V\left(\frac{\partial}{\partial\theta} \ln f(x;\theta)\right)}} = \pm \frac{T - E(T)}{\sqrt{V(T)}} .$$

$$\text{iff } \frac{\frac{\partial}{\partial\theta} \ln f(x;\theta) - 0}{\sqrt{I(\theta)}} = \pm \frac{T - E(T)}{\psi'(\theta)/\sqrt{I(\theta)}}$$

$$\text{iff } \frac{\partial}{\partial\theta} \ln f(x;\theta) = \pm \frac{I(\theta)}{\psi'(\theta)} \{T - E(T)\} \longrightarrow (*)$$

This is the necessary and sufficient condition for attaining the CRLB by the UE T of $\psi(\theta)$.

<u>Remark</u>:- From (*),

$$\frac{\partial}{\partial\theta} \ln f(x;\theta) = k(\theta)\{T - \psi(\theta)\}, \text{ say.}$$

Assuming $k(\theta)$, $k(\theta)\psi(\theta)$ are integrable with respect to θ, and integrating w.r.t. θ, we get

$$\ln f(x;\theta) = T\int k(\theta)d\theta - \int k(\theta)\psi(\theta)d\theta$$

$$\Rightarrow \ln f(x;\theta) = u(\theta)T + v(\theta) + w(x)$$

as "the equality case" in the CR inequality.

Under suitable regularity conditions, CRLB is attained by T iff the family of distns $\{f(x;\theta) : \theta \in \Omega\}$ is an OPEF with

$$f(x;\theta) = \exp\left[u(\theta) \cdot T(x) + v(\theta) + w(x)\right]$$

Definition:- MVBUE/OR/BRUE

Let the family $\{f(x;\theta) : \theta \in \Omega\}$ of distributions satisfies all the regularity conditions (i)-(v), then an UE $T \in U_\psi$, with $Var(T) = \dfrac{\{\psi'(\theta)\}^2}{I(\theta)}$

i.e., T is an UE of $\psi(\theta)$ which attains CRLB is called Minimum Variance Bound Unbiased Estimator (MVBUE) or Best Regular Unbiased Estimator (BRUE).

In this case, the MVBUE has the minimum variance among all UEs of $\psi(\theta)$, i.e. T is UMVUE of $\psi(\theta)$.

Remark:- T is UMVUE iff T attains CRLB iff
$$\ln f(x,\theta) = u(\theta).T(x) + v(\theta) + \omega(x).$$

Hence a MVBUE $T(x)$ is a complete sufficient statistic and is the UMVUE of $E(T) = \psi(\theta)$, say.

It follows that even if OPEF the only parametric function which admits a UMVUE whose variance attains the CRLB is the functions $\psi(\theta) = E(T)$, where T is a complete sufficient statistic.

Ex.(1):- Let $X_1, \ldots, X_n$ be a r.s. from $B(1,p)$, then obtain the CRLB for the variance of an UE of $\psi(p) = p$. Hence obtain the UMVUE of p.

Solution:- The PMF of $\underset{\sim}{X} = (X_1, \ldots, X_n)$ is
$$f(x,p) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \quad \text{if } x_i = 0, 1, \ldots \quad \forall\, i = 1(1)n.$$
where $p \in \Omega = \{p : 0<p<1\}$

Clearly, the family $\{f(x,p) : p \in \Omega\}$ is an OPEF and it satisfies all the regularity conditions (i)-(v).

Then, for any UE of $\psi(p) = p$
$$V(T) = \frac{\{\psi'(p)\}^2}{I(p)} = \frac{1}{I(p)} = \text{CRLB},$$
where, $I(p) = E\left(-\dfrac{\partial^2}{\partial p^2} \ln f(x,p)\right)$

Now, $\ln f(x,p) = \sum_{i=1}^{n} x_i \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p)$

$$\frac{\partial}{\partial p} \ln f(x,p) = \frac{\Sigma x_i}{p} + \frac{n - \Sigma x_i}{1-p}(-1)$$

and $\frac{\partial^2}{\partial p^2} \ln f(x,p) = -\frac{n\bar{x}}{p^2} - \frac{n(1-\bar{x})}{(1-p)^2}$

Hence, $I(p) = E\left\{-\frac{\partial^2}{\partial p^2} \ln f(x,p)\right\}$

$$= \frac{nE(\bar{X})}{p^2} + \frac{n\{1-E(\bar{x})\}}{(1-p)^2}$$

$\left[\because X_i \sim B(1,p)\right.$
$\Sigma X_i \sim Bin(n,p)$
$\therefore E(\Sigma X_i) = np$
$\left.\therefore E(\bar{X}) = p\right]$

$$= \frac{n}{p} + \frac{n}{1-p}$$

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

Hence the CR inequality reduces to

$$var(T) \geq \frac{p(1-p)}{n} = CRLB$$

As $var(\bar{X}) = \frac{p(1-p)}{n}$, it follows that variance of $\bar{X}$, attains CRLB and $\bar{X}$ has the minimum variance among all UE of $\psi(p) = p$.

Hence, $\bar{X}$ is UMVUE of $\psi(p) = p$.

<u>Ex.(2):-</u> Let $X_1, \ldots, X_n$ be a r.s. from $P(\lambda)$. Obtain the CRLB for the variance of an UE of $\psi(\lambda) = \lambda$. Hence find UMVUE of λ.

*
Ex.(3):- An example where CRLB is not attained by the variance of an UE, or, an example of a UMVUE whose variance does not attain CRLB.

Solution:- Let $X \sim P(\lambda)$

Consider the problem of estimation of $\psi(\lambda) = e^{-\lambda}$ based on a single observation X.

Clearly, the family $\{P(\lambda) : \lambda > 0\}$ is an OPEF and it satisfies all the regularity conditions required for CR inequality, then for any UE T of $\psi(\lambda) = e^{-\lambda}$,

$$Var(T) \geq \frac{\{\psi'(\lambda)\}^2}{I(\lambda)} = CRLB$$

Note that, $f(x, \lambda) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$

$$\ln f(x,\lambda) = -\lambda + x\ln\lambda - \ln x!$$

$$\frac{\partial}{\partial\lambda} \ln f(x,\lambda) = -1 + \frac{x}{\lambda}$$

$$\frac{\partial^2}{\partial\lambda^2} \ln f(x,\lambda) = -\frac{x}{\lambda^2}$$

$$I(\lambda) = E\left(-\frac{\partial^2}{\partial\lambda^2} \ln f(x,\lambda)\right)$$

$$= \frac{E(X)}{\lambda^2} = \frac{1}{\lambda}.$$

Then CR inequality reduces to $V(T) = \dfrac{e^{-2\lambda}}{\frac{1}{\lambda}} = \lambda e^{-2\lambda} = CRLB$.

If an UE T attains CRLB, that is (the MVBUE, if exists for $\psi(\lambda) = e^{-\lambda}$) given by $T = \psi(\lambda) \pm \dfrac{\psi'(\lambda)}{I(\lambda)} \cdot \dfrac{\partial}{\partial\lambda} \ln f(x,\lambda)$

$$= e^{-\lambda} \pm \frac{-e^{-\lambda}}{1/\lambda}\left(-1 + \frac{x}{\lambda}\right)$$

$$= e^{-\lambda} \mp e^{-\lambda}(x - \lambda)$$

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

whether we take +ve or −ve sign, T is a function of X and λ. Hence, it's not a statistic.

Thus there does not exist a statistic which attains CRLB, that is in this case CRLB is not an attainable lower bound.

☒ Note that $T = X$ is a complete sufficient statistic.

here $\psi(\lambda) = e^{-\lambda} = P[X=0]$

then $h(X) = \begin{cases} 1 & \text{if } X=0 \\ 0 & \text{ow} \end{cases}$

is an UE of $P[X=0] = e^{-\lambda}$.

By L-S theorem, $h(X)$ is the UMVUE of $\psi(\lambda) = e^{-\lambda}$.

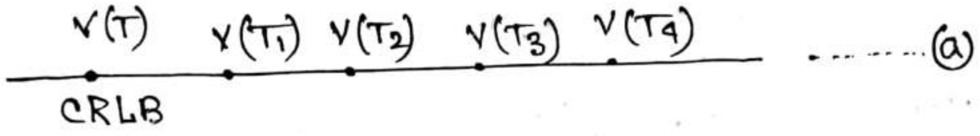
Now, $V(h(X)) = E\{h^2(X)\} - E^2\{h(X)\}$

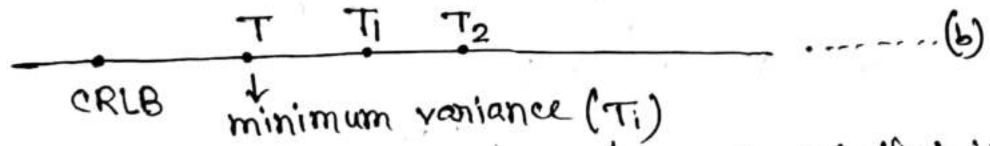
$$= 1^2 . P[X=0] - [1 . P[X=0]]^2$$

$$= e^{-\lambda} - e^{-2\lambda} = e^{-2\lambda}(e^{\lambda}-1)$$

$$= e^{-2\lambda}\left\{\frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots\cdots\right\}$$

$$> \lambda e^{-2\lambda} = \text{CRLB}$$





In general, CRLB is not attainable lower bound, that is, in a case, satisfying the regularity conditions (i)-(v), an UMVUE may not exist. Therefore the variance of UMVUE, whose variance is the least attainable variance in the class of unbiased estimators, exceeds the CRLB.

Part (a): —

Here $V(T) = \text{CRLB}$, therefore T is MVBUE as well as UMVUE.

Part (b): — Here, there is no T for which $V(T)$ is CRLB $\Rightarrow$ there does not exist an MVBUE

$\Rightarrow$ Variance UMVUE $>$ CRLB.

Ex.(4):— Let $X_1, X_2, \ldots, X_n$ be a r.s. from $U(0,\theta), \theta>0$.
Compute the quantity $\dfrac{1}{nE\left(\frac{\partial}{\partial\theta}\ln f(X_1,\theta)\right)^2}$.

Also, obtain the variance of the UEs $\frac{n+1}{n}X_{(n)}$ or $2\bar{X}$. Compute their variance with the above quantity and comment.

Solution:— Here $f(x_1,\theta) = \begin{cases} \frac{1}{\theta}, & 0<x<\theta \\ 0, & \text{OW} \end{cases}$; where $\theta>0$

$$\ln f(x_1,\theta) = -\ln\theta, \quad 0<x_1<\theta$$

and $\frac{\partial}{\partial\theta}\ln f(x_1,\theta) = -\frac{1}{\theta}$

Hence, $E\left(\frac{\partial}{\partial\theta}\ln f(x_1,\theta)\right)^2 = E\left(-\frac{1}{\theta}\right)^2$
$= \theta^{-2}$

$$\therefore \frac{1}{nE\left(\frac{\partial}{\partial\theta}\ln f(x_1,\theta)\right)^2} = \frac{\theta^2}{n}.$$

Note that $\text{Var}\left(\frac{n+1}{n}X_{(n)}\right) = \frac{\theta^2}{n(n+2)}$

and $\text{Var}(2\bar{X}) = 4\,\text{Var}(\bar{X})$
$= 4\cdot\frac{V(X_1)}{n}$
$= \frac{4}{n}\cdot\frac{\theta^2}{12} = \frac{\theta^2}{3n}$

Here, $\text{Var}\left(\frac{n+1}{n}X_{(n)}\right)$ or $V(2\bar{X})$ is less than the given quantity.

Comment:- The family $\{U(0,\theta):\theta>0\}$ does not satisfy the regularity condition (ii) & (iv), since the support $S=\{x: f(x,\theta)>0\} = (0,\theta)$ depends on θ and

" $\frac{\partial}{\partial\theta}\int_S f(x,\theta)\,dx = \int_S \frac{\partial}{\partial\theta} f(x,\theta)\,dx$ "

$\Rightarrow \frac{\partial}{\partial\theta}(1) = \int_0^\theta \frac{\partial}{\partial\theta}\left(\frac{1}{\theta}\right)dx$

$\Rightarrow 0 = -\int_0^\theta \frac{1}{\theta^2}\,dx = -\frac{1}{\theta}$, not possible.

Hence, CR inequality does not exist in the non-regular case, the variance of UMVUE or any other UE may be lower than the quantity $\frac{1}{nE\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2}$ in that non-regular case.

[which, when CR inequality exists, is CRLB]

Ex.(5):- Let $X_1, \ldots, X_n$ be a r.s. from

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

Compute $\dfrac{1}{nE\left(\frac{\partial}{\partial\theta} \ln f(x,\theta)\right)^2}$.

Also find the variance of an UE of θ based on $X_{(1)}$. which one is smaller? Give reasons.

Remark:– Regularity condition (ii) is unnecessarily restrictive. An examination of the proof shows that it is only necessary that (i), (iii) to (v) holds for the CR inequality. Condition (ii) excludes the distributions such as

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

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

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

Note that for (a) and (c), condition (iv) fails to hold. For (b), condition (v) fails to hold.

