Sufficient Statistic:

Definition → ①: Let, $X_1, X_2, \dots, X_n$ be a random sample of size 'n' drawn from a population having an unknown parameter θ. Let, $T = T(X_1, X_2, \dots, X_n)$ be a statistic that estimates the parameter θ. T will be said to be sufficient for θ if no other statistic can give any further information regarding θ.

Definition → ② T will be said to be sufficient for the parameter θ, if the conditional distribution of any other statistic given a value of T is independent of the parameter θ.

Definition → ③ T will be said to be sufficient for the parameter θ, if the conditional dist^n of $X_1, X_2, \dots, X_n$ given a value of T is independent of θ.

Let, $X_1, X_2, \dots, X_n$ be a sample from $N(\mu, 1)$, where, μ is unknown. Suppose, that we transform variables $X_1, X_2, \dots, X_n$ to $Y_1, Y_2, \dots, Y_n$ with the help of an orthogonal transformation so that Y_1 is $N(\sqrt{n}\mu, 1)$; $Y_2, \dots, Y_n$ are iid $N(0,1)$ & also $Y_1, Y_2, \dots, Y_n$ are independent, [Taking $y_1 = \sqrt{n}\,\bar{x}$ & $\forall\, k = 2(1)n$, $y_k = \{(k-1)x_k - (x_1 + x_2 + \dots + x_{k-1})\}/\sqrt{k(k-1)}$]. To estimate μ we can use either the observed values of $X_1, X_2, \dots, X_n$ or simply the observed value of $Y_1 = \sqrt{n}\bar{X}$. The random variables $Y_2, Y_3, \dots, Y_n$ provide no information about μ. Clearly, Y_1 is preferable since one need not keep a record of all the ~~obs~~ observations; it suffices to accumulate the observations & compute y_1. Any analysis of the data based on y_1 is just as effective as any analysis that could be based on x_i's.

Problem 2: $X_1 \sim P(\lambda)$, $X_2 \sim P(\lambda)$ & X_1 & X_2 are independent. Show that (X_1+2X_2) is not sufficient for λ.

Proof: Let, $T = X_1 + 2X_2$

$\therefore P(X_1=0, X_2=0 \mid T=0)=1$; $P(X_1=1, X_2=0 \mid T=1)=1$.

$$P(X_1=0, X_2=1 \mid T=2) = \frac{P(X_1=0, X_2=1)}{P(T=2)} = \frac{P(X=0)\,P(X_2=1)}{P(X_1=2, X_2=0)+P(X_1=0, X_2=1)}$$

$$= \frac{P(X_1=0)\,P(X_2=1)}{P(X_1=2)P(X_2=0)+P(X_1=0)\,P(X_2=1)}$$, since X_1 & X_2 are independent

$$= \frac{e^{-\lambda}.\frac{\lambda^0}{0!}.\,e^{-\lambda}.\frac{\lambda^1}{1!}}{e^{-\lambda}.\frac{\lambda^2}{2!}.e^{-\lambda}.\frac{\lambda^0}{0!}+e^{-\lambda}.\frac{\lambda^0}{0!}.\,e^{-\lambda}.\frac{\lambda^1}{1!}} = \frac{e^{-\lambda}.\,e^{-\lambda}\,\lambda}{e^{-\lambda}.e^{-\lambda}\,\lambda + e^{-\lambda}\frac{\lambda^2}{2}.e^{-\lambda}}$$

$$= \frac{\lambda}{\lambda+\frac{\lambda^2}{2}} = \frac{2\lambda}{2\lambda+\lambda^2} = \frac{2}{2+\lambda}.$$ & this expression is dependent on λ.

As such, $T = X_1+2X_2$ is not a sufficient statistic for 'λ'.

