Example 4:—

$X_1, \dots, X_n \sim N(\theta, \theta^2)$.

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

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

$$= \underbrace{e^{-\frac{1}{2\theta^2}\left(\sum_{i=1}^{n} x_i^2 - 2\theta\sum_{i=1}^{n} x_i\right)} \times \frac{1}{\theta^n} \times e^{-\frac{n\theta^2}{2\theta^2}}}_{g(\theta;\, T_n(x_1, \dots, x_n))} \times \underbrace{\frac{1}{(\sqrt{2\pi})^n}}_{h(x_1, \dots, x_n)}.$$

Hence, by NFT, $T_n = \left(\sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} X_i\right)$ is jointly sufficient statistic for θ.

Rao - Blackwellization Theorem:-

Statement:- Let $U_n(X_1, .., X_n)$ be an unbiased estimator of $g(\theta)$. And, suppose that $T_n(X_1, .., X_n)$ is a sufficient statistic for θ.

Define $\eta(T_n) = E_{U|T_n}[U_n(X_1, .., X_n) \mid T_n(X_1, .., X_n)]$, which is indep of θ since $T_n(X_1, .., X_n)$ is sufficient statistic for θ.

Then $E_{T_n}[\eta(T_n)] = g(\theta)$ and

$$Var(\eta(T_n)) \leq Var(U_n(X_1, .., X_n)).$$

Justification of the unbiasedness of $\eta(T_n)$:-

$(X, Y) \sim f_{X,Y}$

Result :- $EY = E_X \left[E_{Y|X}(Y) \right]$

$$EY = \int y \, f(x, y) \, dx \, dy$$

conditional density, Marginal density of X

$$= \int y \left[f_{Y|X}(y) \, g_X(x) \right] dx \, dy$$

$$= \int \left[\int y \, f_{Y|X}(y) \, dy \right] g_X(x) \, dx$$

$$= \int \left[E_{Y|X}(Y) \right] g_X(x) \, dx$$

$$= E_X \left[E_{Y|X}(Y) \right].$$

Note that

$$\eta(T_n) = E_{U_n|T_n}\left[U_n(X_1, \ldots, X_n) \mid T_n(X_1, \ldots, X_n)\right]$$

$$E_{T_n}\left[\eta(T_n)\right] = E_{T_n}\left[E_{U_n|T_n} U_n(X_1, \ldots, X_n) \mid T_n(X_1, \ldots, X_n)\right]$$

$$= E_{U_n}\left[U_n(X_1, \ldots, X_n)\right]$$

$$= g(\theta).$$

Hence, $\eta(T_n)$ is an unbiased estimator of $g(\theta)$.

$P[X=0] = 1-\theta$

Example: Let $X_1, \ldots, X_n \sim \text{Ber}(1, \theta)$, $\theta \in (0,1)$. $P[X=1] = \theta$

$\mapsto P[X = x] = \theta^{x}(1-\theta)^{1-x}$,

$x = 0$ or 1.

Note that $U_n^{(X_1, \ldots, X_n)} = X_1$ is an unbiased estimator of θ.

Already, we know $T_n(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$ is sufficient statistic for θ.

Observation: $E[U_n] = \theta$ & $Var[U_n] = \theta(1-\theta)$.

$$\eta(T_n) = E[U_n \mid T_n]$$

$$= E\left[X_1 \mid \sum_{i=1}^{n} X_i = t\right]$$

$$= 0 \times P\left[X_1 = 0 \mid \sum_{i=1}^{n} X_i = t\right] \times 1 \times P\left[X_1 = 1 \mid \sum_{i=1}^{n} X_i = t\right]$$

$$= P\left[X_1 = 1 \mid \sum_{i=1}^{n} X_i = t\right]$$

$$= \frac{P\left[X_1 = 1 \cap \sum_{i=1}^{n} X_i = t\right]}{P\left[\sum_{i=1}^{n} X_i = t\right]}$$

$X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Ber}(1, \theta)$

$X_1 \perp\!\!\!\perp (X_2, \ldots, X_n)$.

$$= \frac{P\left[X_1 = 1 \cap \sum_{i=2}^{n} X_i = t-1\right]}{P\left[\sum_{i=1}^{n} X_i = t\right]}$$

because.

$$= \frac{P[X_1 = 1] \times P\left[\underbrace{\sum_{i=2}^{n} X_i}_{\text{Bin}(n-1,\theta)} = t-1\right]}{P\left[\underbrace{\sum_{i=1}^{n} X_i}_{\text{Bin}(n,\theta)} = t\right]}$$

$$= \frac{\theta \times \binom{n-1}{t-1} \theta^{t-1} (1-\theta)^{(n-1)-(t-1)}}{\binom{n}{t} \theta^{t} (1-\theta)^{n-t}} \overset{\text{check}}{=} \frac{\binom{n-1}{t-1}}{\binom{n}{t}} \overset{\text{check}}{=} \frac{t}{n}.$$

Hence, the Rao-Blackwellised improved estimator is $= \frac{\sum_i X_i}{n}$.