Ex. (6):– Are the following families of distn.s regular in the sense of Cramer & Rao?
If so, find the lower bound for the variance of an UE of θ based on a sample of size n. Also, find the UMVUEs of θ.

(a) $f(x,\theta) = \dfrac{e^{-\frac{x^2}{2\theta}}}{\sqrt{2\pi\theta}}$; $-\infty<x<\infty$, $\theta>0$

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

Solution:–

(a) As we know that '=' holds in CR inequality, whenever the family of distributions is OPEF.
The given PDF is OPEF and it satisfies the regularity conditions for CR inequality that is, it is regular in the sense of Cramer–Rao.

By CR inequality, for an UE $\hat{T}$ of θ,

$$\text{Var}(T) \geq \frac{1}{I_n(\theta)} = \text{CRLB}.$$

Here, $f(x,\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{x^2}{2\theta}}$, $x \in \mathbb{R}$, $\theta>0$

$$\Rightarrow \ln f(x,\theta) = -\frac{1}{2}\ln(2\pi\theta) - \frac{x^2}{2\theta}$$

$$\frac{\partial}{\partial\theta}\ln f(x_1,\theta) = -\frac{1}{2\theta} + \frac{x_1^2}{2\theta^2}$$

and $$\frac{\partial^2}{\partial\theta^2}\ln f(x_1,\theta) = \frac{1}{2\theta^2} - \frac{x_1^2}{\theta^3}$$

$$I_n(\theta) = n . I_1(\theta)$$
$$= n . E\left(-\frac{\partial^2}{\partial\theta^2} \ln f(x_1,\theta)\right)$$
$$= n . \left\{-\frac{1}{2\theta^2} + \frac{E(X_1^2)}{\theta^3}\right\}$$
$$= n\left\{-\frac{1}{2\theta^2} + \frac{\theta}{\theta^3}\right\}$$
$$= \frac{n}{2\theta^2}$$

Hence, $Var(T) \geq \frac{2\theta^2}{n} = CRLB$

The MVBUE, if exists, for θ, is given by

$$T = \psi(\theta) \pm \frac{\psi'(\theta)}{I_n(\theta)} \cdot \frac{\partial}{\partial\theta} \ln L(\underset{\sim}{x},\theta)$$
$$= \theta \pm \frac{1}{\frac{n}{2\theta^2}} \cdot \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(x_i,\theta)$$
$$= \theta \pm \frac{2\theta^2}{n}\left\{-\frac{n}{2\theta} + \frac{\Sigma x_i^2}{2\theta^2}\right\}$$
$$= \theta + \frac{2\theta^2}{n}\left\{-\frac{n}{2\theta} + \frac{\Sigma x_i^2}{2\theta^2}\right\}, \text{ taking +ve sign only.}$$
$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Hence, $T = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ attains CRLB and $\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)$ is the MVBUE as well as UMVUE of θ.

Ex.(7):- Based on a r.s. $X_1, X_2, \dots, X_n$ from Gamma (α). Obtain an estimator of $\psi_\alpha = \frac{\partial}{\partial\alpha} \ln\Gamma(\alpha)$ which attains CRLB and its variance.

Solution:- The PDF of $\underset{\sim}{X} = (X_1, \dots, X_n)$ is

$$f(\underset{\sim}{x};\alpha) = \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)} e^{-x_i} . x_i^{\alpha-1}$$
$$= \frac{1}{\{\Gamma(\alpha)\}^n} . e^{-\sum_{i=1}^{n} x_i} . \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1}, \text{ if } x_i > 0 \ \forall i = 1(1)n.$$

$$\Rightarrow \ln f(\underset{\sim}{x},\alpha) = -n\ln\Gamma(\alpha) - \sum_{i=1}^{n} x_i + (\alpha-1)\sum_{i=1}^{n} \ln x_i, \text{ if } x_i > 0$$

$$\frac{\partial}{\partial\alpha} \ln f(\underset{\sim}{x},\alpha) = -n\frac{\partial}{\partial\alpha} \ln\Gamma(\alpha) + \sum_{i=1}^{n} \ln x_i$$

and $\frac{\partial^2}{\partial\alpha^2} \ln f(\underset{\sim}{x},\alpha) = -n . \frac{\partial^2}{\partial\alpha^2} . \ln\Gamma(\alpha)$

Here, $I(\alpha) = E\left(-\frac{\partial^2}{\partial\alpha^2} \ln f(\underset{\sim}{x},\alpha)\right) = n . \frac{\partial^2}{\partial\alpha^2} \ln\Gamma(\alpha).$

An UE which attains CRLB, if exists, is given by,

$$T = \psi(\alpha) \pm \frac{\psi'(\alpha)}{I(\alpha)} \cdot \frac{\partial}{\partial\alpha} \ln f(\underset{\sim}{x}, \alpha)$$

$$= \frac{\partial}{\partial\alpha} \ln\Gamma(\alpha) \pm \frac{\frac{\partial^2}{\partial\alpha^2} \ln\Gamma(\alpha)}{n.\frac{\partial^2}{\partial\alpha^2} \ln\Gamma(\alpha)} \left\{ -n\frac{\partial}{\partial\alpha} \ln\Gamma(\alpha) + \sum \ln x_i \right\}$$

$$= \frac{\partial}{\partial\alpha} \ln\Gamma(\alpha) \pm \left\{ - \frac{\partial}{\partial\alpha} \ln\Gamma(\alpha) + \frac{1}{n}\sum \ln x_i \right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \ln x_i \text{ , taking +ve sign only.}$$

$= \ln G$, where $G = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$ is the GM of $x_1, x_2, \cdots, x_n$.

Clearly, $\text{var}(T) = \text{CRLB} = \dfrac{\{\psi'(\alpha)\}^2}{I(\alpha)}$

$$= \frac{\left\{\frac{\partial^2}{\partial\alpha^2} \ln\Gamma(\alpha)\right\}^2}{n.\frac{\partial^2}{\partial\alpha^2} \ln\Gamma(\alpha)}$$

$$= \frac{\frac{\partial^2}{\partial\alpha^2} \ln\Gamma(\alpha)}{n}$$

☒ Use of CR inequality in finding UMVUE:—

If a MVBUE T exists, then it is given by,

$$T = \psi(\theta) \pm \frac{\psi'(\theta)}{I(\theta)} \cdot \frac{\partial}{\partial\theta} \ln f(x;\theta) \quad \cdots\cdots\cdots (*)$$

Note that, the RHS of (*) can be computed once the distn.: $f(x,\theta)$, $\theta \in \Omega$ and $\psi(\theta)$ are specified and we can immediately check whether or not $\exists$ a statistic T satisfying (*).

If RHS of (*) determines a statistic T, then T is MVBUE as well as UMVUE of $\psi(\theta)$.

The above theory presents a complete solution to the problem of finding UMVUE, in the case of family of distn.s satisfying the regularity conditions and (*) for some statistic T, then T is the UMVUE of $\psi(\theta)$.

Ex.(8):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from the PMF

$$P[X=0] = 1 - \frac{\theta}{2}, \quad P[X=1] = \frac{1}{2}, \quad P[X=2] = \frac{\theta}{2}; \ 0<\theta<1$$

Find the CRLB for θ.

Ex.(9):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(\mu, 1)$ and $\psi(m) = \mu^2$.

(a) S.T. the lower bound of the variance of an UE of μ^2 from CR inequality is $\frac{4\mu^2}{m}$.

(b) S.T. $T = \bar{X}^2 - \frac{1}{n}$ is a UMVUE of $\psi(\mu) = \mu^2$ with variance $\left(\frac{4\mu^2}{n} + \frac{2}{n^2}\right)$

Compare (a) & (b) and comment.

Ex.(10):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from exp distn. with mean $1/\alpha$.

(a) S.T. $T = \frac{n-1}{n\bar{X}}$ is the UMVUE of α with variance $\frac{\alpha^2}{n-2}$.

(b) S.T. the CRLB is $\frac{\alpha^2}{n}$.

Compare (a) and (b) and comment.

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 $p'(x)$ 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 $P[x-h/2 < X < x+\frac{h}{2}] \simeq f(x;\theta).h$ 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 $L(\theta/\underset{\sim}{x})$~~, 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

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

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

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

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

$L(p/\underset{\sim}{x} = 0) = (1-p)^n \downarrow p$ 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 $(x_1, x_2, \ldots, x_n)$ be a r.s. from Bernoulli(p), $0<p<1$. If $(x_1, \ldots, x_n) = (0, 0, \ldots, 0)$ or $(1, 1, \ldots, 1)$ then MLE of p does not exist.

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

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

The Likelihood function is

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

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

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

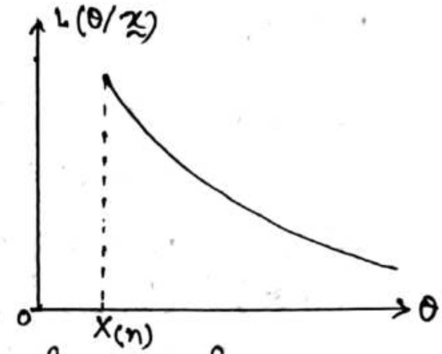
$$= \frac{n}{\lambda}(\bar{x}-\lambda) \begin{cases} >0 & \text{if } \bar{x}>\lambda \\ <0 & \text{if } \bar{x}<\lambda \end{cases}$$

Hence, $L(\lambda/\underset{\sim}{x})$ first increases, then achives its 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 $X_1, \ldots, X_n$ 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 $\theta \geqslant x_{(n)}$, $L(\theta/\underset{\sim}{x}) = \frac{1}{\theta^n}$ is a decreasing function of θ.

Hence, $L(\theta/\underset{\sim}{x})$ is maximum iff $\theta\,(\geqslant x_{(n)})$ 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 $X_1, \ldots, X_n$ be a r.s. from $U(\alpha, \beta)$. Find the MLE of (α, β).

Solution:- The Likelihood function is

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

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

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

$L(\alpha, \beta / \underset{\sim}{x}) = \dfrac{1}{(\beta-\alpha)^n}$ 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 $X_1, \ldots, X_n$ 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 $\underset{\sim}{x} = (x_1, \ldots, x_n)$ 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 $0 \le \alpha \le 1$,

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

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. (6):- Let $X_1, \ldots, X_n$ be a r.s. from $U(-\theta, \theta)$; $\theta > 0$. Find the MLE of θ. Is it unique?

Solution:- $X_i \overset{iid}{\sim} U(-\theta, \theta)$, $i = 1(1)n$

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

$\Rightarrow Y_1, \ldots, Y_n$ is a r.s. from $U(0, \theta)$.

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.

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

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

iff $\dfrac{1}{\sqrt{4^n G^n}} \gtreqless 1$, 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 parameter 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 $X_1, \dots, X_n$ be a r.s. from $N(\mu, \sigma^2)$, $\mu \in R, \sigma > 0$

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 $\mu \in R, \sigma > 0$.

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

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

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

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

, 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

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

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

Solution:- The log-likelihood function is

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

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

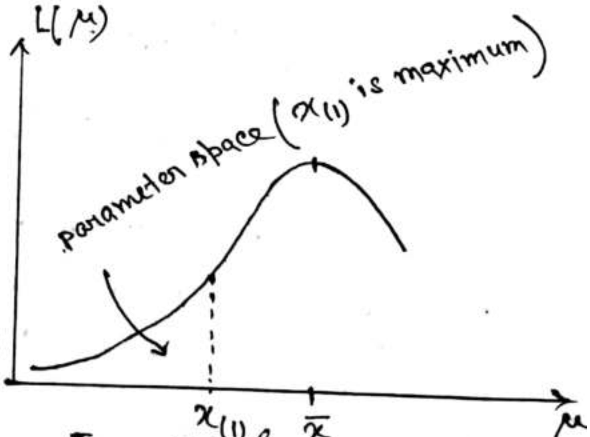
$$\frac{\delta}{\delta u} \ln L = \frac{\delta}{\delta u} \left\{ -n \ln \mu - \frac{1}{\mu} \Sigma (x_i - \mu) \right\}$$

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

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

$\Rightarrow L(\mu / \underset{\sim}{x})$ 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 $\Omega = \left[\frac{1}{4}, \frac{3}{4}\right]$.

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 $p \in \left[\frac{1}{4}, \frac{3}{4}\right]$

iff $p = 1/4$.

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

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

$\therefore$ 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 $\delta(X) = E[\delta(X) - p]^2$
$$= E\left(\tfrac{1}{2} - p\right)^2$$
$$\le \tfrac{1}{16}$$
$\left[\because \tfrac{1}{4} \le p \le \tfrac{3}{4} \Rightarrow -\tfrac{1}{4} \le p - \tfrac{1}{2} \le \tfrac{1}{4}\right]$

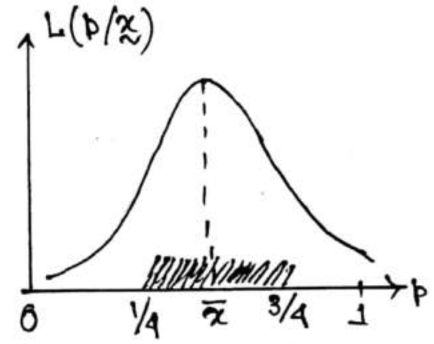
Hence, $MSE\{\delta(X)\} \le MSE(\hat{p})$.