Neyman - Fisher Factorisation Theorem

Let, $X_1, X_2, \dots, X_n$ be a random sample of size 'n' drawn from a population having an unknown parameter θ. Let, $f_\theta(x_1, x_2, \dots, x_n)$ be the joint p.m.f. or p.d.f. of $X_1, X_2, \dots, X_n$. A statistic T is said to be sufficient for the parameter θ iff we can write

$$f_\theta(x_1, x_2, \dots, x_n) = g(\theta, t) \cdot h(x_1, x_2, \dots, x_n),$$

where, $g(\theta, t)$ = is a function of θ & t and $h(x_1, x_2, \dots, x_n)$ is independent of θ. Also, $t = T(x_1, x_2, \dots, x_n)$.

... sufficient statistic for an unknown θ

Problem ③: [illegible]

Theorem: ① Let, T be a sufficient statistic for an unknown θ & $\psi(T)$ be a 1-1 function of T. Show that $\psi(T)$ is also sufficient for θ.

Proof: Let, $T' = \psi(T)$.

$\Rightarrow T = \psi^{-1}(T')$ ∵ since, $\psi(T)$ is a 1-1 function, so inverse also exists

$\Rightarrow T = \phi(T')$, say.

Since, T is sufficient for θ, from Neyman-Fisher Factorisation theorem, we have

$$f_\theta(x_1, x_2, \ldots, x_n) = g(t, \theta)\, h(x_1, x_2, \ldots, x_n),$$ where, $h(x_1, x_2, \ldots, x_n)$ is independent of θ.

$$= g[\phi(t'), \theta]\, h(x_1, x_2, \ldots, x_n).$$

$$\Rightarrow f_\theta(x_1, x_2, \ldots, x_n) = k(t', \theta)\, h(x_1, x_2, \ldots, x_n).$$

As such, $T' = \psi(T)$ is sufficient for θ. [Proved].

Theorem: (2) If T be sufficient for θ & $T = \psi(T^*)$, then, T^* is sufficient for θ.

Proof: $\because$ T is sufficient for θ; $\therefore$ from Neyman-Fisher factorisation theorem we have, $f_\theta(x) = g(t,\theta)\,h(x) = g(\theta, \psi(t^*))\,h(x) = g^*(\theta, t^*)\,h(x)$

$\therefore$ T^* is sufficient for θ. [Proved]

[Vol. ②]

1.3) For a random ~~variable~~ sample X_i ($i=1(1)n$) from an exponential distribution with p.d.f.

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & \text{if } 0<x<\infty \\ 0, & \text{otherwise.} \end{cases}$$

where, $0<\theta<\infty$, show that $\sum_i X_i$ is a sufficient statistic for θ.

Ans/- By question, X_i follows exponential distribution with parameter θ & the p.d.f. of X_i is given by

$$f_\theta(x_i) = \begin{cases} \frac{1}{\theta} \cdot e^{-x_i/\theta}, & 0<x_i<\infty, \ \forall\, i=1(1)n \\ 0, & \text{otherwise} \end{cases}$$

Being a random sample, X_i's are mutually independent, $\forall\, i=1(1)n$.
Hence, the joint p.d.f. of $X_1, X_2, \ldots, X_n$ is given by

$$f_\theta(x_1, x_2, \ldots, x_n) = \begin{cases} \frac{1}{\theta^n} \cdot e^{-\sum_i x_i/\theta}, & 0<x_i<\infty \\ 0, & \text{o.w.} \end{cases}$$

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

$g(t,\theta) = \frac{1}{\theta^n} e^{-\sum_i x_i/\theta}$ which is dependent on θ & t

$h(x_1, x_2, \ldots, x_n) = 1$, which is independent of θ. Also $t = \sum_{i=1}^{n} x_i$

As such, the Neyman-Fisher factorisation theorem is satisfied for the statistic $T = \sum_{i=1}^{n} X_i$ & hence, it is sufficient for θ. [Proved].

1.7) Let, X_i, $\forall i=1(1)n$ be a random sample from a distribution with p.d.f.

$$f_\theta(x) = \begin{cases} \theta x^{\theta-1}, & \text{if } 0<x<1 \\ 0, & \text{o.w} \end{cases}, \quad \text{where, } 0<\theta<\infty.$$

Find a sufficient statistic for θ.

