BSc II Year

Paper Code: 294

Statistical Inference

Unit 1

UNIFIED SYLLABUS OF STATISTICS
B.Sc. Part- II

Paper I : Statistical Inference

UNIT – I

Point estimation. Characteristics of a good estimator: Unbiasedness, consistency, sufficiency and efficiency. Method of maximum likelihood and properties of maximum likelihood estimators (without proof). Method of minimum Chi-square. Method of Least squares and method of moments for estimation of parameters. Problems and examples.

UNIT – II

Sufficient Statistics, Cramer-Rao inequality and its use in finding MVU estimators. Statistical Hypothesis (simple and composite). Testing of hypothesis. Type I and Type II errors, significance level, power of a test. Definitions of Most Powerful (MP), Uniformly Most Powerful (UMP) and Uniformly Most Powerful Unbiased (UMPU) tests.

UNIT – III

Neyman-Pearson's lemma and its applications for finding most powerful tests for simple hypothesis against simple alternative. Interval estimation – concept of interval estimation confidence interval for mean & variance in case of normal population only.

UNIT-IV

Test of significance – large sample test for proportions and means : (i) single sample, (ii) two independent samples. Tests based on chi-square, t and F distributions.

Statistical Inference

Statistical Inference

- Estimation
  - Point Estimation
  - Interval Estimation
- Testing of Hypothesis

INTRODUCTION:- A sample from the distribution of a population is useful in making inferences about the population characteristics. The process of going from known sample to the unknown popln. has been called statistical inference.

(1) Estimation:— Some features of the popln. in which an investigator is interested, may be knownto him and he may want to make a guess about this features, on the basis of a random sample drawn from the popln. This type of problem is called problem of estimation.

(2) Testing of Hypothesis:— Some tentative information on a feature of the population may be available to the investigator and he may want to see whether the information is tenable in the light of the random sample taken from the population. This type of problem is called the problem of testing of hypothesis.

(1) Concept of Estimation:— The problem of estimation is loosely defined as: assume that some characteristics of the elements of the popln. can be represented by a r.v. X whose PMF or PDF is $f(x, \theta)$ where the functional form of the PMF or PDF is known except the parameter θ, $\theta \in \Omega$. The set Ω is called the parameter space. Let $(x_1, x_2, \dots, x_n)$ be an observed random sample from $f(x, \theta)$. On the basis of the observed random sample, it is desired to estimate the value of the parameter θ. This estimation is done in two ways.

(a) Point Estimation:- The problem of point estimation is to pick or select a statistic $T(\underset{\sim}{x_1}) = T$ that best estimates the parameters.

The numerical value of $T(\underset{\sim}{x})$ when an observed value of $\underset{\sim}{X}$ is $\underset{\sim}{x}$ is called an estimate of θ while such a statistic $T(\underset{\sim}{x})$ is called an estimator of θ. Let (X_1, X_2, X_3) be a random sample from $f(x, \theta)$. Then $\overline{X} = \dfrac{X_1 + X_2 + X_3}{3}$ is an estimator of θ. If the observed sample is $(-1, 1, 3)$, then the sample mean, $\overline{x} = 1$ is an estimate of θ.

(b) Interval Estimation :- The problem of interval estimation is to define 2 statistic $T_1(\underset{\sim}{x})$ and $T_2(\underset{\sim}{x})$ such that (T_1+T_2) constitutes an interval for which the probability can be determined that it contains the parameter θ.

(a) Point Estimation :- It is clear that if any given problem of estimation, we may have a large, often an infinite no. of estimators, we may choose from.

Requirement of good estimator / Measures of quality of the estimator

Clearly we could like the estimator $T(\underset{\sim}{x}) = T$ to be close to θ. Since T is a R.V., the usual measures of closeness $|T-\theta|$ is also a R.V. Example of such measure of closeness are

Part : 1: $P[\,|T-\theta| < \epsilon\,] \quad \forall\ \epsilon > 0$ ——— (1)

Part : 2: $E\left[\,|T-\theta|^r\right]$, for some $r > 0$ ——— (2)

$$\left[\ P[\,|T-\theta|<\epsilon\,] > 1 - \frac{E[\,|T-\theta|^r]}{\epsilon^r}\ \right]$$

We want to be large (1) but to be small (2).

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 (2) by putting $r=2$.

Definition :- Let T is an estimator of θ. The quantity $E(T-\theta)^2$ is defined to be the MSE of estimator T.

Notation :- $MSE_\theta(T) = E[T-\theta]^2$.

Note that, $E[T-\theta]^2$ is a measure of spread of the values of T about the parameter θ. If we are to compare estimators by looking at there respective MSE's, naturally we would prefer (1) with small or smallest MSE.

Here the requirement is to choose T_0 such that $MSE_\theta(T_0) \le MSE_\theta(T)$. for all T, for $\theta \in \Omega$. But such estimator rearly exists.

Note that,
$$\begin{aligned} MSE_\theta(T) &= E(T-\theta)^2 \\ &= E\left[T-E(T)+E(T)-\theta\right]^2 \\ &= E\{T-E(T)\}^2 + \{E(T)-\theta\}^2 \\ &\quad + 2E\{T-E(T)\}\{E(T)-\theta\} \\ &= Var(T) + \{b(\theta,T)\}^2 \end{aligned}$$

Hence, to control MSE, we need to control both $Var(T)$ and $\{b(\theta,T)\}^2$, the quantity $b(\theta,T) = E(T)-\theta$, is called the bias of T in estimating θ.

One approach is to restrict attention to those estimator which have zero bias, i.e. $E(T)=\theta \ \forall \ \theta \in \Omega$.

If $b(\theta,T)=0$, then T is called an unbiased estimator of θ and

$MSE_\theta(T)=Var(T)$.

Now, it is required to find an estimator with uniformly minimum MSE among all unbiased estimator, which is equivalent to finding an estimator with uniformly minimum variance among all unbiased estimator. This is the concept of unbiasedness and minimum variance.

Unbiasedness :-

Definition :- An estimator T is defined to be an unbiased estimator (UE) of θ if $E(T)=\theta \ \forall \ \theta \in \Omega$.