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

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

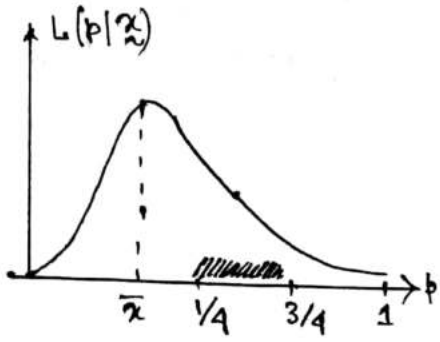
Hence, $L(p|\underset{\sim}{x})$ first increases, then achieves its maximum at $p = \bar{x}$ and finally decreases.



Case I:- Let, $\tfrac{1}{4} \le \bar{x} \le \tfrac{3}{4}$

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

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



Case II:-

Let, $\bar{X} < \tfrac{1}{4}$

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

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

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

- Hence, the MLE of p is
$$\hat{p} = \begin{cases} \tfrac{1}{4} & \text{if } \bar{X} < \tfrac{1}{4} \\ \bar{X} & \text{if } \tfrac{1}{4} \le \bar{X} \le \tfrac{3}{4} \\ \tfrac{3}{4} & \text{if } \bar{X} > \tfrac{3}{4} \end{cases}$$

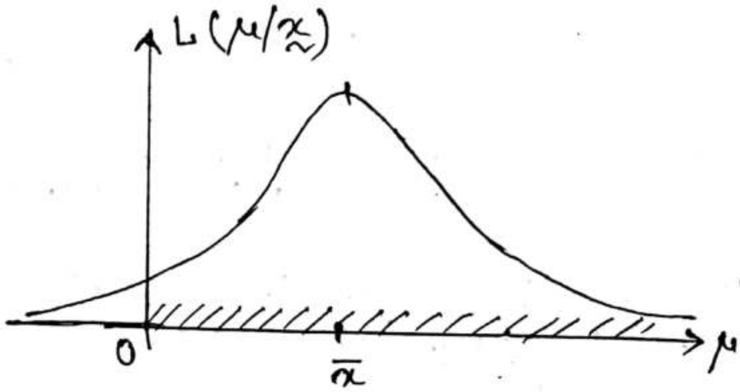
Ex. (14):- Let $X_1, \ldots, X_n$ be a r.s. from $N(\mu, 1)$, $\mu \geq 0$. Find the MLE of μ.

Solution:-

$$L(\mu/\underset{\sim}{x}) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2} \text{ ; where } \mu \geq 0.$$

$$\frac{\delta}{\delta\mu} \ln L(\mu/\underset{\sim}{x}) = n(\bar{x}-\mu) \begin{cases} >0 & \text{if } \mu < \bar{x} \\ <0 & \text{if } \mu > \bar{x} \end{cases}$$

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



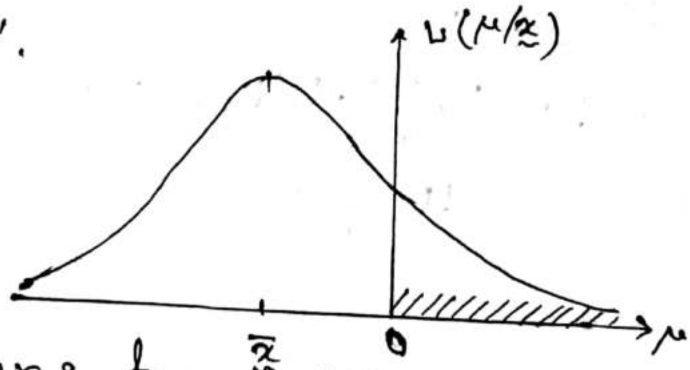
Case I:- $\bar{x} \geq 0$

Hence the MLE of μ is

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

Case II:- $\bar{x} < 0$

Then the MLE of μ is '0'.



- Hence the MLE of μ is

$$\hat{\mu} = \begin{cases} \bar{X} & , \text{ if } \bar{X} \geq 0 \\ 0 & , \text{ if } \bar{X} < 0 \end{cases}$$

Ex. (15):- Let $X_1, \ldots, X_n$ be a r.s. from the PMF

$P[X=0] = (1-\theta)/2$, $P[X=1] = 1/2$, $P[X=2] = \theta/2$; $0 < \theta < 1$, with atleast one value with 0 and 2. Find the MLE of θ.

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 $L(\underset{\sim}{x};\theta) = g(T(\underset{\sim}{x}),\theta)\cdot h(\underset{\sim}{x})$; where, $h(\underset{\sim}{x})$ is independent of θ, by factorization criterion.

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

Now, the likelihood equation is

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

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

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

$$0 = \frac{\partial}{\partial\theta}\ln L(\underset{\sim}{x},\theta)$$
$$= \frac{\partial}{\partial\theta}\ln g(T(\underset{\sim}{x}),\theta)$$ will be a function of T.

[Maximizing $\ln L(\underset{\sim}{x};\theta)$ w.r.t. θ is equivalent to maximizing $\ln g(T(\underset{\sim}{x}),\theta)$ w.r.t. θ. Here, $g(T(\underset{\sim}{x}),\theta)$ depends only on θ and $T(\underset{\sim}{x})$. The MLE of θ is the value of θ for which $\ln L(\underset{\sim}{x};\theta)$ or $\ln g(T(\underset{\sim}{x}),\theta)$ 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.

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

The likelihood equation is

$$\frac{\partial}{\partial\theta}\ln L(\underset{\sim}{x};\theta) = 0$$
$$\Rightarrow \Lambda(\theta)\{T-\theta\} = 0$$
$$\Rightarrow \theta = T \text{ is the unique solution.}$$

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

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

and $\left.\frac{\partial^2}{\partial\theta^2}\ln L(\underset{\sim}{x};\theta)\right|_{\theta=T} = -\Lambda(T) < 0$

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

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

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

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 $N(\mu, \hat{\sigma}^2)$ popl^n., $\hat{\sigma}^2 = \frac{1}{n}\Sigma(x_i - \bar{x})^2$, which is biased but $E\left(\frac{n}{n-1}\hat{\sigma}^2\right) = \sigma^2$, i.e. $\frac{n}{n-1}.\hat{\sigma}^2$ 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 $\sup_{\lambda} L^*(\lambda; \underset{\sim}{x}) = \sup_{\lambda} L(h^{-1}(\lambda); \underset{\sim}{x}) = \sup_{\theta} L(\theta; \underset{\sim}{x})$

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

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

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 $nI_{X_1}(\lambda)=\frac{n}{\lambda}$ $\quad\left[V(\bar{X})=\frac{1}{nI_{X_1}(\lambda)}\right]$

$\therefore e^{-\hat{\lambda}} \overset{a}{\sim} N\left(e^{-\lambda}, \frac{\lambda e^{-2\lambda}}{n}\right)$ 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.

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

<u>Solution</u>:- 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

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

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

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

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

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

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. $X_1,\dots,X_n$ 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 $X_1,\dots,X_n$ 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 :-

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

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

Likelihood Equation: -

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

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

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

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

$$\Rightarrow \theta=\beta=\frac{-1+\sqrt{1+\frac{4}{n}\sum X_i^2}}{2}$$; 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 $\theta=\beta$.

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

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 $M \gtrless \dfrac{nN}{x}$.

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.

A Practical Method of solution of Likelihood Equation

[Fisher's Method of Scoring]

In case of a single parameter Cramer family such as Cauchy, the variables $x_1, x_2, \ldots, x_n$ and θ are not seperable and the likelihood equation is given by

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

This is an algebraic equation of degree $(2n-1)$ in θ and explicit solution is not available. We then use classical iteration procedure to obtain a numerical solution for the observed values $x_1, x_2, \ldots, x_n$. In Newton-Raphson method, we start the iterative procedure with T_1 as a trial value and obtain successive iteration by

$$T_{r+1} = T_r - \left(\frac{\frac{\partial \ln L}{\partial\theta}}{\frac{\partial^2 \ln L}{\partial\theta^2}} \right)_{\theta = T_r}$$

Fisher proposed a modification of the NR method

$$T_{r+1} = T_r + \left\{ \frac{\left(\frac{\partial \ln L}{\partial\theta}\right)}{n.I_{X_1}(\theta)} \right\}_{\theta = T_r} \qquad \text{———} (*)$$

Note that, fisher modification consists in using $E\left(-\frac{\partial^2 \ln L}{\partial\theta^2}\right) = n.I_{X_1}(\theta)$ and iterative procedure given by (*) is known as <u>Fisher's method of scoring</u>.

Example:- Describe the method of finding the MLE of θ in the Cauchy $(\theta, 1)$ distribution for a r.s. $x_1, \ldots, x_n$.

Solution:- Here, the sample median $\tilde{X}$ is consistent for θ and

$I_{X_1}(\theta) = \frac{1}{2}$ (find it).

Hence, considering $T_1 = \tilde{x}$ as a trial root, the fisher's method of scoring gives

$$T_2 = T_1 + \left\{ \frac{\frac{\partial}{\partial\theta} \ln L}{n.I_{X_1}(\theta)} \right\}_{\theta = T_1} = \tilde{x} - \frac{4}{n} \sum_{i=1}^{n} \frac{(x_i - \tilde{x})}{\{1+(x_i-\tilde{x})^2\}}$$

as an improved estimate of θ over $T_1 = \tilde{x}$ and the successive improved estimates are

$$T_{r+1} = T_r + \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - T_r)}{\{1+(x_i-T_r)^2\}}.$$

(II) Method of Minimum Chi-Square (MCS): —

Suppose, we have a sample of size n from a popln distn. which can be classified as a multinomial popln. with K mutually exclusive, exhaustive classes with probabilities $p_1, p_2, \ldots, p_K$, and the observed frequency of the ith class being f_i, $i=1(1)K$, where $\sum_{i=1}^{K} p_i = 1$ and $\sum_{i=1}^{n} f_i = n$.

Then p_i's are functions of the parameters $\theta_1, \theta_2, \ldots, \theta_l$ so that $p_i = p_i(\theta_1, \ldots, \theta_l)$, $i=1(1)K$. The expected frequency of the ith class is $n.p_i$.

As a measure of goodness of fit between the observed frequency and the expected frequency, Karl Pearson suggested the following statistic:

$$\chi^2 = \sum_{i=1}^{K} \frac{(f_i - np_i)^2}{np_i}.$$

One may ask the question in this connection: What procedure of estimation should be used? To answer the question, one will be inclined to estimate the unknown parameters so as the "measures of goodness of fit" or "measure of discrepency between the observed and expected frequencies", i.e. the χ^2 as small as possible. This procedure of estimation may be called the '<u>minimum χ^2 method</u>'.

To minimize χ^2 by Calculas, we have to solve the equations: —

$$0 = \frac{\partial \chi^2}{\partial \theta_r} = \frac{\partial}{\partial \theta_r}\left\{\sum_{i=1}^{K} \frac{(f_i - np_i)^2}{np_i}\right\}$$

$$= -2\sum_{i=1}^{K}\left\{\frac{f_i - np_i}{p_i} + \frac{(f_i - np_i)^2}{2np_i^2}\right\}\frac{\partial p_i}{\partial \theta_r} \quad \text{----} (*)$$

for $r = 1(1)l$.

Even in simple cases, the system of equations (*) are often very difficult to solve. It is, however, intuitively plausible that if the <u>hypothesis</u> is true, the terms $\frac{(f_i - np_i)^2}{2np_i^2}$ will, for large n, have little effect on the value of the roots of (*). We shall omit these terms and thus replace (*) by the simpler system of equations:

$$\sum_{i=1}^{K} \frac{f_i - np_i}{p_i} \cdot \frac{\partial p_i}{\partial \theta_r} = 0 \Rightarrow \sum_{i=1}^{K} \frac{f_i}{p_i} \cdot \frac{\partial p_i}{\partial \theta_r} - n.\frac{\partial}{\partial \theta_r}\left(\sum_{i=1}^{K} p_i\right) = 0$$

$$\Rightarrow \sum_{i=1}^{K} \frac{f_i}{p_i} \cdot \frac{\partial p_i}{\partial \theta_r} = 0 \quad \text{as } \sum_{i=1}^{K} p_i = 1, \text{ for } r = 1, 2, \ldots, l \quad \ldots\ldots (**)$$

The procedure of estimating $\theta_1, \ldots, \theta_l$ by solving (**) will be called the "<u>Modified Minimum χ^2 method</u>".

EFFICIENCY AND ASYMPTOTIC EFFICIENCY

(A) Efficiency :— Let T_1, T_2 be two UEs for $\psi(\theta)$ and $V(T_1)$, $V(T_2)$ are finite. We define the efficiency of T_1 w.r.t. T_2 by
$$eff(T_1/T_2) = \frac{V(T_2)}{V(T_1)}.$$
The precision of an UE T is defined as $\frac{1}{var(T)}$.

