* Factorization Criterion
or Fisher – Neyman Criterion :

Let $x_1, x_2, ---, x_n$ be a r.s. from a population with p.d.f. $f(x;\theta)$, where $\theta \in \Omega$. A Statistic $T(x)$ is said to be sufficient for the parameter θ iff the joint density function of the random sample $x_1, x_2, ---, x_n$ can be expressed in the following form :-

$$f(x_1, x_2, \dots, x_n, \theta) = g(T(x), \theta) \cdot h(x_1, x_2, ---, x_n)$$

where the non-negative function $g(T(x), \theta)$ depends on only through the statistic $T(x)$ and the parameter θ and the non-negative function $h(x_1, x_2, ---, x_n)$ is independent of θ.

* Example :

Let $x_1, x_2, --, x_n$ be a r.s. from the p.d.f. $f(x;\theta) = \theta^x (1-\theta)^{1-x} \; I_{\{0,1\}}(x)$. Find the sufficient statistic for θ by using the Factorization Critierion

Solu :

$$f(x_1, x_2, --, x_n ; \theta) = \prod_{i=1}^{n} f(x_i ; \theta)$$

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

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

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

where $g(T(x) = \sum_{i=1}^{n} x_i , \theta) = \left(\frac{\theta}{1-\theta}\right)^{\sum_{i=1}^{n} x_i} (1-\theta)^n$

and $h(x_1, ..., x_n) = 1$

Hence $\sum_{i=1}^{n} x_i$ is a sufficient statistic for θ.

* Example:

Let $x_1, x_2, ..., x_n$ be a r.s. from $N(\mu, 1)$. Find the sufficient statistic for μ

Solu:

To find the sufficient statistic for μ, the joint density is given by:-

$$f(x_1, x_2, ..., x_n, \mu) = \prod_{i=1}^{n} f(x_i; \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-\mu)^2}$$

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

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

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

where $g(T(x), \mu) = e^{\mu\sum_{i=1}^{n} x_i - \frac{n}{2}\mu^2}$

and $h(x_1, x_2, ..., x_n) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2}$

then the joint density has been factored with $T(x) = \sum_{i=1}^{n} x_i$. Hence $\sum_{i=1}^{n} x_i$ is a sufficient statistic for the parameter μ.

* **Example:**
Find a sufficient statistic for mean of $N(\theta, \sigma^2)$, (σ^2 = known)

Solu:

$$f(x_1, x_2, \dots, x_n, t(x)) = \prod_{i=1}^{n} f(x_i; \theta, \sigma^2)$$

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

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}x_i^2 - 2\sum x_i\theta + n\theta^2\right]}$$

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

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

$$= g\left(\sum_{i=1}^{n}x_i, \theta\right) \cdot h(x_1, x_2, \dots, x_n)$$

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

* **Example:**
Show that $\bar{X}$ is a sufficient statistic for mean of $N(\theta, \sigma^2)$, (where σ^2 = known)

Solu:

$$F(x_1, x_2, \dots, x_n, \theta) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum(x_i-\theta)^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\left[\sum(x_i-\bar{x})^2 + n(\bar{x}-\theta)^2\right]}$$

$$= e^{-\frac{n(\bar{x}-\theta)^2}{2\sigma^2}} \cdot \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum(x_i-\bar{x})^2}$$

$$= g(\bar{x}, \theta) \cdot h(x_1, x_2, \dots, x_n)$$

$\therefore \bar{X}$ is a sufficient statistic for θ.

* **Example:**

Let $x_1, x_2, \ldots, x_n$ be a r.s. from $N(\theta, \sigma^2)$ (θ = known) show that $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\theta)^2$ is a sufficient statistic for σ^2, using Factorization criterion.

Soln:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\theta)^2$$

$$f(x_1, x_2, \ldots, x_n; \theta) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2}$$

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

$$= \left(\frac{1}{\sigma}\right)^n e^{-\frac{n\hat{\sigma}^2}{2\sigma^2}} \left(\frac{1}{\sqrt{2\pi}}\right)^n$$

$$= g(t(x), \sigma^2) \cdot h(x_1, \ldots, x_n)$$

$$= g(\hat{\sigma}^2, \sigma^2) \cdot h(x_1, x_2, \ldots, x_n)$$

where $g(\hat{\sigma}^2, \sigma^2) = \left(\frac{1}{\sigma}\right)^n e^{-\frac{n\hat{\sigma}^2}{2\sigma^2}}$

$$h(x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n$$

$\therefore$ $\hat{\sigma}^2$ is a sufficient statistic for σ^2.

* **Example:**

Let $x_1, x_2, \ldots, x_n$ be a r.s. from a population with p.d.f. $f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$

show that $\sum_{i=1}^{n} x_i$ and $\bar{X}$ are sufficient statistics for θ.

Solu:

$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$

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

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

$$= g\left(\sum_{i=1}^{n} x_i; \theta\right) \cdot h(x_1, x_2, \dots, x_n)$$

$\therefore \sum x_i$ is a sufficient statistic for θ.

$$f(x_1, x_2, \dots, x_n; \theta) = \left(\frac{1}{\theta}\right)^n e^{-\frac{n\bar{x}}{\theta}} \cdot 1$$

$$= g(\bar{x}, \theta) \cdot h(x_1, x_2, \dots, x_n)$$

$\Longrightarrow \bar{x}$ is a sufficient statistic for θ.

* Example:

Let $x_1, x_2, \dots, x_n$ be a r.s. from $N(\theta, \sigma^2)$ show that $\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ and $\bar{x}$ are set of two sufficient statistics estimates for σ^2 and θ.

Solu:

In general, if

$$f(x_1, x_2, \dots, x_n, \theta_1, \theta_2, \dots, \theta_k) = g(\hat{\theta}_1, \dots, \hat{\theta}_k, \theta_1, \dots, \theta_k)\, h(x_1, x_2, \dots, x_n)$$

Then we say that $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ are jointly sufficient statistic for $\theta_1, \theta_2, \dots, \theta_k$

$$f(x_1, x_2, \cdots, x_n; \theta, \sigma^2) = \prod_{i=1}^{n} f(x_i; \theta, \sigma^2)$$

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

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

$$= \underbrace{\left(\frac{1}{\sigma}\right)^n e^{-\frac{1}{2\sigma^2}(n-1)\hat{\sigma}^2} \cdot e^{-\frac{1}{2\sigma^2}n(\bar{x}-\theta)^2}}_{g(\hat{\theta}=\bar{x},\, \hat{\sigma}^2;\, \theta, \sigma^2)} \cdot \underbrace{\left(\frac{1}{\sqrt{2\pi}}\right)^n}_{h(x_1, \cdots, x_n)}$$

$\therefore$ $\bar{X}$ and $\hat{\sigma}^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$ are set of sufficient statistics for θ and σ^2 respectively.