Unbiasedness of T says that T has no systematic error, it neither overestimates nor underestimates θ on an average.

Biasedness :-

Definition :- An estimator T is said to be biased for the parameter θ if $E(T) \neq \theta$ for some $\theta \in \Omega$.

Ex. 1. Unbiased Estimator of population moments :—

Let $X_1, X_2, \ldots, X_n$ be a r.s from a popln. with finite kth order moment $\mu_k' = E(X_1^k)$. Nothing else is known about the popln distribution. Find an unbiased estimator of μ_r', $1 \le r \le k$.

Solution :- Define $m_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r$

Then, $E(m_r') = \frac{1}{n}\sum_{i=1}^{n} E(X_i^r)$

$= \frac{1}{n}\cdot n \cdot E(X_1^r)$, as X_i's are i.i.d. $\Leftrightarrow X_i^r$'s are i.i.d.

$= E(X_1^r)$

$= \mu_r'$, $1 \le r \le k$.

Hence, the sample rth order raw moment is an unbiased estimator (UE) of μ_r', $r=1(1)k$.

Ex. 2. Let $X_1, X_2, \ldots, X_n$ be the random sample from an infinite population with mean μ and variance $\sigma^2 (<\infty)$. Show that

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$ is a biased estimator of σ^2.

Hence, find an UE of σ^2.

Solution: — $\because$

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

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

$$= \frac{1}{n}\left\{\sum_{i=1}^{n} Var(X_i) - n\,Var(\bar{X})\right\}$$

$$= \frac{1}{n}\cdot\left\{n\sigma^2 - \frac{n.\sigma^2}{n}\right\} = \frac{n-1}{n}\sigma^2$$

$$\left[\begin{array}{l} \text{Here, } E(X_i) = \mu, \ Var(X_i) = \sigma^2 \\ E(\bar{X}) = \mu, \ Var(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{\sigma^2}{n}. \end{array}\right]$$

Here, $E(S^2) = \frac{n-1}{n}\cdot\sigma^2 \neq \sigma^2$ $\therefore$ Bias $= E(S^2) - \sigma^2$

$= -\sigma^2/n$

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

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

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

$$\therefore \text{Bias}(\sigma^2, S^2) = E(S^2) - \sigma^2$$

$$= -\frac{1}{n}\sigma^2 \longrightarrow 0 \text{ as } n \to \infty.$$

Ex. 3. Let $X_1, X_2, \ldots, X_n$ be a r.s. from $P(\lambda)$ distn. S.T. for $0 \le \alpha \le 1$,

$T_\alpha = \alpha\bar{X} + (1-\alpha)S^2$ is an UE of λ and comment.

Solution: — We know that $\bar{X}$ and S^2 are UEs of the popln. mean and variance, respectively. Since for $P(\lambda)$ distn., $\bar{X} = S^2 = \lambda$.

Hence, $$E(T_\alpha) = \alpha E(\bar{X}) + (1-\alpha)E(S^2)$$

$$= \alpha.\lambda + (1-\alpha)\lambda$$

$$= \lambda, \quad \alpha \in [0,1]$$

For each $\alpha \in [0,1]$, T_α is an UE of λ. Hence there are infinitely many UEs of λ of the form

$$T_\alpha = \alpha\bar{X} + (1-\alpha)S^2.$$

Uniformly Minimum Variance Unbiased Estimator (UMVUE) :-

[WBSU'11]

Let T_1 and T_2 be two different UEs of θ. then there exists an infinitely many UEs of θ of the form :

$$T_\alpha = \alpha . T_1 + (1-\alpha) T_2 \quad , \quad 0 \le \alpha \le 1 .$$

which of these should we choose?

Here comes the concept of UMVUE.

Definition:—

(a) An estimator T^* is defined to be UMVUE of θ iff

i) $E(T^*) = \theta \quad \forall \theta \in \Omega$.

ii) $Var_\theta(T^*) \le Var_\theta(T) \quad \forall \theta \in \Omega$,

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

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

Hence $T_1 = 2\bar{X}$ is an UE of θ.

$$E[X_{(n)}] = \int_0^\theta x . \frac{nx^{n-1}}{\theta^n} dx \quad \left[\because f_{X_{(n)}}(x) = \begin{cases} \frac{nx^{n-1}}{\theta^n} & , 0<x<\theta \\ 0 & , OW \end{cases} \right]$$

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

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

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

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

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

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

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

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

for $X_{(1)}$:-

$P[X_{(1)} \le x] = 1 - P[X_{(1)} > x]$

$= 1 - P[X_1, \ldots, X_n > x]$

$= 1 - \prod_{i=1}^n P[X_i > x]$ [due to indep.]

$= \{1 - (1 - \frac{x}{\theta})\}^n , 0<x<\theta$

$\therefore f_{X_{(1)}}(x) = \frac{n}{\theta}\left(1 - \frac{x}{\theta}\right)^{n-1}$

$\therefore E(X_{(1)}) = \frac{\theta}{n+1}$

$\therefore E((n+1) X_{(1)}) = \theta$

$\therefore \theta$ is unbiasedly estimated by $(n+1) X_{(1)}$.

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

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

Theorem:- The UMVUE of a 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 $T = \frac{T_1+T_2}{2}$ is also an UE of θ.

Hence, $var(T) \geq \gamma$

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

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

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

$\Rightarrow \rho \geq 1$, but $|\rho| \leq 1$.

Hence, $\rho = 1$.

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

Now, $E(T_1) = a + b.E(T_2)$

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

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

$[\; V(T_1) = b^2 V(T_2) \Rightarrow b^2 = 1, b>0 \Rightarrow b=1$. and

$E(T_1) = a + bE(T_2) \Rightarrow \theta = a + 1.\theta \Rightarrow a=0\;]$

Hence $T_1 = T_2$ with prob. 1.

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

Ex.2. Let T_1, 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, $V(T) \geq \sigma^2$

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

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

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

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

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

FURTHER PROBLEMS :—

Ex. 1. Estimating p^2 for Bernoulli distribution

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

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

Solution:-

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

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

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

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

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

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

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

Then by definition,

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

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

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

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

Clearly, (i) is an identity in p.