Most Efficient Estimator:— An UE T of $\psi(\theta)$ is called most efficient among all UEs of $\psi(\theta)$, if T is UMVUE of $\psi(\theta)$.

Efficiency:— Let T be the most efficient estimator, i.e. UMVUE of $\psi(\theta)$. Then the efficiency of any UE T_1 of $\psi(\theta)$ is defined as
$$eff(T_1) = eff(T_1/T) = \frac{V(T)}{V(T_1)}.$$

Alternative concept:— If there exists MVBUE of $\psi(\theta)$ which is also UMVUE and most efficient, then
$$eff(T_1) = \frac{\{\psi'(\theta)\}^2}{I_n(\theta)} \Big/ Var(T_1).$$

Ex.(1):— If T_1 and T_2 are two UEs of $\psi(\theta)$ having the same variance and ρ is the correlation coefficient between them. Show that,— $\rho \geq 2e-1$, where, e is the efficiency of each estimators.

Solution:— Let T be the most efficient/UMVUE of $\psi(\theta)$.

Here $V(T_1) = V(T_2) = v$, say.

Then $e = \frac{V(T)}{v}$

Define, $T_3 = \frac{T_1+T_2}{2}$ as an UE of $\psi(\theta)$.

Hence, $V(T_3) \geq V(T)$ $\Rightarrow \frac{1}{4}\{V(T_1)+V(T_2)+2cov(T_1,T_2)\} \geq V(T)$

$\Rightarrow \frac{1}{4}\{v+v+2\rho v\} \geq ev$

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

Ex.(2):— Show that the correlation coefficient between a most efficient or UMVUE and any other UE with efficiency e is $\sqrt{e}$.

Solution:— Let T, T_1 be the UMVUE and any other UE of $\psi(\theta)$, respectively. Then, $e = \frac{V(T)}{V(T_1)}$.

Note that, $E(T-T_1)=0 \ \forall \theta$

$\Rightarrow u = T - T_1$ is an UE of zero.

As, T is UMVUE, $cov(T, u) = 0 \ \forall\theta$ and for any UE u of zero.

Hence, $cov(T, T-T_1) = 0 \ \forall \theta$

$\Rightarrow V(T) = cov(T, T_1)$

Now, $$\rho = \frac{cov(T,T_1)}{\sqrt{V(T)}\sqrt{V(T_1)}} = \sqrt{\frac{V(T)}{V(T_1)}} = \sqrt{e}.$$

Ex.(3):- Let T_1, T_2 be two UEs of $\psi(\theta)$ with efficiencies e_1 and e_2, respectively and $\rho = \rho(T_1, T_2)$. Then show that

$$\sqrt{e_1 e_2} - \sqrt{(1-e_1)(1-e_2)} \le \rho \le \sqrt{e_1 e_2} + \sqrt{(1-e_1)(1-e_2)}.$$

Solution:- Let T be the UMVUE of $\psi(\theta)$.

Then $e_i = \dfrac{V(T)}{V(T_i)}$, $i = 1, 2$.

Define, $T_3 = \alpha T_1 + \beta T_2$, $(\alpha + \beta = 1)$, as an UE of $\psi(\theta)$.

Hence, $V(T_3) \geq V(T)$, $\forall (\alpha, \beta)$.

$$\Rightarrow \alpha^2 V(T_1) + \beta^2 V(T_2) + 2\alpha\beta \operatorname{cov}(T_1, T_2) \geq Var(T)$$

$$\Rightarrow Var(T) \left\{ \frac{\alpha^2}{e_1} + \frac{\beta^2}{e_2} + 2\alpha\beta \cdot \frac{\rho}{\sqrt{e_1 e_2}} \right\} \geq V(T)$$

$$\Rightarrow \frac{\alpha^2}{e_1} + \frac{\beta^2}{e_2} + 2\alpha\beta \cdot \frac{\rho}{\sqrt{e_1 e_2}} \geq 1 = (\alpha + \beta)^2$$

$$\Rightarrow \alpha^2 \left(\frac{1}{e_1} - 1 \right) + \beta^2 \left(\frac{1}{e_2} - 1 \right) + 2\alpha\beta \left(\frac{\rho}{\sqrt{e_1 e_2}} - 1 \right) \geq 0 \ \forall (\alpha, \beta)$$

The LHS is a quadratic in (α, β) and it is n.n.d.

Hence,
$$\begin{vmatrix} \frac{1}{e_1} - 1 & \frac{\rho}{\sqrt{e_1 e_2}} - 1 \\ \frac{\rho}{\sqrt{e_1 e_2}} - 1 & \frac{1}{e_2} - 1 \end{vmatrix} \geq 0$$

$$\Rightarrow \left(\frac{\rho}{\sqrt{e_1 e_2}} - 1 \right)^2 \le \left(\frac{1}{e_1} - 1 \right) \left(\frac{1}{e_2} - 1 \right)$$

$$\Rightarrow -\sqrt{(1-e_1)(1-e_2)} \le \rho - \sqrt{e_1 e_2} \le \sqrt{(1-e_1)(1-e_2)}$$

$$\Rightarrow \sqrt{e_1 e_2} - \sqrt{(1-e_1)(1-e_2)} \le \rho \le \sqrt{e_1 e_2} + \sqrt{(1-e_1)(1-e_2)}$$

Remark:-

i) In ex.(1); $e_1 = e_2 = e$

$$e - (1-e) \le \rho \le e + (1-e)$$
$$\Rightarrow 2e - 1 \le \rho \le 1.$$

ii) In ex.(2); $e_1 = e$, $e_2 = 1$.

$$\sqrt{e} \le \rho \le \sqrt{e}$$
$$\Rightarrow \rho = \sqrt{e}.$$

Ex. (4):- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(0, \sigma^2)$. Find the most efficient estimator of σ^2. Also, obtain an UE of σ based on $\sum_{i=1}^{n}|X_i|$ and its efficiency.

Hints:- The MVBUE of σ^2 is $T = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

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

$$\mathrm{Eff}(T_1) = \frac{\text{CRLB for } \psi(\sigma) = \sigma}{V(T_1)}.$$

(B) Asymptotic Efficiency :— There may be a large no. of consistent estimators $\psi(\theta)$. To make a choice among the estimators which are equivalent so far as the criterion of consistency is concerned, we should have some further criterion. If we confined ourselves to those consistent estimators that are asymptotically normally distributed, then the concept of asymptotic efficiency is based on the asymptotic variance of an estimator.

Consistent Asymptotically Normal (CAN) Estimator :—

An estimator $\{T_n\}$ is said to be CAN of $\psi(\theta)$ if T_n is consistent and $\sqrt{n}\{T_n - \psi(\theta)\} \overset{a}{\sim} N(0, \sigma_T^2(\theta))$.

If $\{T_{1n}\}$ and $\{T_{2n}\}$ are two CAN estimators of $\psi(\theta)$, then one with smaller variance will be preferable.

Asymptotic Relative Efficiency (ARE) :—

If $\{T_{1n}\}$ and $\{T_{2n}\}$ are two consistent estimators and $\sqrt{n}\{T_{1n} - \psi(\theta)\} \overset{a}{\sim} N(0, \sigma_{T_1}^2(\theta))$,

$$\sqrt{n}\{T_{2n} - \psi(\theta)\} \overset{a}{\sim} N(0, \sigma_{T_2}^2(\theta)),$$

then ARE of T_1 w.r.t. T_2 is defined as

$$ARE(T_1/T_2) = \frac{\sigma_{T_2}^2(\theta)}{\sigma_{T_1}^2(\theta)}.$$

Remark :- To estimate $\psi(\theta)$, by CAN estimated $\{T_{1n}\}$ and $\{T_{2n}\}$ with precision $\frac{1}{v}$, i.e. with variance v.

Let $\sqrt{n}\{T_{1n_1} - \psi(\theta)\} \overset{a}{\sim} N(0, \sigma_{T_1}^2(\theta))$

$$\sqrt{n}\{T_{2n_2} - \psi(\theta)\} \overset{a}{\sim} N(0, \sigma_{T_2}^2(\theta))$$

$$\Rightarrow T_{1n_1} \overset{a}{\sim} N\left(\psi(\theta), \frac{\sigma_{T_1}^2(\theta)}{n_1}\right)$$

$$T_{2n_2} \overset{a}{\sim} N\left(\psi(\theta), \frac{\sigma_{T_2}^2(\theta)}{n_2}\right)$$

Here $\frac{\sigma_{T_1}^2(\theta)}{n_1} = v = \frac{\sigma_{T_2}^2(\theta)}{n_2}$

$$\Rightarrow \frac{n_2}{n_1} = \frac{\sigma_{T_2}^2(\theta)}{\sigma_{T_1}^2(\theta)}.$$

The smaller the sample size required to achieve the same precession, the better the estimator.

Hence, $eff(T_1/T_2) = \frac{n_2}{n_1} = \frac{\sigma_{T_2}^2(\theta)}{\sigma_{T_1}^2(\theta)}$.

Best Asymptotically Normal Estimator [BAN] :—

An estimator $\{T_n\}$ is said to be BAN estimator for $\psi(\theta)$ if $\{T_n\}$ is consistent for $\psi(\theta)$ and the variance of the limiting distribution.

$\sqrt{n}\{T_n-\psi(\theta)\}$ has the least possible value.

Asymptotic Efficiency :— Let $\{T_n\}$ be BAN estimator of $\psi(\theta)$. Then asymptotic efficiency of CAN estimator $\{T_{1n}\}$ of $\psi(\theta)$ is defined as $AE(T_1/T) = \dfrac{\sigma_T^2(\theta)}{\sigma_{T_1}^2(\theta)}$

Alternative concept :— Let $X_1,\ldots,X_n$ be a r.s. from a PDF or PMF satisfying the regularity conditions in CR inequality. Suppose that

$$\sqrt{n}\{T_n-\psi(\theta)\} \overset{a}{\sim} N(0,\sigma_T^2(\theta)) \longrightarrow (i)$$

and under some additional conditions it can be shown that

$$\sigma_T^2(\theta) \geq \frac{\{\psi'(\theta)\}^2}{I_{X_1}(\theta)} \longrightarrow (ii)$$

In any such regular cases, we define the asymptotic efficiency $\{T_n\}$ satisfying (i) and (ii), as the limiting value of

$$\left\{\frac{\dfrac{\psi'(\theta)}{I_{X_1}(\theta)}}{\sigma_T^2(\theta)}\right\}^2.$$

Ex.(1):— Let $X_1,\ldots,X_n$ be a r.s. from $N(\mu,\sigma^2)$. Find the asymptotic efficiency of the sample median relative to sample mean and comment.

Solution:— Here $\bar{X} \sim N\left(\mu,\frac{\sigma^2}{n}\right)\ \forall n$

and $\tilde{X} \overset{a}{\sim} N\left(\mu, \frac{1}{4nf^2(\mu)}\right)$, for large n.

$\Rightarrow \tilde{X} \overset{a}{\sim} N\left(\mu, \frac{\pi\sigma^2}{2n}\right)$, for large n.

Clearly, $\bar{X}$ and $\tilde{X}$ are CAN estimators of μ.

Asymptotic Relative efficiency of $\tilde{X}$ w.r.t. $\bar{X}$ is

$$ARE(\tilde{X}/\bar{X}) = \frac{\sigma^2_{\bar{X}}(\mu)}{\sigma^2_{\tilde{X}}(\mu)} = \frac{\sigma^2}{\frac{\pi\sigma^2}{2}} = \frac{2}{\pi} \simeq 0.64$$

$$\left[\begin{array}{l} \because \sqrt{n}(\bar{X}-\mu) \sim N(0,\sigma^2) \\ \text{and } \sqrt{n}(\tilde{X}-\mu) \sim N\left(0,\frac{\pi\sigma^2}{2}\right)\end{array}\right]$$

Note that $\bar{X}$ is the UMVUE of μ, hence it is most efficient for μ, Now, $ARE(\tilde{X}/\bar{X}) = \frac{n_2}{n_1} \simeq \frac{64}{100}$;
this means that an estimate of μ from a sample of $n_2 = 64$ observations using $\bar{x}$ is just as reliable as an estimate from a sample of $n_1 = 100$ observations using $\tilde{x}$.

<u>Ex.(2)</u>:- Let $X_1, \dots, X_n$ be a r.s. from

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

. Find the asymptotic efficiency of the sample median to estimate θ.