Ans/ By question, X_i is a continuous r.v. with p.d.f.

$$f_\theta(x_i) = \begin{cases} \theta x_i^{\theta-1}, & \text{if } 0<x<1 \\ 0, & \text{o.w.} \end{cases}, \quad \forall i=1(1)n.$$

Being a random sample, X_i's, $\forall i=1(1)n$, are mutually independent.
Hence, the joint p.d.f. of $X_1, X_2, \dots, X_n$ is given by,

$$f_\theta(x_1, x_2, \dots, x_n) = \begin{cases} \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta-1}, & 0<x_i<1. \\ 0, & \text{o.w} \end{cases}$$

$\Rightarrow f_\theta(x_1, x_2, \dots, x_n) = g(t, \theta)\, h(x_1, x_2, \dots, x_n)$ where

$g(t, \theta) = \theta^n \cdot \left(\prod_{i=1}^{n} x_i\right)^{\theta - 1}$ which is dependent on θ, & $t = \prod_{i=1}^{n} x_i$

& $h(x_1, x_2, \ldots, x_n) = 1$, which is independent of θ.

$\therefore$ The statistic $T = \prod_{i=1}^{n} x_i$ satisfies the Neyman-Fisher factorisation theorem

& hence it is sufficient for θ.

2) Let, $\underset{\sim}{X} = (X_1, X_2, \dots, X_n)$ be a sample from $N(\alpha\sigma, \sigma^2)$, where α is a known real number. Show that the statistic $T(\underset{\sim}{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ is sufficient for σ.

A) The joint p.d.f of $X_1, X_2, \dots, X_n$ is given by

$$f_\sigma(x_1, x_2, \dots, x_n) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \alpha\sigma)^2} \quad -\infty < x_i < \infty \quad \forall\, i = 1(1)n.$$

$$= \frac{1}{\sigma^n(\sqrt{2\pi})^n} e^{-\frac{\sum_i x_i^2}{2\sigma^2} + \frac{2\alpha\sigma}{2\sigma^2}\sum_{i=1}^{n} x_i - n\frac{\sigma^2\alpha^2}{2\sigma^2}}$$

$$= \frac{1}{\sigma^n(\sqrt{2\pi})^n} \cdot e^{-\frac{1}{2\sigma^2}\sum_i x_i^2 + \frac{\alpha}{\sigma}\sum_{i=1}^{n} x_i} \cdot e^{-n\alpha^2/2}.$$

$$\Rightarrow f_\sigma(x_1, x_2, \dots, x_n) = \cancel{g(t,\sigma)}\; g(t_1, t_2, \sigma)\, h(x_1, x_2, \dots, x_n),$$

where, $g(t_1, t_2, \sigma) = \frac{1}{\sigma^n} e^{-\frac{t_1}{2\sigma^2} + \frac{\alpha}{\sigma} t_2}$, where, $t_1 = \sum_{i=1}^{n} x_i^2$ & $t_2 = \sum_{i=1}^{n} x_i$ &

$$h(x_1, x_2, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-n\alpha^2/2}.$$

$\therefore$ $T_1 = \sum_{i=1}^{n} X_i^2$ & $T_2 = \sum_{i=1}^{n} X_i$ jointly ~~follow~~ satisfy the Neyman-Fisher factorisation theorem & hence, they are jointly sufficient for σ.

Problem: $f(x) = \frac{1}{2\theta} e^{-\frac{|x|}{\theta}}$, $-\infty < x < \infty$
$\theta > 0$.

If X is a single obs. from the above distn, then prove that $|X|$ is sufficient for θ.

Soln: Method-1: $f_\theta(x) = \frac{1}{2\theta} e^{-\frac{|x|}{\theta}}$, $-\infty < x < \infty$
$\theta > 0$

$\Rightarrow f_\theta(x) = g(t, \theta)\, h(x_1, x_2, \dots, x_n)$, where,

$g(t, \theta) = \frac{1}{2\theta} e^{-t/\theta}$, $t = |x|$, $h(x_1, x_2, \dots, x_n) = 1$. $\therefore$ The statistic $T = |X|$ satisfies the Neyman-Fisher factorisation theorem. Hence, it is sufficient for θ.