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

$1 = 0 \rightarrow$ absurd

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

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

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

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

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

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

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

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

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

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

By uniqueness of powers series, we have

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

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

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

Ex. 3.

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

Solution:—

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[For a normal sample, $\bar{X}$ and s^2 are independently distributed]

N.T. $\alpha \bar{X} + (1-\alpha) s^2$ is an UE of μ, $0 \le \alpha \le 1$.

Ex.7. Let $X_1, X_2, \ldots, X_n$ be a r.s. from the PDF

$$f(x) = \begin{cases} \theta x^{\theta - 1}, & 0 < x < 1 \\ 0, & \text{ow} \end{cases}, \quad \text{where } \theta > 0.$$

Find an UE of (i) $\frac{1}{\theta}$, (ii) θ.

Solution:- i) Let $Z_i = -2\theta \ln X_i$, then $X_i = e^{-\frac{Z_i}{2\theta}}$

The PDF of Z_i is,

$$f_{Z_i}(z_i) = \begin{cases} \theta \left(e^{-\frac{z_i}{2\theta}} \right)^{\theta - 1} \left| \frac{d}{dz_i} \left(e^{-\frac{z_i}{2\theta}} \right) \right|, & \text{if } 0 < z_i < \infty \\ 0, & \text{ow} \end{cases}$$

$$= \begin{cases} \frac{1}{2} e^{-z_i/2}, & 0 < z_i < \infty \\ 0, & \text{ow} \end{cases}$$

$$\Rightarrow Z_i \overset{iid}{\sim} \chi_2^2 \quad \forall\, i = 1(1)n.$$

$$\Rightarrow \sum_{i=1}^{n} Z_i \sim \chi^2_{2n}$$

i.e. $Y_i = \sum_{i=1}^{n} (-2\theta \ln X_i) \sim \chi^2_{2n}$

Now, $E\left(\sum_{i=1}^{n} -2\theta \ln X_i \right) = 2n$

$$\Rightarrow E\left(-\frac{1}{n} \sum_{i=1}^{n} \ln X_i \right) = \frac{1}{\theta}$$

$$\Rightarrow T_1 = \frac{1}{n} \sum_{i=1}^{n} -\ln X_i \text{ is an UE of } \frac{1}{\theta}.$$

ii) Now, $E\left(\frac{1}{Y}\right) = E\left(\frac{1}{\chi^2_{2n}} \right) = 2^{-1} \frac{\Gamma\left(\frac{2n}{2} - 1\right)}{\Gamma\left(\frac{2n}{2}\right)}$ if $n > 1$

$$= \frac{1}{2} \cdot \frac{\Gamma(n-1)}{\Gamma(n)} = \frac{1}{2(n-1)}, \quad n > 1.$$

$$\Rightarrow E\left(\frac{1}{\sum_{i=1}^{n} -2\theta \ln X_i} \right) = \frac{1}{2(n-1)}, \quad n > 1$$

$$\Rightarrow E\left(\frac{n-1}{\sum_{i=1}^{n} -\ln X_i} \right) = \theta, \quad n > 1.$$

$$\Rightarrow T_2 = \frac{n-1}{\sum_{i=1}^{n} -\ln X_i} \text{ is an UE of } \theta.$$

Ex. 8. Unbiased estimator may sometimes be absurd.

Give an example of Absurd Unbiased estimator.

Solution:- Let X be a single observation of $P(\lambda)$. If possible, let, $T(X)$ be an UE of $e^{-3\lambda}$.

Then $E[T(X)] = e^{-3\lambda}$, $\forall \lambda > 0$

$$\Rightarrow \sum_{x=0}^{\infty} T(x) \cdot e^{-\lambda} \cdot \frac{\lambda^x}{x!} = e^{-3\lambda}$$

$$\Rightarrow \sum_{x=0}^{\infty} T(x) \cdot \frac{\lambda^x}{x!} = e^{-2\lambda} = \sum_{x=0}^{\infty} \frac{(-2\lambda)^x}{x!}, \lambda > 0$$

By uniqueness of Powers series, we have

$$\frac{T(x)}{x!} = \frac{(-2)^x}{x!} \quad \forall x = 0,1,2,\ldots$$

$$\Rightarrow T(x) = (-2)^x \quad \forall x = 0,1,2,\ldots$$

Hence, $T(x) = (-2)^x$ is the unique UE of $e^{-3\lambda}$.

N.T. $T(x) = (-2)^x = \begin{cases} 2^x, & x = 0,2,4,\ldots \\ -2^x, & x = 1,3,5,\ldots \end{cases}$

Hence, $T(X)$ is UE but it takes negative values in estimating a positive parameter $e^{-3\lambda}$. This is an example of absurd UE.

Remark:- (1) Here $T(X) = (-2)^X$ is the only ~~on~~ unique UE of $e^{-3\lambda}$. Hence, $T(X) = (-2)^X$ is the UMVUE of $e^{-3\lambda}$.

(2) For $X \sim P(\lambda)$, $P_X(t) = e^{\lambda(t-1)}$, $t \in R$

$$\Rightarrow E[t^X] = e^{\lambda(t-1)}, t \in R$$

Put, $t = -2$,

$$E[(-2)^X] = e^{-3\lambda}.$$

Ex. 9. If $X \sim Bin(n, p)$, then show that only polynomial in p of degree $\leq n$ are unbiasedly estimable.

Solution:- [A parametric function $\psi(\theta)$ is unbiasedly estimable if $E\{T(X)\} = \psi(\theta)$, for some $T(X)$, $\forall \theta \in \Omega$.]

Let $\psi(p)$ be an unbiasedly estimable parametric function.

Then $\exists$ a statistic $T(X) \ni$

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

$$= \sum_{x=0}^{n} T(x) \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} T(x) \cdot \binom{n}{x} p^x \left\{ \sum_{k=0}^{n-x} \binom{n-x}{k} (-p)^k \right\}$$

$$= \sum_{x=0}^{n} \sum_{k=0}^{n-x} (-1)^k T(x) \binom{n}{x} \binom{n-x}{k} p^{x+k},$$ which is a polynomial in p of degree $\leq n$.