<u>Soln</u>:- Here $\frac{\partial \ln f(x_1;\theta)}{\partial\theta} = \frac{2(x_1-\theta)}{1+(x_1-\theta)^2}$

Now,
$$I_{X_1}(\theta) = E\left(\frac{\partial}{\partial\theta}\ln f(x_1;\theta)\right)^2$$

$$= \int_{-\infty}^{\infty} \frac{4(x_1-\theta)^2}{\{1+(x_1-\theta)\}^2} \cdot \frac{1}{\pi\{1+(x_1-\theta)^2\}} dx_1$$

$$= \int_{-\infty}^{\infty} \frac{4z^2}{\pi\{1+z^2\}^3} dz \text{ ; where } z = x_1-\theta.$$

$$= 8\int_0^{\infty} \frac{z^2}{\pi(1+z^2)^3} dz$$

$$= \frac{4}{\pi}\int_0^{\infty} \frac{t^{3/2-1}}{(1+t)^{3/2+3/2}} dt \text{ ; where } t = z^2$$

$$= \frac{4}{\pi}\cdot B\left(\frac{3}{2}, \frac{3}{2}\right)$$

$$= \frac{4}{\pi}\cdot \frac{\Gamma(3/2)\Gamma(3/2)}{\Gamma(3)}$$

$$= \frac{4}{\pi}\frac{\left(\frac{1}{2}\sqrt{\pi}\right)^2}{2} = \frac{1}{2}$$

$$\Rightarrow I_n(\theta) = \frac{n}{2}.$$

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

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

$$\Rightarrow \sqrt{n}(\tilde{X}-\theta) \overset{a}{\sim} N\left(0, \frac{\pi^2}{4} = \sigma^2_{\tilde{X}}(\theta)\right)$$

Hence, $AE(\tilde{X}) = \left\{\frac{1}{I_{X_1}(\theta)}\right\} / \sigma^2_{\tilde{X}}(\theta) = \frac{2}{\frac{\pi^2}{4}} = \frac{8}{\pi^2} \simeq 0.8104.$

——— x ———

[C.U] **WORKED OUT PROBLEMS ON ESTIMATION**

① Let $X_1, \ldots, X_n$ be a random sample of size n from Poisson(λ) popln.. Show that the sample mean is UMVUE for λ.

'10 (5)

Solution:- $X_1, \ldots, X_n$ be a r.s. from $P(\lambda)$ of size n, where λ being unknown.

Now, poisson distribution belongs to the exponential family of distribution.

Thus, $T'(X) = \sum_{i=1}^{n} X_i$ is a complete statistic.

We prove it as follows:

$$E(T') = 0$$

$$\Rightarrow \sum_{x=0}^{\infty} T'(x) \cdot \frac{e^{-\lambda} \cdot \lambda^x}{x!} = 0$$

$$\Rightarrow T'(0)e^{-\lambda} + T'(1) \cdot e^{-\lambda} \cdot \lambda + T'(2) \cdot \frac{e^{-\lambda} \cdot \lambda^2}{2!} + \cdots\cdots = 0$$

Now, each coefficient of $T(\underset{\sim}{x})$ is non-zero.

Here to satisfy the RHS of the equation

$$T'(x) = 0 \quad \forall x$$

$\therefore T'(\underset{\sim}{x})$ is a complete statistic.

Now, let $T = \frac{1}{n} \sum_{i=1}^{n} X_i$

$$E_\lambda(T) = \frac{1}{n} \sum_{i=1}^{n} E_\lambda(X_i) = \frac{1}{n}(n\lambda) = \lambda$$

$\therefore T$ is an unbiased estimator for λ.

Again, $T = \frac{T'}{n}$, i.e. T is a function of complete statistic T'.

$\therefore$ By Lehmann-Scheffe theorem T = sample mean is UMVUE for λ.

② Show that the correlation coefficient between an MVUE and any unbiased estimator is non-negative. Make your comments.

'10 (5)

Solution:- Let T be an unbiased estimator for a parametric function $\gamma(\theta)$ and T_0 be the UMVUE of $\gamma(\theta)$.

Now, $E_\theta(T - T_0) = 0$

$\therefore T - T_0$ is an unbiased estimator for '0'.

Now, $Cov(T_0, T - T_0) = 0$ [The condition of MVUE]

$\Rightarrow Cov(T, T_0) = Var_\theta(T_0)$

Now, the correlation coefficient between T and T_0 is given by,

$$\rho_{T,T_0} = \frac{Cov_\theta(T, T_0)}{\sqrt{V_\theta(T_0)}\sqrt{V_\theta(T)}} = \sqrt{\frac{V_\theta(T_0)}{V_\theta(T)}} \geq 0 \quad [\because V_\theta(T_0) \text{ and } V_\theta(T) \text{ are non-negative}]$$

Hence, the result is proved.

▨ Since the correlation coefficient between MVUE and an unbiased estimator is always non-negative and we can comment that they have a positive correlation, i.e. the estimate of both the estimators will not differ much.

③. State the important properties of a maximum likelihood estimator. (3) '10

Solution:– The important properties of maximum likelihood estimator is as follows:

i) Let us consider a one parameter exponential family labelled by parameter θ. Here if MVBUE exists, then it will be an MLE for θ. The fact immediately follows from the condition of existance of an MVBE.

i.e. $\frac{\partial}{\partial\theta} \ln f_\theta(\underset{\sim}{x}) = k(\theta)\left(T(\underset{\sim}{x}) - \theta\right)$

$\therefore \frac{\partial}{\partial\theta} \ln L(\theta) = 0$ [$\because L(\theta)$ is the likelihood function of the estimator]

$\Rightarrow \hat{\theta} = T(\underset{\sim}{x})$

and as a consequence MLE is necessarily a sufficient statistic.

ii) Invariance Property of MLE: — If the r.s. is drawn from $f_\theta(\cdot)$ popl^n. and if $T(\underset{\sim}{x})$ be an MLE of θ, then $g(T(\underset{\sim}{x}))$ will be the MLE of $g(\theta)$.

iii) Maximum likelihood estimator of a parameter is not unique.

iv) MLE may be absurd even when exist.

④. Write a short note on minimum χ^2-estimator. (3). ['09]

Solution:– Suppose a sample of size n is drawn from a popl^n labelled by parameter $\underset{\sim}{\theta}$. Further assume that the popl^n. is classified into K mutually exclusive and exhaustive classes $A_1, \ldots, A_K$. Let, $\pi_i = P(A_i)$, $\pi_i > 0$, $\sum_{i=1}^{K} \pi_i = 1$.

Clearly $\pi_i = \pi_i(\underset{\sim}{\theta})$.

If n_i observations fall in A_i $\left(\sum_{i=1}^{K} n_i = n\right)$ then,

$(n_1, n_2, \ldots, n_{K-1}) \sim$ multinomial $(n, \pi_1, \pi_2, \ldots, \pi_{K-1})$,

which implies, $n_i \sim bin(n, \pi_i)$, $i = 1(1)K$.

$E(n_i) = n\pi_i$.

A measure of discrepancy between the observed and expected frequency is given by,

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n\pi_i)^2}{n\pi_i}.$$

As $\pi_i = \pi_i(\theta)$, then clearly, $\chi^2 = \chi^2(\theta)$.

An estimate of θ can be obtained by minimizing $\chi^2(\theta)$. Clearly, the estimate of θ can be obtained by solving the following equation,

$$\frac{\partial}{\partial \theta} \chi^2(\theta) = 0 \quad \forall\ i = 1(1)k,$$

is called the minimum χ^2 equation, provided $\chi^2(\theta)$ is completely differentiable and $\left(\left(\frac{\partial^2 \chi^2(\theta)}{\partial \theta_i \partial \theta_j}\right)\right)$ is positive definite.

This method is too much cumbersome since it is very difficult to deal with the minimum χ^2 equation as $\underset{\sim}{\theta}$ occurs in the denominators of the terms under the sum through π_i's.

⑤ "A sufficient statistic provides a reduction of the data without loss of information" – Explain. (S)'<u>10</u>

<u>Solution:–</u> An experimenter uses the information in a r.s. $X_1, X_2, \ldots, X_n$ to make inference about an unknown population parameter θ. If the sample size n is large then the observed sample $(x_1, x_2, \ldots, x_n)$ is a long list of members that may be hard to interpret. An experimenter might wish to summerize the data in a sample determining a few key features of the sample values. This is usually done by computing a statistic, a function of the random sample. A statistic is also a r.s; it condenses the n r.v. ; it condenses the n r.v.s $X_1, \ldots, X_n$ into a single r.v.. We shall be interesting in watching if we lost any "information" by this condensing process. Let $\mathscr{X}$ denotes the sample space, i.e. the range of the values that $(X_1, \ldots, X_n)$ may assume. Now a statistic defines a partition of $\mathscr{X}$ or induces partition.

A sufficient statistic is a particular kind of statistics. It is said that, "A sufficient statistic provides a reduction of the data without any loss of information". We justify this statement by the definition of sufficient statistic given as follows:

Let $X_1, \ldots, X_n$ be a r.s. from the density $f_\theta(\cdot)$, where θ is unknown parameter. A statistic $T(\underset{\sim}{X})$ is defined to be a sufficient statistic if and only if the conditional distribution of $X_1, \ldots, X_n$ given $T(\underset{\sim}{X}) = t$ does not depend on θ for any value t of T. The definition says, that a statistic $T(\underset{\sim}{X})$ is sufficient if the conditional distribution of the sample given the statistic value does not depend on θ. The idea is that if the value of the sufficient statistic is known, then the sample values are not needed and it can not tell nothing more about θ and this is true since the distribution of the sample given the sufficient statistic does not depend on θ.

Hence we can say that sufficient statistic condenses $\mathfrak{X}$ in such a way that 'no information about θ is lost'.

(6). State and prove Lehmann-Scheffe theorem in the theory of point estimation. Suppose X has the Poisson distribution with unknown variance λ. Discuss how this theorem can be applied in finding the uniformly minimum variance unbiased estimator of $\lambda + \lambda^2$ on the basis of n independent observations on X. Prove that the non-availability of an unbiased estimator of $\frac{1}{\lambda}$. Suggest any possible estimator of $\frac{1}{\lambda}$ with reasons. (5+5+3+2) '08

Solution:-

Lehmann Scheffe Theorem:-

Statement:- Let h be an unbiased estimator for a real valued parametric function $\gamma(\theta)$. Then if T be a complete sufficient statistic then $E[h/T]$ will be a UMVUE of $\gamma(\theta)$.

Proof:- Let $X_1, \ldots, X_n$ be a r.s. from a popln. $f_\theta(\cdot)$, θ is an unknown parameter and $\theta \in \Theta$. Now, if $h(\underset{\sim}{x})$ be an unbiased estimator for the parametric function $\gamma(\theta)$, then by Rao Blackwell theorem, we know that for any other sufficient statistic T, the conditional distribution of h given T is an improvement over T, i.e.

$$E\left(E[h|T]\right) = \gamma(\theta) .$$

Now, here we are required to show that if T is a complete sufficient statistic then $E[h/T]$ is UMVUE for $\gamma(\theta)$, i.e. here it is enough to prove if T is complete sufficient, then $E[h/T]$ is unique.

For this we consider that h_1 and h_2 be two unbiased estimator for $\gamma(\theta)$.

$$\therefore E[E(h_1/T)] = E[E(h_2/T)] = \gamma(\theta)$$

Let $\psi(T) = E[h_1/T] - E[h_2/T]$

Now, since T is complete ~~sufficient~~ then $\psi(T)$ is also a complete statistic.

$$\text{Now, } E[E(h_1/T)] - E[E(h_2/T)] = 0$$

$$\Rightarrow E\{E(h_1|T) - E(h_2|T)\} = 0$$

$$\Rightarrow E(\psi(T)) = 0.$$

Now, since $\psi(T)$ is a complete sufficient statistic.

$$\psi(T) = 0 \quad a.e.$$

$$\Rightarrow E(h_1|T) = E(h_2|T)$$

This completes the proof.

▨ Here $X_1, \ldots, X_n$ be a r.s. from a $P(\lambda)$ distn.

Let $T(\bar{X}) = \sum_{i=1}^{n} X_i$ be a statistic. Now, we will check whether the statistic is complete or not.

$$E[T(X)] = 0$$

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

$$\Rightarrow T(0) + T(1) \cdot \lambda + T(2) \cdot \frac{\lambda^2}{2!} + \cdots\cdots = 0 \text{ ——— } \langle i \rangle$$

Since coefficient of $T(X)$ in the LHS of $\langle i \rangle$ is non-zero, $T(X) = 0 \ \forall \ x$, then $T(X) = \sum X_i$ is a complete statistic. Again $T(X)$ is also a sufficient statistic for λ.

$$\text{Here } E(T) = n\lambda$$

$$\Rightarrow E\left(\frac{T}{n}\right) = \lambda$$

$$\therefore Var(T) = n\lambda$$

$$\therefore E(T^2) - E^2(T) = n\lambda$$

$$\Rightarrow E(T^2) = n\lambda + n^2\lambda^2 + n^2\lambda - n^2\lambda.$$

$\Rightarrow E(T^2) - E(T) + nE(T) = n^2(\lambda+\lambda^2)$

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

$\Rightarrow E\left\{\frac{T^2}{n^2} - \frac{n-1}{n^2}T\right\} = \lambda+\lambda^2$

since T is complete sufficient and $\frac{1}{n^2}\left(T^2 - (n-1)T\right)$ is a function of T, then by Lehmann-Scheffe theorem, we can conclude that,

$\frac{1}{n^2}\left\{T^2 - (n-1)T\right\}$ is the UMVUE for $\lambda+\lambda^2$.