Observation :-

$$\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{Var}(X_i) + \frac{1}{n^2}\sum_{i\neq j} \operatorname{Cov}(X_i, X_j).$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \theta(1-\theta) + \frac{1}{n^2}\sum_{i\neq j} 0$$

$$= \frac{1}{n^2}\times n\, \theta(1-\theta) = \frac{\theta(1-\theta)}{n}.$$

Hence, for any $n \geq 1$, $\operatorname{Var}(\eta(T_n)) = \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$

$$\leq \operatorname{Var}(X_1) = \operatorname{Var}[U_n].$$

Remark :- However, Rao-Blackwellization does NOT give us uniformly minimum variance estimator. It helps us to lower the variance only.

Another concept:—

Complete statistic:—

Def^{n}: A statistic $T_n(X_1, \ldots, X_n)$ is said to be complete if for any function g, we have

$$E\, g(T_n) = 0 \quad \forall \ \theta \in \Theta.$$

$$\Rightarrow g(T_n) = 0 \text{ almost everywhere}$$

$$\Downarrow$$

$$P[g(T_n) \neq 0] = 0$$

If a complete statistic is sufficient, then it is called complete sufficient statistic.

A few examples:-

Example:- $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Ber}(1, \theta)$, $\theta \in (0,1)$.

Already we known $T_n = \sum_{i=1}^{n} X_i$ is suffient for θ.

Want to check whether T_n is complete suffient stat (CSS) or NOT.

Take any arbitrary $f^{n} =$ g.

Suppose $E\, g(T_n) = 0 \quad \forall \quad \theta \in \Theta = (0,1)$.

Further, note that $T_n \sim \text{Bin}(n, \theta)$.

$$E\, g(T_n) = 0 \quad \forall\ \theta \in \Theta$$

$$\Leftrightarrow \sum_{t=1}^{n} g(t) \binom{n}{t} \theta^{t} (1-\theta)^{n-t} = 0 \quad \forall\ \theta \in \Theta$$

$$\Leftrightarrow \underbrace{(1-\theta)^{n}}_{>0} \sum_{t=1}^{n} g(t) \binom{n}{t} \left(\frac{\theta}{1-\theta}\right)^{t} = 0 \quad \forall\ \theta \in \Theta$$

$$\Leftrightarrow \sum_{t=1}^{n} g(t) \binom{n}{t} \left(\frac{\theta}{1-\theta}\right)^{t} = 0 \quad \forall\ \theta \in \Theta$$

$$\Leftrightarrow \sum_{t=1}^{n} g(t) \binom{n}{t} \delta^{t} = 0, \text{ when } \delta \in (0, \infty).$$

$$\delta = \frac{\theta}{1-\theta}.$$

Hence, $g(t)\binom{n}{t} = 0$

$\Rightarrow g(t) = 0$ since $\binom{n}{t} > 0$.

Hence, $T_n = \sum_{i=1}^{n} x_i$ is CSS.

Example:- Not all sufficient statistic is complete.

$X_1, \ldots, X_n \sim N(\theta, \theta^2)$.

$$T_n = \Big(\underbrace{\sum_{i=1}^{n} X_i}_{T_{1,n}},\ \underbrace{\sum_{i=1}^{n} X_i^2}_{T_{2,n}}\Big)$$ is "jointly" sufficient for θ.

Note that $E(T_{1,n}^2) = \theta^2 (n-1) n \Rightarrow E\left(\frac{T_{1,n}^2}{n(n-1)}\right) = \theta^2$. — ①

& $E(T_{2,n}) = 2n\theta^2 \Rightarrow E\left(\frac{T_{2,n}}{2n}\right) = \theta^2$ — ②

① − ② $\Rightarrow$

$$E\left(\frac{T_{1,n}^2}{n(n-1)} - \frac{T_{2,n}}{2n}\right) = \theta^2 - \theta^2 = 0$$

However, $\frac{T_{1,n}^2}{n(n-1)} - \frac{T_{2,n}}{2n} \neq 0$ a.e. Hence $T_n = (T_{1,n}, T_{2,n})$ is NOT CSS