Remark:- N.T. (i) $\sqrt{p}$, (ii) $\frac{1}{p}$, (iii) e^p, (iv) $\log p$ are not polynomials and hence not unbiasedly estimable. If $X \sim B(1, p)$, then only linear function in p are unbiasedly estimable. Hence, p^2, a 2nd degree polynomial is not unbiasedly estimable.

Best Linear Unbiased Estimator (BLUE) :—

Let $x_1, x_2, \ldots, x_n$ be a r.s. from a population with mean μ and variance $\sigma^2 (< \infty)$. Then an estimator $T = \sum_{i=1}^{n} a_i x_i$ is called a linear estimator. A linear estimator $T = \sum_{i=1}^{n} a_i x_i$ is unbiased for μ

iff $E(T) = \mu \;\forall\; \mu$

iff $\left(\sum_{i=1}^{n} a_i\right)\mu = \mu \;\forall\; \mu$

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

[The estimator, $T = \sum_{i=1}^{n} a_i e^{x_i}$ is not linear estimator also, $T_3 = \bar{x}^2$, $T_4 = s^2$ are linear estimators.]

Definition:— A linear unbiased estimator $T = \sum_{i=1}^{n} a_i x_i$ with $\sum_{i=1}^{n} a_i = 1$ of μ that has the minimum variance among all linear unbiased estimators of μ, is called the BLUE of μ.

Theorem:— If $x_1, x_2, \ldots, x_n$ be a r.s. from a population with mean μ and variance σ^2, show that the sample mean $\bar{x}$ is the BLUE of μ. [WBSU'11]

Proof:— BLUE of μ is the estimator which has the minimum variance in the class $\mathcal{C} = \left\{ T : T = \sum_{i=1}^{n} a_i x_i,\; \sum_{i=1}^{n} a_i = 1 \right\}$ of all linear UEs of μ.

Note that, $Var(T) = \left(\sum_{i=1}^{n} a_i^2\right)\sigma^2$, as X_i's are iid and $\sum_{i=1}^{n} a_i = 1$.

To minimize $Var(T) = \sigma^2\left(\sum_{i=1}^{n} a_i^2\right)$ subject to $\sum_{i=1}^{n} a_i = 1$,

By C-S inequality,

$$\left(\sum_{i=1}^{n} a_i^2 . 1\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} 1^2\right)$$

$$\Rightarrow \sum_{i=1}^{n} a_i^2 \geq \frac{1}{n} \quad \text{as} \quad \sum_{i=1}^{n} a_i = 1.$$

N.T. with $\sum_{i=1}^{n} a_i = 1$, $\sum_{i=1}^{n} a_i^2$ attains its minimum

iff '=' holds in C-S inequality.

iff $a_i \propto 1 \;\forall\; i = 1(1)n$

iff $a_i = k \;\forall\; i = 1(1)n$

iff $a_i = \frac{1}{n} \;\forall\; i$ as $1 = \sum_{i=1}^{n} a_i = nk$

Hence, $T = \frac{1}{n}\sum_{i=1}^{n} X_i$ has the minimum variance among all linear UEs of μ.

$\Leftrightarrow T = \bar{x}$ is the BLUE of μ.

Ex. 1. Let $X_1, X_2, \ldots, X_n$ be n independent variables with common mean μ and variances $\sigma_i^2 = V(X_i)$, $i = 1(1)n$. Find the BLUE of μ.

Solution:- To find an estimator T such that it has the minimum variance in the class $\mathcal{C} = \left\{T : T = \sum_{i=1}^{n} a_i X_i, \sum_{i=1}^{n} a_i = 1\right\}$ of all UEs of μ.

Note that $Var(T) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$ where $\sum_{i=1}^{n} a_i = 1$.

By C-S inequality,

$$\left(\sum_{i=1}^{n} a_i \cdot \sigma_i \cdot \frac{1}{\sigma_i}\right)^n \le \left(\sum_{i=1}^{n} a_i^2 \sigma_i^2\right)\left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\right)$$

$$\Rightarrow \sum_{i=1}^{n} a_i^2 \sigma_i^2 \geq \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} \text{, as } \sum_{i=1}^{n} a_i = 1$$

Now, $\sum_{i=1}^{n} a_i = 1$, $\sum_{i=1}^{n} a_i^2 \sigma_i^2$ attains its minimum value

iff '=' holds in cauchy-schwartz inequality,

iff $a_i \sigma_i \propto \frac{1}{\sigma_i}$

iff $a_i = \frac{k}{\sigma_i^2} \quad \forall i$

iff $a_i = \left(\frac{1}{\sigma_i^2}\right) \Big/ \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\right)$ $\quad \left[\because 1 = \sum_{i=1}^{n} a_i = k \cdot \sum_{i=1}^{n} \frac{1}{\sigma_i^2} \Rightarrow k = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}\right]$

Hence, $T = \dfrac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \cdot X_i}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$ is the BLUE of μ,

Ex. 2. Let $X_1, X_2, \ldots, X_n$ be a r.s. from a popln. with mean μ and variance σ^2. Suggest two UEs based on all X_i's and compare their performances.

Solution:- Note that any weighted average of X_i's is an UE of μ based on all X_i's.

$$T = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i} \text{ is an UE of } \mu.$$

(i) $T_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$, (ii) $T_2 = \dfrac{\sum_{i=1}^{n} i X_i}{\frac{n(n+1)}{2}}$

Now, $V(T_1) = \frac{\sigma^2}{n}$, and $Var(T_2) = \frac{4}{\{n(n+1)\}^2} \cdot \sum_{i=1}^{n} i^2 \sigma^2$

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

$$= \frac{2(2n+1)}{3n(n+1)} \cdot \sigma^2 = \frac{\sigma^2}{n}\left(\frac{4n+2}{3n+3}\right) > \frac{\sigma^2}{n}$$

$\therefore Var(T_2) > Var(T_1)$

Hence, T_1 has smaller variance than T_2 and T_1 is better than T_2. Infact $T_1 = \overline{X}$ is the **BLUE** of μ.