Let $h(X)$ be an unbiased estimator of $\frac{1}{\lambda}$.

$\therefore E(h(X)) = \frac{1}{\lambda}$

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

$\Rightarrow e^{-\lambda}\cdot h(0) + \frac{e^{-\lambda}\cdot\lambda}{1}\cdot h(1) + \frac{e^{-\lambda}\cdot\lambda^2}{2} h(2) + \cdots = \frac{1}{\lambda}$ ——— (1)

In equation (1), the coefficient of λ^{-1} does not match in both the sides. Hence the above equation is inconsistent.

$\therefore h$ can't be an unbiased estimator of $\frac{1}{\lambda}$.

If $X_1, \ldots, X_n$ be a r.s. from $P(\lambda)$, then $\bar{X} = \frac{1}{n}\sum X_i$ is the MLE of λ.

$\therefore$ By invariance property of MLE, $\frac{1}{\bar{X}}$ is the MLE of $\frac{1}{\lambda}$, provided $\bar{X} > 0$.

(7). Let $X_1, \ldots, X_n$ be a r.s. from $R(-\theta, \theta)$, $\theta > 0$. Find an MLE for θ. Verify whether it is consistent or not. '10 (4).

Solution:— The Likelihood function of $X_1, \ldots, X_n$ is given by

$L(\theta|\underset{\sim}{x}) = \left(\frac{1}{2\theta}\right)^n, \quad -\theta < x_i < \theta \;\forall\; i = 1(1)n.$

Now, note that $L(\theta|\underset{\sim}{x})$ is maximum whenever θ is minimum.

Here $x_i < |\theta|$ iff $|x_i| < \theta \;\forall\; i = 1(1)n$

$\therefore \theta > \max\{|x_1|, \ldots, |x_n|\}$

$\therefore$ MLE of θ is $|X_{(n)}|$.

Now, we have to check whether $|X_{(n)}|$ is consistent or not.

$$P_\theta\left[\,|X_{(n)}-\theta| < \epsilon\,\right], t>0$$

$$= P_\theta\left[\theta-\epsilon < |X_{(n)}| < \theta+\epsilon\right]$$

$$= P_\theta\left(|X_{(n)}| < \theta\right) - P_\theta\left(|X_{(n)}| < \theta-\epsilon\right)$$

$$= 1 - P_\theta\left(-\theta+\epsilon < X_{(n)} < \theta-\epsilon\right)$$

Here note that,

$$P(X_{(n)} < x) = \left\{\int_{-\theta}^{x} \frac{dx}{2\theta}\right\}^n = \frac{(x+\theta)^n}{(2\theta)^n}$$

$$\therefore f_{X_{(n)}}(x) = \frac{n(x+\theta)^{n-1}}{(2\theta)^n}.$$

$$\therefore P_\theta\left(-\theta+\epsilon < X_{(n)} < \theta-\epsilon\right)$$

$$= n\int_{-\theta+\epsilon}^{\theta-\epsilon} \frac{(x+\theta)^{n-1}}{(2\theta)^n}\,dx = \frac{n(x+\theta)^n}{n(2\theta)^n}\Bigg|_{-\theta+\epsilon}^{\theta-\epsilon}$$

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

$$\longrightarrow 0 \quad \text{as } n\to\infty$$

$$\therefore P_\theta\left[\,|X_{(n)}-\theta| < \epsilon\,\right] \longrightarrow 1 \text{ as } n\to\infty$$

$\therefore X_{(n)}$ is consistent for θ.

(8). Find under what conditions the variance of an unbiased estimator attains the Cramer-Rao lower bound. (5)'08

Solution:- Let $x_1, \ldots, x_n$ be a r.s. drawn from a popln. with p.d.f. $f(\underset{\sim}{x}, \theta)$, where θ is the unknown parameter, $\theta \in \Theta$.

Let $T(\underset{\sim}{X})$ be an unbiased estimator for a real valued parametric function $\gamma(\theta)$.

We make assumptions and following regularity conditions:

i) $\frac{\partial}{\partial\theta} f_\theta(\underset{\sim}{x})$ exists for all $\underset{\sim}{x} \in \mathcal{X}$ and $\theta \in \Theta$.

ii) $\frac{\partial}{\partial\theta}\int_{\mathcal{X}} f_\theta(\underset{\sim}{x}) = \int_{\mathcal{X}} \frac{\partial}{\partial\theta} f_\theta(\underset{\sim}{x})$

iii) $I_\theta = E_\theta\left[\frac{\partial}{\partial\theta}\ln f_\theta(\underset{\sim}{x})\right]^2 < \infty$

iv) The support of $\underset{\sim}{x}$ is independent of the parameter θ.

If the conditions hold then the Cramer-Rao Lower bound for the variance of the unbiased estimator is given by,

$$V_\theta(T) \geq \frac{[\gamma'(\theta)]^2}{I(\theta)} \quad \text{———} (*)$$

To prove the above result we proceed in the following way,

Since $f_\theta(\underset{\sim}{x})$ is a pdf,

$$\int_{\mathscr{X}} f_\theta(\underset{\sim}{x})\, d\underset{\sim}{x} = 1$$

$$\Rightarrow \frac{\partial}{\partial\theta} \int_{\mathscr{X}} f_\theta(\underset{\sim}{x})\, d\underset{\sim}{x} = 0$$

$$\Rightarrow \int_{\mathscr{X}} \frac{\partial}{\partial\theta} f_\theta(\underset{\sim}{x})\, d\underset{\sim}{x} = 0$$

$$\Rightarrow \int_{\mathscr{X}} \left\{\frac{\partial}{\partial\theta} \ln f_\theta(\underset{\sim}{x})\right\} f_\theta(\underset{\sim}{x})\, d\underset{\sim}{x} = 0$$

$$\Rightarrow E_\theta\left[\frac{\partial}{\partial\theta} \ln f_\theta(\underset{\sim}{x})\right] = 0$$

Now, $T(\underset{\sim}{X})$ be the unbiased estimator of $\gamma(\theta)$,

$$E_\theta(T) = \gamma(\theta)$$

$$\Rightarrow \int_{\mathscr{X}} T(\underset{\sim}{X}) f_\theta(\underset{\sim}{x})\, d\underset{\sim}{x} = \gamma(\theta)$$

$$\Rightarrow \int_{\mathscr{X}} T(\underset{\sim}{X}) \left\{\frac{\partial}{\partial\theta} \ln f_\theta(\underset{\sim}{x})\right\} f_\theta(\underset{\sim}{x})\, d\underset{\sim}{x} = \gamma'(\theta)$$

$$\Rightarrow E_\theta\left[T(\underset{\sim}{X}) \frac{\partial}{\partial\theta} \ln f_\theta(\underset{\sim}{x})\right] = \gamma'(\theta)$$

Now,

$$\mathrm{Cov}_\theta\left(T(\underset{\sim}{X}), \frac{\partial}{\partial\theta} \ln f_\theta(\underset{\sim}{x})\right) = \gamma'(\theta)$$

Now, by C-S inequality

$$\mathrm{Var}_\theta(T)\, V_\theta\left[\frac{\partial}{\partial\theta} \ln f_\theta(\underset{\sim}{x})\right] \geq [\gamma'(\theta)]^2$$

$$\Rightarrow V_\theta(T) \geq \frac{[\gamma'(\theta)]^2}{E_\theta\left[\frac{\partial}{\partial\theta} \ln f_\theta(\underset{\sim}{x})\right]^2}$$

Now for C-S inequality, the equality arises if for the two variables x and y, $x = ky$

So, the equality arises in Cramer - Rao inequality if

$$\frac{\partial}{\partial\theta} \ln f_\theta(\underset{\sim}{x}) = k(\theta)\left[T(\underset{\sim}{X}) - \gamma(\theta)\right].$$

X ——— X

(9). Consider a r.s. of size n from $N(\mu, \sigma^2)$. μ, σ are unknown. Find the UMVUE of σ^2. (5)

Solution:- The joint PDF is given by

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \frac{1}{(\sigma\sqrt{2\pi})^n} \cdot \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right] \quad ; x_i \in \mathbb{R}\ \forall i = 1(1)n.$$

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

$$= g(T(\underset{\sim}{X}), \sigma^2) \cdot h(\underset{\sim}{x}) \longrightarrow (1)$$

where, $h(\underset{\sim}{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n$,

$$g(T(\underset{\sim}{X}), \sigma^2) = \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

and $T(\underset{\sim}{X}) = \frac{1}{\sigma^2}\sum(x_i-\mu)^2 = s^2$, say

Now, $X_i \sim N(\mu, \sigma^2)$ and $\frac{X_i-\mu}{\sigma} \sim N(0,1)$

$\therefore Z_i = \frac{(X_i-\mu)^2}{\sigma^2} \overset{iid}{\sim} \chi^2_1$ [By the defn. of χ^2-distn.]

$\therefore$ By the reproductive property of χ^2-distn.

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$$

$$\therefore E\left(\sum_{i=1}^{n} Z_i^2\right) = n$$

$$\Rightarrow E(s^2) = \sigma^2$$

$\therefore s^2$ is an unbiased estimator for σ^2. Again from (1) and by Neyman-Fisher factorization theorem, we can say that s^2 is a sufficient statistic for σ^2.

Now, since normal distribution belong to the complete family, s^2 is also a complete sufficient statistic and as well as an unbiased estimator of σ^2.

$\therefore$ By Lehmann-Scheffe theorem, we can say that s^2 is the UMVUE for σ^2.

X ——— X

INTERVAL ESTIMATION

Interval Estimation

Introduction:- Estimation of parameter by a single value is reffered to as a point estimation. In a wide variety of inference problems one is not interested in point estimation or testing of hypothesis of the parameter. Rather one wishes to establish a level or an upper bound or both, for the parameter. As alternative procedure is to give an interval within which the parameter may be supposed to lie with certain probability or confidence. This is called Interval Estimation.

Let $X_1, X_2, \ldots, X_n$ be a r.s. from $N(\mu, \sigma^2)$.

Then $$\frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}$$

$$\& \ P\left[-t_{\alpha/2;n-1} < \frac{\sqrt{n}(\bar{X}-\mu)}{S} < t_{\alpha/2;n-1}\right] = 1-\alpha$$

$$\Leftrightarrow P\left[\bar{X} - \frac{S}{\sqrt{n}} t_{\alpha/2;n-1} < \mu < \bar{X} + \frac{S}{\sqrt{n}} t_{\alpha/2;n-1}\right] = 1-\alpha.$$

If a large no. of samples, each of size n, are thrown from a popln. and if for each samples the above interval is determined, then in about $100(1-\alpha)\%$ of cases the interval will include μ.

For an observed sample $x_1, x_2, \ldots, x_n$, one will therefore justify in saying that the interval $\left(\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha/2;n-1} ; \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha/2,n-1}\right)$ provides a guess on estimation regarding μ. The no. of $(1-\alpha)$ is a measure of trust or confidence.

Definition:- 1. An interval $I(\underset{\sim}{x})$ which is a subset of $\Omega \subseteq \mathbb{R}$ is said to constitute a confidence interval with confidence coefficient $(1-\alpha)$, if $P[I(\underset{\sim}{x}) \ni \theta] = 1-\alpha \ \forall \ \theta \in \Omega$.

i.e. the random interval $I(\underset{\sim}{x})$ covers the true parameter with probability $= 1-\alpha$.

2. A subset $S(\underset{\sim}{x})$ of $\Omega \subseteq \mathbb{R}^k$ is said to constitute a confidence set at confidence $(1-\alpha)$ if $P[S(\underset{\sim}{x}) \in \theta] \geq 1-\alpha \ \forall \ \theta \in \Omega$.

Methods of finding Confidence interval:–

Let θ be a parameter & T be a statistic based on a r.s. of size n from a popln. Most often it is possible to find a function $\psi(T,\theta)$ whose distn. is independent of θ.

Then $P\left[\psi_{1-\alpha/2} < \psi(T,\theta) < \psi_{\alpha/2}\right] = 1-\alpha$,

where, ψ_α is independent of θ, as distn. of $\psi(T,\theta)$ is indep. of θ.

Now, $\psi_{1-\alpha/2} < \psi(T,\theta) < \psi_{\alpha/2}$ can often be put in the form

$$\theta_1(T) \le \theta \le \theta_2(T) .$$

Then $P\left[\theta_1(T) \le \theta \le \theta_2(T)\right] = 1-\alpha$ & $[\theta_1(T), \theta_2(T)]$ will be the observed value of the interval $[\theta_1(T), \theta_2(T)]$ the confidence interval for θ with confidence coefficient $(1-\alpha)$.

Example 1:– Let $X_1, \ldots, X_n$ be a r.s. from $N(\mu, \sigma^2)$; μ and σ both are unknown. Find confidence interval for

(i) μ (ii) σ, with confidence coefficient $(1-\alpha)$,

(iii) (μ, σ^2) .

Solution:– (i) For confidence interval of μ, we select the statistic $T = \bar{X}$.