Method of finding Estimators :-

(I) Method of Moments :- [The substitution Principle]
(Due to Karl Pearson)

One of the oldest and simplest method of estimation is the method of moments or the substitution principle. Let $f(x, \theta_1, \theta_2, \ldots, \theta_k)$ be the PDF or PMF of the given popln., whose moments μ_r', $r=1(1)k$, exists. Then, in general, μ_r' will be the function of $\theta_1, \theta_2, \ldots, \theta_k$. Let $X_1, X_2, \ldots, X_n$ be a r.s. from the given popln..

Define, $m_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r$ as the r^{th} order sample raw moment.

The method of moments consists in equating the k sample moments m_r', with the corresponding population moments μ_r' and solving k equations for k unknowns

$$\mu_r'(\theta_1, \theta_2, \ldots, \theta_k) = m_r', \quad r=1(1)k.$$

$$\Rightarrow \theta_i = h(m_1', m_2', \ldots, m_k'), \quad i=1(1)k.$$

Then, by method of moments,

$\hat{\theta}_i = h_i(m_1', \ldots, m_k')$ is the required estimator θ_i, $i=1(1)k$.

This method is quite reasonable if the sample is a good representation of the population.

Rational behind the Method of Moments :-

Note that X_i's are iid RVs.

$\Leftrightarrow X_i^r$'s are iid RVs.

Hence, by Khinchin's WLLN,

$$\frac{1}{n}\sum_{i=1}^{n} X_i^r \xrightarrow{P} E(X_1^r), \text{ provided } \mu_r' = E(X_1^r) \text{ exists.}$$

$$\Leftrightarrow m_r' \xrightarrow{P} \mu_r', \text{ provided } \mu_r' \text{ exists.}$$

Again, $E(m_r') = \mu_r'$

$\Rightarrow m_r'$ is an UE of μ_r'.

It can be shown that, under general conditions, m_r' are asymptotically normal. Based on the above facts, we can equate m_r' to μ_r', quite reasonable.

Remark:- Method of moments may lead to absurd estimators. If we are asked to compute estimators of θ in $N(\theta, \theta)$ or, $N(\theta, \theta^2)$ by the method of moments, then we can verify this assertion.

Example:- Let $X_1, X_2, \ldots, X_n$ be a r.s. from $P(\lambda)$.

Note that, $E(X_i) = \lambda = V(X_i)$

By method of moments,

$$\mu_1' = m_1' \;;\; \mu_2' = m_2'$$

$$\therefore \mu_2' - \mu_1' = m_2' - m_1'$$

$$\Leftrightarrow \lambda = \overline{X} \text{ and } \lambda = m_2 \text{ or } s^2$$

The method of moments leads to using either $\overline{X}$ or s^2, as an estimator of λ.

To avoid ambiguity, we take the estimator involving the lowest order sample moments.

Ex.1. Let X_i's be the r.s from Geometric (p) $\forall\, i = 1(1)n$. Find an MME of the parameter. Comment on the quality of estimator.

Solution:-

By Method of moments,

$$\mu_1' = \overline{X} \Rightarrow \frac{1}{p} = \overline{X}$$

An MME of p is $\hat{p} = \frac{1}{\overline{X}}$

Note that, $0 < \hat{p} = \frac{1}{\overline{X}} \leq 1$

$$\Rightarrow \hat{p} = \frac{1}{\overline{X}} \in \Omega = (0, 1)$$

and $E(\hat{p}) = E\left(\frac{1}{\overline{X}}\right) > \frac{1}{E(\overline{X})} = \frac{1}{1/p} = p.$

$\Rightarrow \hat{p}$ is the unbiased estimator.

Ex.2. Let X_i's $(i = 1(1)n)$ be a r.s. from $B(\alpha, \alpha)$ of 1st kind. Find an MME of α and comment on the quality of the estimator.

Ex.3. Find the estimator for λ by the method of moments in the exponential distribution [WBSU'11]

$$f(x, \lambda) = \frac{1}{\lambda} e^{-x/\lambda}, \; \lambda > 0, x > 0$$
$$= 0 \qquad , \text{ otherwise}$$

Solution:-

For exponential distribution,

$$\mu_1' = E(X) = \int_0^\infty x \cdot \frac{1}{\lambda} e^{-x/\lambda} dx$$
$$= \lambda$$

Now, the sample moment m_1' is given by

$$m_1' = \frac{1}{n} \sum X_i = \overline{X}$$

Equating μ_1' and m_1', we get

$$\hat{\lambda} = \overline{X}$$

(II) Method of Least Squares:- Let $y = f(x, \theta_1, \theta_2, \ldots, \theta_k)$ be the approximate regression equations of Y on X, which is assumed to be linear in parameters $\theta_1, \theta_2, \ldots, \theta_k$.

Let (x_i, y_i), $i = 1(1)n$, be an observed data on (X, Y). Define,
$e_i = y_i - f(x_i, \theta_1, \theta_2, \ldots, \theta_k)$ as the error in the prediction.

for a r.s. (X_i, Y_i), $i = 1(1)n$, we assume that

$$E_i = Y_i - f(x_i, \theta_1, \ldots, \theta_k)$$
$$\sim N(0, \sigma^2), \text{ where } \sigma^2 \text{ is constant.}$$

Then the likelihood of the observed errors $e_1, e_2, \ldots, e_n$ is $L(e_1, \ldots, e_n : \theta_1, \ldots, \theta_k) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} e_i^2}$

The observed sample $\{(x_i, y_i) : i = 1(1)n\}$ may be regarded as the most likely or most probable.

Hence the observed error $(e_1, e_2, \ldots, e_n)$ is also most probable.

Hence, we shall maximize the likelihood L w.r.t. $\theta_1, \theta_2, \ldots, \theta_k$

Now, maximizing L is equivalent to minimizing $\sum_{i=1}^{n} e_i^2$

$$= \sum_{i=1}^{n} \{y_i - f(x_i, \theta_1, \theta_2, \ldots, \theta_k)\}^2.$$

Hence, the principle of Least squares consist in minimizing the sum of squares of errors w.r.t. the parameters $\theta_1, \theta_2, \ldots, \theta_k$.

It can be shown that the least squares estimators are the solutions of $\frac{\partial}{\partial \theta_i} \sum_{i=1}^{n} \{Y_i - f(x_i, \theta_1, \ldots, \theta_k)\}^2 = 0 \;\forall\, i = 1(1)k.$

Ex.1. If $Y \sim N\left(\beta x_i, \frac{\sigma^2}{x_i}\right)$ when $X = x_i$, $i = 1(1)n$, find the LSE of β based on the r.s. (x_i, Y_i).

Solution:- Here $Y/X = x_i \sim N\left(\beta x_i, \frac{\sigma^2}{x_i}\right)$ ~~when $X = x_i$~~.

$$\Rightarrow E(Y/X = x_i) = \beta x_i \;\forall\, i = 1(1)n.$$

Note, $e_i = Y_i - \beta x_i \sim N\left(0, \frac{\sigma^2}{x_i}\right)$, when $X = x_i$.

$$\Rightarrow e_i\sqrt{x_i} \sim N(0, \sigma^2)$$

To maximize $L = \frac{1}{\left(2\pi \frac{\sigma^2}{x_i}\right)^{n/2}} \cdot e^{-\frac{1}{2}\frac{\Sigma e_i^2}{\sigma^2/x_i}}$

i.e. to minimize $\sum_{i=1}^{n} e_i^2 x_i$.

Normal equation is : $\frac{\partial}{\partial \beta}\left\{\sum_{i=1}^{n}(y_i-\beta x_i)^2 x_i\right\}=0$

$$\Rightarrow 2\sum_{i=1}^{n}(y_i-\beta x_i)(-x_i^2)=0$$

$$\Rightarrow \sum x_i^2 y_i = \beta \sum x_i^3$$

$$\Rightarrow \beta = \frac{\sum x_i^2 y_i}{\sum x_i^3}.$$

Ex.2 When $X = x_i$, then $E(Y_i)=\beta x_i$ and $Var(Y_i)=\sigma^2$, $\forall\ i=1(1)n$. Define $\hat{\beta}=\frac{\sum x_i Y_i}{\sum x_i}$, show that $\sum_{i=1}^{n}(Y_i-\beta x_i)^2 \geq \sum_{i=1}^{n}(Y_i-\hat{\beta}x_i)^2$.
Also, show that $E(\hat{\beta})=\beta$ and $Var(\hat{\beta})=\frac{\sigma^2}{\sum x_i^2}$. If each Y_i follows normal distribution, S.T. $\hat{\beta}$ is a normal variable.

Solution:- Here $E(Y/X=x_i)=\beta x_i$. Then $e_i = Y_i-\beta x_i\ \forall\ i=1(1)n$.
By method of least squares, to minimize

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(Y_i-\beta x_i)^2 \text{ w.r.t. } \beta,$$

Normal equation is : $\frac{\partial}{\partial\beta}\sum_{i=1}^{n}(Y_i-\beta x_i)^2 = 0$

$$\Rightarrow \beta = \frac{\sum x_i Y_i}{\sum x_i^2} = \hat{\beta}$$

Hence, $\sum(Y_i-\beta x_i)^2$ is minimum when $\beta=\hat{\beta}$.

$$\Rightarrow \sum(Y_i-\beta x_i)^2 \geq \sum(Y_i-\hat{\beta}x_i)^2$$

$$E(\hat{\beta}) = E\left(\frac{\sum x_i y_i}{\sum x_i^2}\right) = \frac{\sum x_i E(Y_i)}{\sum x_i^2} = \frac{\sum x_i \cdot \beta x_i}{\sum x_i^2} = \beta.$$

and $V(\hat{\beta}) = V\left(\frac{\sum x_i y_i}{\sum x_i^2}\right) = \frac{\sum x_i^2 V(Y_i)}{(\sum x_i^2)^2} = \frac{\sigma^2}{\sum x_i^2}$

Note that, $\hat{\beta} = \sum_{i=1}^{n}\left(\frac{x_i}{\sum x_i^2}\right)Y_i$ is a linear combination of normal variables Y_i, $i=1(1)n$.

Hence, $\hat{\beta}\sim N\left(E(\hat{\beta}), V(\hat{\beta})\right) \Rightarrow \hat{\beta}\sim N\left(\beta, \frac{\sigma^2}{\sum x_i^2}\right)$

[Q.E.D.]

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) $P_\theta\left[|T-\theta| < \epsilon\right]$, 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.

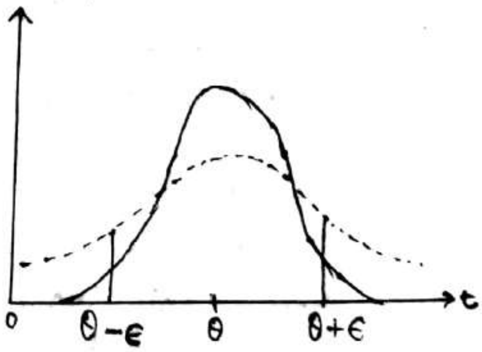
Definition: More concentrated and Most concentrated Estimators:

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.

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 $MSE_\theta(T_0) \le MSE_\theta(T)$, 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 unbiased estimator of a parametric function $\psi(\theta)$ iff $E\{T\} = \psi(\theta) \ \forall \ \theta \in \Omega$.

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

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

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

(3) A parametric function $\psi(\theta)$ is said to be estimable (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 than 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) Consistency :—

Here we shall consider a large sample property of estimators. Define, $T_n = T(X_1, X_2, \cdots, X_n)$, 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 $(x_1, x_2, \cdots, x_n)$ 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 $X_1, X_2, \cdots, X_n$ 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 $X_1, X_2, \cdots, X_n$ 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 $$Var\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1)$$

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

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

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)
$$T_2 = \frac{n+1}{n} T_1$$
$$= b_n T_1, \text{ where } b_n = \frac{n+1}{n} \longrightarrow 1 \text{ as } n \to \infty$$

Clearly, T_n is consistent for θ, since for every $\epsilon > 0$,
$$P\left[|T_2 - \theta| < \epsilon\right]$$
$$= P\left[\left|\frac{n+1}{n} T_1 - \theta\right| < \epsilon\right]$$
$$\simeq P\left[|T_1 - \theta| < \epsilon\right], \text{ for large } n.$$
$$\longrightarrow 1 \text{ as } n \to \infty.$$

(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 $n \to \infty$.