Then $\psi(T,\mu) = \dfrac{\sqrt{n}(\bar{X}-\mu)}{s} \sim t_{n-1}$, which is indep. of μ.

$$\text{Now, } 1-\alpha = P\left[-t_{\alpha/2, n-1} < \frac{\sqrt{n}(\bar{X}-\mu)}{s} < t_{\alpha/2, n-1}\right]$$

$$= P\left[\bar{X} - \frac{s}{\sqrt{n}} t_{\alpha/2, n-1} \le \mu \le \bar{X} + \frac{s}{\sqrt{n}} t_{\alpha/2, n-1}\right]$$

Hence $\left(\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha/2, n-1}, \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha/2, n-1}\right)$ is an observed confidence interval for μ with confidence coefficient $(1-\alpha)$.

(ii) For confidence interval of σ^2, we select the statistic

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

Then, $\psi(s^2, \sigma^2) = (n-1)\dfrac{s^2}{\sigma^2} \sim \chi^2_{n-1}$, the distn. is indep. of σ^2.

$$\text{Now, } 1-\alpha = P\left[\chi^2_{1-\alpha/2, n-1} \le (n-1)\frac{s^2}{\sigma^2} \le \chi^2_{\alpha/2, n-1}\right]$$

$$= P\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2, n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2, n-1}}\right]$$

Hence $\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\chi^2_{\alpha/2,\, n-1}}, \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\chi^2_{1-\alpha/2,\, n-1}} \right)$ is an observed C.I for σ^2 with confidence coefficient $(1-\alpha)$.

(iii) $$P\left[\bar{X} - \frac{s}{\sqrt{n}} t_{\alpha_1/2,\, n-1} \le \mu \le \bar{X} + \frac{s}{\sqrt{n}} t_{\alpha_1/2,\, n-1}\right] = 1-\alpha_1,$$

& $$P\left[\frac{(n-1)s^2}{\chi^2_{\alpha_2/2,\, n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha_2/2,\, n-1}}\right] = 1-\alpha_2.$$

Note that :- (Boolsen Prob.): $P(A\cap B) \geq P(A)+P(B)-1$.

$$\therefore P\left[\bar{X} - \frac{s}{\sqrt{n}} t_{\alpha_1/2,\, n-1} \le \mu \le \bar{X} + \frac{s}{\sqrt{n}} t_{\alpha_1/2,\, n-1};\ \frac{(n-1)s^2}{\chi^2_{\alpha_2/2,\, n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha_2/2,\, n-1}}\right]$$

$$\geq (1-\alpha_1)+(1-\alpha_2)-1$$

$$= 1-\alpha, \text{ where } \alpha = \alpha_1+\alpha_2.$$

Hence, $$S(\underset{\sim}{x}) = \left(\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha_1/2,\, n-1}, \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha_1/2,\, n-1}\right) \times \left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha_2}{2},\, n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha_2}{2},\, n-1}}\right)$$

Example 2:- Let $X_1, \ldots, X_n$ be a r.s. from $U(0,\theta)$, $\theta>0$. Find a confidence interval for θ with confidence coefficient $(1-\alpha)$, based on $X_{(n)}$.

Sol. The p.d.f of $X_{(n)}$ is
$$f_{X_{(n)}}(x) = \begin{cases} \frac{nx^{n-1}}{\theta^n} & \text{if } 0<x<\theta \\ 0 & \text{ow} \end{cases}$$

The pdf $\psi(X_{(n)},\theta) = \frac{X_{(n)}}{\theta} = T$ is
$$g(t) = \begin{cases} nt^{n-1}, & 0<t<1 \\ 0, & \text{ow} \end{cases}$$
which is independent of θ.

Now, $P\left[c < \psi(X_{(n)},\theta) < 1\right] = 1-\alpha.$

$\Rightarrow \int_c^1 nt^{n-1}\,dt = 1-\alpha$, where c is the critical region.

$\Rightarrow 1-c^n = 1-\alpha$, i.e. $c = \alpha^{1/n}$.

Note that, $\alpha^{1/n} < \psi(X_{(n)}, \theta) = \frac{X_{(n)}}{\theta} < 1$

$$\Rightarrow \alpha^{-1/n} > \frac{\theta}{X_{(n)}} > 1$$

i.e. $X_{(n)} < \theta < \alpha^{-1/n} \cdot X_{(n)}$

Hence $\left[X_{(n)}, \alpha^{-1/n} X_{(n)}\right]$ is a C.I. for θ with confidence coefficient $(1-\alpha)$.

<u>Example 3.</u> Consider a r.s. of size n from the rectangular distribution $f(x,\theta) = \begin{cases} 1/\theta & \text{if } 0<x<\theta \\ 0 & \text{ow} \end{cases}$

If r be the sample range then ξ is given by

$$\xi^{n-1}\left[n-(n-1)\xi\right] = \alpha.$$

S.T. r and r/ξ are confidence limit to θ with confidence coefficient $(1-\alpha)$.

<u>Sol.</u> Here, $r = X_{(n)} - X_{(1)}$

The pdf of r, is $f_r(y) = \begin{cases} n(n-1)y^{n-2}(1-y), & \text{if } 0<y<\theta \\ 0 & \text{, ow} \end{cases}$

The pdf of $\psi(r,\theta) = U$ is

$$f_U(u) = \begin{cases} n(n-1)u^{n-2}(1-u) & \text{if } 0<u<1 \\ 0 & \text{, ow} \end{cases}$$

which is independent of θ.

Now, $P\left[\xi \le U \le 1\right] = 1-\alpha$.

$$\Rightarrow \int_{\xi}^{1} n(n-1)u^{n-2}(1-u)\,du = 1-\alpha.$$

$$\Rightarrow n(n-1)\int_{\xi}^{1}\left[u^{n-2}-u^{n-1}\right]du = 1-\alpha.$$

$$\Rightarrow \xi^{n-1}\left[n-(n-1)\xi\right] = \alpha.$$

Note that $\{\xi \le U \le 1\} = \left\{\xi \le \frac{r}{\theta} \le 1\right\} = \left\{r \le \theta \le \frac{r}{\xi}\right\}$

Hence, $(r, r\xi^{-1})$ is a random C.I. for θ with confidence coefficient $1-\alpha$, where ξ is $\ni$

$$\xi^{n-1}\left[n-(n-1)\xi\right] = \alpha.$$

Ex. 4. Consider a r.s. of size n from an exponential distn. with pdf $f_X(x) = \begin{cases} \exp[-(x-\theta)] & , \text{if } \theta < x < \infty \\ 0 & , \text{OW} \end{cases}$

show that $P_\theta\left[X_{(1)} + \frac{1}{n}\log\alpha \le \theta \le X_{(1)}\right] \le 1-\alpha$.

and hence suggest a $100(1-\alpha)\%$ confidence interval for θ.

Solution:- The d.f. of $X_{(1)}$ is

$$F_{X_{(1)}}(x) = 1 - P\left[X_{(1)} > x\right] = P\left[X_{(1)} \le x\right]$$

$$= 1 - \left\{P\left[X_{(1)} > x\right]\right\}^n$$

$$= 1 - \left\{e^{-(x_{(1)}-\theta)}\right\}^n$$

$$= 1 - e^{-n(x_{(1)}-\theta)} \quad \text{if } x_{(1)} > \theta.$$

Hence $U = e^{-n(x_{(1)}-\theta)} = 1 - F(x_{(1)}) \sim U(0,1)$.

p.d.f. $f(x) = \frac{d}{dx} F_{X_{(1)}}(x)$

$$= n e^{-n(x-\theta)} \quad \text{if } x > \theta$$

Let $U = e^{-(x_{(1)}-\theta)}$

& $U = e^{-(x-\theta)}$

$\Rightarrow \log U = -(x-\theta) \Rightarrow \frac{1}{U}.dU = -dx$

$\Rightarrow J = \left|\frac{dx}{du}\right| = \frac{1}{U}$.

& $f_U(u) = \begin{cases} nu^{n-1} & \text{if } 0<U<1 \\ 0 & \text{OW} \end{cases}$

Now, $1-\alpha = P\left[c \le U \le 1\right] = \int_c^1 nU^{n-1} du$ $\quad\Big|\quad \alpha = P\left[0 \le U \le c\right] \Rightarrow c = \alpha^{1/n}$.

$\Rightarrow 1-\alpha = 1-c^n$

$\Rightarrow c = \alpha^{1/n}$.

Note that $\alpha^{1/n} \le u \le 1$

$\Rightarrow \alpha^{1/n} \le e^{-(X_{(1)}-\theta)} \le 1$.

$\Rightarrow \frac{1}{n}\log\alpha \le -(X_{(1)}-\theta) \le 0$

$\Rightarrow X_{(1)} + \frac{1}{n}\log\alpha \le \theta \le X_{(1)}$.

Wilk Optimum Criterion Based on the Expected length:–

Let $x_1, \ldots, x_n$ be a r.s. from a pdf $f_\theta(x)$ and $T(x_1, \ldots, x_n; \theta) = T_\theta$ be a r.v. where distn. is independent of θ.

Then $P[\lambda_1 < T_\theta < \lambda_2] = 1-\alpha \Rightarrow P[\underline{\theta}(\underset{\sim}{X}) < \theta < \bar{\theta}(\underset{\sim}{X})] = 1-\alpha.$

For each T_θ, λ_1 and λ_2 can be chosen in many ways. We could like to choose λ_1 & λ_2 so that $(\bar{\theta} - \theta)$ is minimum. Such an interval is a $(1-\alpha)$ level shortest length confidence interval based on T_θ. An alternative to minimize the length of C.I. is to minimize the expected length $E[\bar{\theta}(\underset{\sim}{X}) - \underline{\theta}(\underset{\sim}{X})]$.

<u>Definition I</u>:– A $(1-\alpha)$ level of CI $[\underline{\theta}(\underset{\sim}{X}), \bar{\theta}(\underset{\sim}{X})]$ is said to be <u>shorter</u> that another $(1-\alpha)$ level of C.I. $[\underline{\theta}^*(\underset{\sim}{X}), \bar{\theta}^*(\underset{\sim}{X})]$ if

$$E[\bar{\theta}(\underset{\sim}{X}) - \underline{\theta}(\underset{\sim}{X})] < E_\theta[\bar{\theta}^*(\underset{\sim}{X}) - \underline{\theta}^*(\underset{\sim}{X})] \quad \forall\ \theta \in \Omega.$$

<u>Example</u>:– Let $(x_1, \ldots, x_n)$ be a r.s. from $N(\mu, \sigma^2)$. Find the shortest length C.I. for (i) μ based on $\bar{X}$
(ii) σ^2 based on s^2

<u>Solution</u>:– (i) <u>Pivotal Statistic</u>:– $T_\mu = \dfrac{\sqrt{n}(\bar{X}-\mu)}{s} \sim t_{n-1}$, which is independent of μ.

Then
$$1-\alpha = P\left[a < \frac{\sqrt{n}(\bar{X}-\mu)}{s} < b\right]$$
$$= P\left[\bar{X} - b.\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} - a.\frac{s}{\sqrt{n}}\right]$$

$\therefore$ Expected length, $E(\hat{L}) = (b-a)\dfrac{E(s)}{\sqrt{n}}$.

To minimize the expected length subject to the restriction
$$\int_a^b f_{n-1}(t)\,dt = 1-\alpha.$$

Now,
$$\frac{\partial E(\hat{L})}{\partial a} = \left(\frac{\partial b}{\partial a} - 1\right)\frac{E(s)}{\sqrt{n}}.$$

$$\& \quad f_{n-1}(b)\frac{\partial b}{\partial a} - f_{n-1}(a) = 0 \Rightarrow \frac{\partial b}{\partial a} = \frac{f_{n-1}(a)}{f_{n-1}(b)}.$$

$$\therefore \frac{\partial E(\hat{L})}{\partial a} = \left[\frac{f_{n-1}(a)}{f_{n-1}(b)} - 1\right]\frac{E(s)}{\sqrt{n}} = 0$$

$$\Rightarrow f_{n-1}(a) = f_{n-1}(b) \Rightarrow a = -b.$$

$$\Rightarrow b = t_{\alpha/2, n-1} \quad [\because a = -b]$$

Note that shortest expected length C.I. from μ with CI $(1-\alpha)$ based on $\bar{X}$ is $\left(\bar{X} - t_{\alpha/2, n-1}\cdot\dfrac{s}{\sqrt{n}},\ \bar{X} + t_{\alpha/2, n-1}\cdot\dfrac{s}{\sqrt{n}}\right)$.