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

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

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

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

$= \mu$

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

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

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

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

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

Hence, T_1 is consistent for μ.

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

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

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:-

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

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

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

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

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

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

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

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

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

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

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

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

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:-

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

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

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

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

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

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

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:-

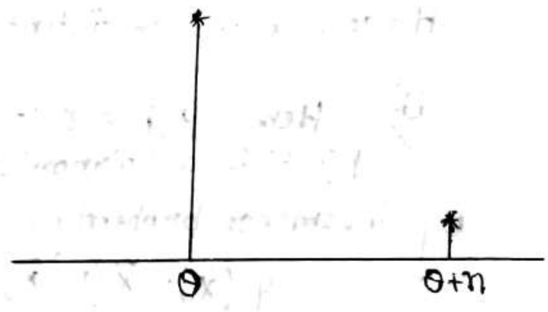
$$T_n = \begin{cases} \theta & \text{with probability } \left(1 - \frac{1}{n}\right) \\ \theta + n & \text{with probability } \frac{1}{n} \end{cases}$$

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

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

As $\{T_n\}$ is consistent for θ,

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

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

$\Rightarrow \{\psi(T_n)\}$ 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 $\rightarrow$ (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 $X_1, X_2, \dots, X_n$ 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 $X_1, X_2, \dots, X_n$ 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,

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

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

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

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

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 popln mean does not exist and μ is not the popln mean but it is the popln. 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 popln 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:-

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

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

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

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

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

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

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

$\therefore X_{(1)}$ 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) 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 $(X_1, X_2, \dots, X_n)$ be a random sample drawn from F_θ. A statistic $S = S(X_1, X_2, \dots, X_n)$ 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 $(X_1, X_2, \dots, X_n)$ 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 $(X_1, \dots, X_n)$ 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 $x_1, x_2, \dots, x_n$; 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 where as 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 $X_1, X_2, \dots, X_n$ 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.

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

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

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

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

Scanned by CamScanner

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

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:-

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

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

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

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

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 $f(x;\theta)$ 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.

Ex. (7). 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.

ANS:- 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

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

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.

Scanned by CamScanner

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

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

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

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

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

, 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.

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:-
$$\prod_{i=1}^{n} f(x_i;\theta) = g_1(t_1,\theta)\, h_1(\underset{\sim}{x})$$
$$= g_2(t_2,\theta)\, h_2(\underset{\sim}{x})$$
$$\Rightarrow \frac{g_1(t_1,\theta)}{g_2(t_2,\theta)} = \frac{h_2(\underset{\sim}{x})}{h_1(\underset{\sim}{x})}, \text{ which is independent of } \theta.$$
$$\Rightarrow \psi(t_1,t_2) = h^*(\underset{\sim}{x})$$
$\Rightarrow$ 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 or 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 or 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 $h(\underset{\sim}{x}) = \frac{1}{n!}$ 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 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 $T : \mathscr{X} \to \tau$, where $\tau = \{t : t = T(\underset{\sim}{x}), \underset{\sim}{x} \in \mathscr{X}\}$, then $T(\underset{\sim}{x})$ partitions the sample space into sets $A_t ; t \in \tau$ 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
$$\underset{\sim}{1}' \underset{\sim}{T_2} = \sum_{i=1}^{n} X_i$$

⋮

(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. :—
$$\prod_{i=1}^{n} f(x_i; p) = \left\{\theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum x_i}\right\} \times 1$$
$$= g\{T(\underset{\sim}{x}), \theta\} \cdot h(\underset{\sim}{x}), \text{ where } h(\underset{\sim}{x}) = 1$$
$$\text{and } T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$$

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 $\bar{x}$.

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

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

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

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

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

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

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

$\therefore$ 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 $X_1, X_2, \ldots, X_n$ be a r.s. from

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

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

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

Find a non-trivial sufficient statistic in each case.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 PMFs

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

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

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

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

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

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

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

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

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

$= T\left(\sum_{i=1}^{n}\ln x_i, \sum_{i=1}^{n}(\ln x_i)^2; \mu,\sigma\right) \cdot h(\underset{\sim}{x})$; 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 $a > b$
$= 0$ ow

$= g\left(\prod_{i=1}^{n} x_i, x_{(1)}; \theta, \alpha\right) \cdot h(x)$; 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) $f(\underset{\sim}{x}) = \dfrac{2^n}{\theta^{2n}} \prod_{i=1}^{n}(\theta - x_i)$; $0 < x_i < \theta$

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

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

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

<u>Ex. 11.</u> Let $X_1, \dots, X_n$ 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).

<u>Soln</u>:→ $f(\underset{\sim}{x}) = \left\{\dfrac{\alpha^p}{\Gamma p}\right\}^n \cdot \exp\left[-\alpha \sum_i x_i\right] \cdot \left(\prod_i x_i\right)^{p-1}$

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

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

<u>Soln</u>:→ $f(\underset{\sim}{x}) = \dfrac{1}{\theta^n} \cdot \exp\left\{-\dfrac{1}{\theta}\sum_{i=1}^{n} x_i\right\}$

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

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

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

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

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

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

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

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

then can you have a single sufficient statistic for θ?

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

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

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

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

$= 1 +$ 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 $g(T(\underset{\sim}{x}), \theta) \cdot h(\underset{\sim}{x})$ for a statistic other than the trivial choices

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

Hence there is no non-trivial sufficient statistic

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

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

Ex.(15). Let X_1 and X_2 be iid RVs having the discrete uniform distribution on $\{1,2,\ldots,N\}$, 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 RV 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:-

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

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

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

$\therefore \sum_{i=1}^{n} |x_i-\mu|$ 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 $X_1, \dots, X_n$ 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 $X_1, \dots, X_n$ be independently distributed RV's with PDFs