(ii) Pivotal statistic:–

$$T_{\sigma^2} = \frac{\Sigma (x_i - \bar{x})^2}{\sigma^2} = (n-1)\frac{s^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Since, $P\left[a < (n-1)\frac{s^2}{\sigma^2} < b\right] = 1-\alpha.$

$$\Rightarrow P\left[(n-1)\frac{s^2}{b} < \sigma^2 < (n-1)\frac{s^2}{a}\right] = 1-\alpha.$$

$\therefore$ Expected length, $E(\hat{L}) = \left(\frac{1}{a} - \frac{1}{b}\right) E\left[\overline{n-1}\, s^2\right]$

To minimize $E(\hat{L})$ subject to the condition

$$\int_a^b f_{n-1}(t)\,dt = 1-\alpha.$$

$$\Rightarrow \frac{\partial b}{\partial a} = \frac{f_{n-1}(a)}{f_{n-1}(b)}.$$

We have, $\frac{\partial E(\hat{L})}{\partial a} = \left[-\frac{1}{a^2} + \frac{1}{b^2}\cdot\frac{\partial b}{\partial a}\right] E\left[\overline{n-1}\, s^2\right]$

$$= \left[-\frac{1}{a^2} + \frac{1}{b^2}\cdot\frac{f_{n-1}(a)}{f_{n-1}(b)}\right] E\left[\overline{n-1}\, s^2\right] = 0$$

if, $a^2 f_{n-1}(a) = b^2 f_{n-1}(b).$

$$\Leftrightarrow e^{-a/2}\, a^{\frac{n+1}{2}} = e^{-b/2}\, b^{\frac{n+1}{2}}.$$

Numerical solution may be used for finding a & b. Let $\hat{a}, \hat{b}$ be the solution, then $\left[\frac{(n-1)s^2}{b}, \frac{(n-1)s^2}{a}\right]$ is the shortest expected length C.I. of σ^2.

Ex. (2):- Let $X_1, \ldots, X_n$ be r.s. from $U(0,\theta)$. Find the shortest expected length C.I. of θ based on $X_{(n)}$.

Solution:-

$$T_\theta = \frac{X_{(n)}}{\theta} \text{ has the d.f. } F(t) = t^n,\ 0<t<1,$$

which is independent of θ.

Now, $$1-\alpha = P\left[a < \frac{X_{(n)}}{\theta} < b\right]$$

$$= P\left[\frac{X_{(n)}}{b} < \theta < \frac{X_{(n)}}{a}\right]$$

$$E(\hat{L}) = \left(\frac{1}{a} - \frac{1}{b}\right) E(X_n)$$

To minimize the expected length C.I. for θ based on $X_{(n)}$ $\ni$

$$E(\hat{L}) = F(b) - F(a) = 1-\alpha.$$

$$\Rightarrow b^n - a^n = 1-\alpha \quad \text{——— (1)}$$

For $(1-\alpha)^{1/n} \le b \le 1$

$$\frac{\partial E(\hat{L})}{\partial b} = \left(-\frac{1}{a^2}\cdot\frac{\partial a}{\partial b} + \frac{1}{b^2}\right) E(X_{(n)})$$

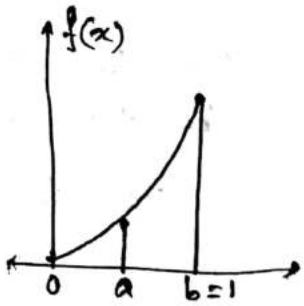
& $nb^{n-1} - na^{n-1} \cdot \frac{\partial a}{\partial b} = 0$ [from (1)]

$$\Rightarrow \frac{\partial a}{\partial b} = \frac{b^{n-1}}{a^{n-1}}.$$

$$\therefore \frac{\partial E(\hat{L})}{\partial b} = \left(-\frac{1}{a^2}\cdot\frac{b^{n-1}}{a^{n-1}} + \frac{1}{b^2}\right) E(X_n)$$

$$= \left(\frac{1}{b^2} - \frac{b^{n-1}}{a^{n+1}}\right) E(X_n)$$

$$= \frac{a^{n+1} - b^{n+1}}{b^2 a^{n+1}} E(X_n) < 0$$



$\Rightarrow E(\hat{L}) \downarrow$ at b.

$\Rightarrow$ min. of $E(\hat{L})$ occurs at $b=1$.

& $1 - a^n = 1-\alpha$

$\Rightarrow a = \alpha^{1/n}$.

Hence the shortest expected length C.I. of θ based on $X_{(n)}$ is $\left[X_{(n)},\ X_{(n)} \cdot \alpha^{-1/\alpha}\right]$.

Scanned by CamScanner

Confidence Estimation

Definition ((1-α) level confidence sets)

Let $\theta \in \Theta \subseteq \mathbb{R}^k$ and $0 < \alpha < 1$. A family of random subsets $S(\underset{\sim}{X})$ of Θ is called a family of confidence sets at confidence level $(1-\alpha)$ if

$$P_\theta \{ S(\underset{\sim}{X}) \ni \theta \} \geq 1-\alpha \;\; \forall \underset{\sim}{\theta} \in \Theta . \quad \text{——— (1)}$$

The quantity $\inf\limits_{\theta \in \Theta} P_\theta \{ S(\underset{\sim}{X}) \ni \underset{\sim}{\theta} \}$

is called confidence coefficient associated with random set $S(\underset{\sim}{X})$.

Definition:– (Uniformly Most Accurate Family of Confidence Sets)

A family of confidence sets $\{ S(\underset{\sim}{X}) \}$ is said to be a UMA family of confidence sets if

$$P_\theta \{ S(\underset{\sim}{X}) \ni \underset{\sim}{\theta} \} \geq 1-\alpha \;\; \forall \underset{\sim}{\theta} \in \Theta$$

and $P_{\theta'} \{ S(\underset{\sim}{X}) \ni \underset{\sim}{\theta} \} \leq P_{\theta'} \{ S'(\underset{\sim}{X}) \ni \underset{\sim}{\theta} \} \;\; \forall \underset{\sim}{\theta}, \underset{\sim}{\theta}' \in \Theta$.

for all $S'(\underset{\sim}{X})$ satisfying equation (1), i.e. $S'(\underset{\sim}{X})$ is any other family of $(1-\alpha)$ level confidence sets.

UMAU Confidence Sets:– A family $\{ S(\underset{\sim}{x}) \}$ of confidence sets for a parameter θ is said to be unbiased at level $(1-\alpha)$ if

$$P_\theta \{ S(\underset{\sim}{X}) \ni \theta \} \geq 1-\alpha \;\; \forall \theta \in \Theta .$$

and $P_{\theta'} \{ S(\underset{\sim}{X}) \ni \theta \} \leq 1-\alpha \;\; \forall \theta, \theta' \in \Theta$.

If $S^*(\underset{\sim}{X})$ is a family of $(1-\alpha)$ level unbiased confidence sets that minimizes $P_\theta \{ S(\underset{\sim}{X}) \ni \theta' \} \;\; \forall \theta, \theta' \in \Theta$.

Then $S^*(\underset{\sim}{X})$ is called uniformly most accurate unbiased (UMAU) family of confidence sets at level $(1-\alpha)$.

▨ **Discuss by theorem the relationship between UMP unbiased size-α acceptance region and UMAU family of confidence set at let $1-\alpha$.**

Solution:– **Theorem:–** Consider the testing problem $H_0 : \theta = \theta_0$ Vs. $H_1 : \theta \neq \theta_0$ for each $\theta_0 \in \Theta$. Let $A(\theta_0)$ be the UMP unbiased size α acceptance region for this problem. Then $S(\underset{\sim}{x}) = \{ \theta \mid \underset{\sim}{x} \in A(\theta) \}$ is a UMP unbiased family of confidence sets at level $(1-\alpha)$.

Proof:- Let the UMP unbiased size-α test be given by $\phi_0(\underset{\sim}{x})$.

Unbiasedness gives $E_{\theta'}\phi_0(\underset{\sim}{X}) \geq \alpha \ \forall \theta' \in H_1(\theta_0)$

$$\Rightarrow E_{\theta'}(1-\phi_0(\underset{\sim}{X})) \leq 1-\alpha \ \forall \theta' \in H_1(\theta_0)$$

$$\Rightarrow P_{\theta'}(\underset{\sim}{X} \in A(\theta)) \leq 1-\alpha$$

$$\Rightarrow P_{\theta'}(S(\underset{\sim}{X}) \ni \theta) \leq 1-\alpha.$$

shows that $S(\underset{\sim}{X})$ is unbiased.

Next, consider any other unbiased size-α test $\phi^*(\underset{\sim}{X})$, with acceptance region $A^*(\theta)$; we get a corresponding $(1-\alpha)$ level family of unbiased confidence sets $S^*(\underset{\sim}{X})$, i.e.

$$P_{\theta'}(S^*(\underset{\sim}{X}) \ni \theta) \leq 1-\alpha \ \forall \theta' \in H_1(\theta_0)$$

The test $\phi_0(\underset{\sim}{X})$ has been given to be UMP, therefore

$$E_{\theta'}[\phi_0(\underset{\sim}{X})] \geq E_{\theta'}[\phi^*(\underset{\sim}{X})] \ \forall \theta' \in H_1(\theta_0)$$

or, $E_{\theta'}[1-\phi_0(\underset{\sim}{X})] \leq E_{\theta'}[1-\phi^*(\underset{\sim}{X})]$

or, $P_{\theta'}[\underset{\sim}{X} \in A(\theta)] \leq P_{\theta'}[\underset{\sim}{X} \in A^*(\theta)]$

or, $P_{\theta'}[S(\underset{\sim}{X}) \ni \theta] \leq P_{\theta'}\{S^*(\underset{\sim}{X}) \ni \theta\} \ \forall \theta' \in H_1(\theta_0)$

This follows $S(\underset{\sim}{X})$ is UMA unbiased family of confidence sets at level $(1-\alpha)$.

Ex.(1):- Let $X_1, \dots, X_n$ be a sample from $N(\mu, \sigma^2)$, where σ^2 is known for testing $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$. Find a UMA $(1-\alpha)$ level confidence sets for μ.

Solution:- For testing of Hypothesis $H_0: \mu=\mu_0$ Vs. $H_1: \mu \neq \mu_0$ the UMP unbiased size-α test is given by

$$\phi(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \dfrac{\sqrt{n}|(\bar{x}-\mu_0)|}{\sigma} > c \\ 0 & \text{ow} \end{cases}$$

This test is known as z-test. The constant c can be determined by the size condition

$$E_{\mu_0}[\phi(\underset{\sim}{X})] = \alpha$$

$$\text{or, } P_{\mu_0}\left\{\frac{\sqrt{n}|\bar{X}-\mu_0|}{\sigma} > c\right\} = \alpha/2$$

which gives $c = z_{\alpha/2}$.

Thus, the acceptance region corresponding to this UMP unbiased size-α-test is given by

$$A(\mu_0) = \left\{\underset{\sim}{X} : \frac{\sqrt{n}|\bar{x}-\mu_0|}{\sigma} \leq z_{\alpha/2}\right\}$$

By the above theorem, the UMA unbiased family of confidence sets $S(\underset{\sim}{X})$ at level $(1-\alpha)$ is finally given by

$$S(\underset{\sim}{x}) = \{\mu : \underset{\sim}{x} \in A(\mu)\}$$
$$= \left\{ -z_{\alpha/2} \le \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} \le z_{\alpha/2} \right\}$$
$$= \left\{ -\frac{\sigma}{\sqrt{n}} z_{\alpha/2} \le (\mu - \bar{x}) \le \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \right\}$$
$$= \left(\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \le \mu \le \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \right).$$

Ex.(2):- Let X be a r.v. with the density

$$f_X(x|\theta) = \begin{cases} \frac{1}{\theta} \cdot e^{-x/\theta} & ; \text{if } x > 0 \\ 0 & \text{o/w} \end{cases}$$

where $\theta > 0$. Consider the testing problem $H_0 : \theta = \theta_0$ vs. $H_1 : \theta < \theta_0$. Find out a UMA $(1-\alpha)$ level family of confidence sets corresponding to size-α UMP test.

Sol. The given family belong to the OPEF.
The UMP size-α acceptance region is given by

$$A(\theta) = \{x : T(x) \ge c(\theta)\}$$
$$= \{x : x \ge c(\theta)\}$$

where, we choose $c(\theta)$ by

$$P_{\theta_0}(A(\theta_0)) = 1 - \alpha.$$

or,
$$\int_0^{c(\theta_0)} \frac{1}{\theta_0} \cdot e^{-x/\theta_0} dx = \alpha$$

or,
$$\frac{1}{\theta_0} \left[-\theta_0 e^{-x/\theta_0} \right]_0^{c(\theta_0)} = \alpha$$

$$\Rightarrow e^{-c(\theta_0)/\theta_0} + 1 = \alpha$$

$$\Rightarrow c(\theta_0) = \theta_0 \cdot \log \frac{1}{1-\alpha}, \quad 0 < \alpha < 1.$$

Therefore, the corresponding UMA family of $1-\alpha$ level of confidence sets is given by

$$S(x) = \{\theta : x \in A(\theta)\} = \left\{\theta : x \ge \theta \log \frac{1}{1-\alpha}\right\}$$
$$= \left\{\theta : \theta \le \frac{x}{\log \frac{1}{1-\alpha}}\right\} = \left(0, \frac{x}{\log\left(\frac{1}{1-\alpha}\right)}\right] \quad (\text{since } \theta > 0)$$