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

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

Solution:-

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

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

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

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

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

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

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

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

(ii) Hints:-

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

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

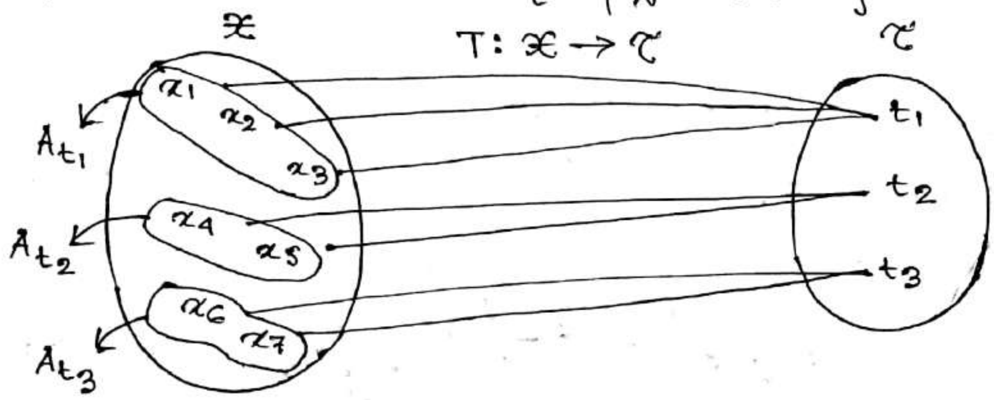
$$Y_i - 1 \sim U(-\theta, \theta).$$

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

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

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

Wherever '<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 in stead 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.

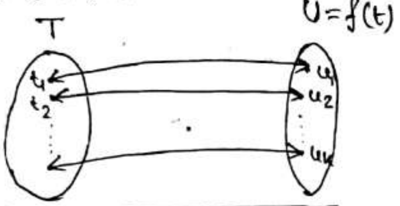
Definition:– A statistic T is called a minimal sufficient statistic for θ, provided

(i) T is sufficient for θ.

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

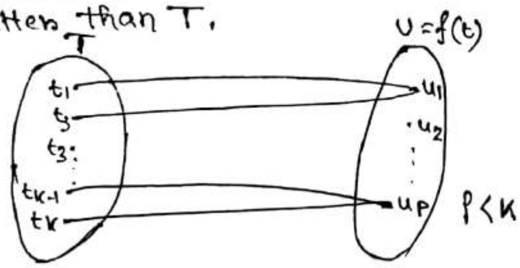
Remark:– If T and U are two sufficient statistics and $U = f(T)$. Which one is better?

⇒ If $f(\cdot)$ is one-to-one then T and $U = f(T)$ are eauivalent 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 $T(X)$ 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' ∋ $f(\underset{\sim}{x};\theta) = g'(T(\underset{\sim}{x});\theta) \cdot h'(\underset{\sim}{x})$. 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)

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

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

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) $T'(\underset{\sim}{X}) = (X_{(n)} - X_{(1)}, (X_{(n)}+X_{(1)})/2)$ 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.).

Scanned by CamScanner

(IV) $\boxed{\text{COMPLETENESS}}$:—

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.

Definition:— 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 distn.s $\{g(t;\theta) : \theta \in \Omega\}$ is complete.

Ex.(1). 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..

soln.:→ 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, $E(h(T)) = 0 \ \forall \ p \in (0,1)$

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

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

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

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. →

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

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

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

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

Differentiating w.r.t. θ, we get

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

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

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

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

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

[**Leibnitz Rule:-**

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

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

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

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

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

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

$\Rightarrow T = \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right)$ 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,

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

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

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

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

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

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

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

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

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

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

$\Rightarrow h_1(T)-h_2(T)=0$, 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.

Scanned by CamScanner

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 $T \sim NB(n, p)$.

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

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

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

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

$f(x;\theta) = \exp\left[u(\theta)\cdot T(x) + v(\theta) + w(x)\right]$, 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 distns.

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

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

$$= \exp\left[-\lambda + x\ln\lambda - \ln \underline{|x}\right]$$

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

where, $u(\lambda) = \ln\lambda$, $T(x) = x$, $v(\lambda) = -\lambda$, $w(x) = -\ln \underline{|x}$.

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

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

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

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

$\exp[u(\theta).T(x)+v(\theta)+w(x)]$ 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) $\{f(x;\theta):\theta\in\mathbb{R}\}$ where $f(x;\theta)=\begin{cases} e^{-(x-\theta)}, & x>\theta \\ 0, & \text{ow}\end{cases}$

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

$$\begin{aligned}\prod_{i=1}^{n} f(x_i;\theta) &= \exp\left[u(\theta).\sum_{i=1}^{n}T(x_i)+nv(\theta)+\sum_{i=1}^{n}w(x_i)\right]\\ &= \exp\left[u(\theta)\cdot\left(\sum_{i=1}^{n}T(x_i)\right)+nv(\theta)\right]\times\exp\left[\sum_{i=1}^{n}w(x_i)\right]\\ &= g\left(\sum_{i=1}^{n}T(x_i);\theta\right)\cdot h(\underset{\sim}{x})\end{aligned}$$

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,

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

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

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

$T(x) = \ln x$, $u(\theta) = \theta$, 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:

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

C_2:- 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.

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

Remark:-

⟨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.

Ex.(1):- 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. $(X_1, X_2, \ldots, X_n)$.

Solution:- Here $\underset{\sim}{\theta} = (\mu, \sigma)$, $\Omega = \{(\mu, \sigma) : \mu \in R, 0 < \sigma < \infty\}$
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.

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

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

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

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

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

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

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

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

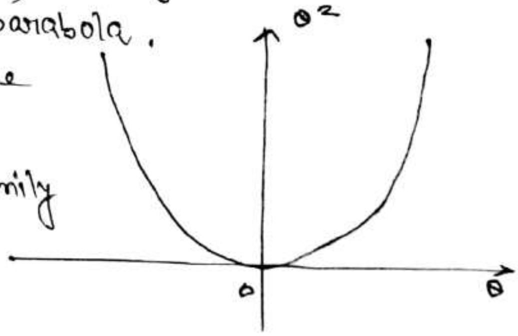
$$f(x;\theta) = \begin{cases} \frac{1}{\sqrt{2\pi}\,\theta} e^{-\frac{(x-\theta)^2}{2\theta^2}} & ; x \; i > \theta \\ 0 & ; \text{OW} \end{cases}$$

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

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

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

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

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

Show that they are not two-parameter exponential families.