Joint Sufficient Statistics and Minimal sufficient

Jointly sufficient statistics

Consider

$$\left.\begin{array}{l} T_1 = T_1(X_1, \cdots, X_n) \\ T_2 = T_2(X_1, \cdots, X_n) \\ \cdots \\ T_k = T_k(X_1, \cdots, X_n) \end{array}\right\} \text{ - functions of the sample } (X_1, \cdots, X_n).$$

Very similarly to the case when we have only one function T, a vector $(T_1, \cdots, T_k)$ is called *jointly sufficient statistics* if the distribution of the sample given T's

$$\mathbb{P}_\theta(X_1, \ldots, X_n | T_1, \ldots, T_k)$$

does not depend on θ. The Neyman-Fisher factorization criterion says in this case that $(T_1, \ldots, T_k)$ is jointly sufficient if and only if

$$f(x_1, \ldots, x_n | \theta) = u(x_1, \ldots, x_n) v(T_1, \ldots, T_k, \theta).$$

Example 1. Let us consider a family of normal distributions $N(\alpha, \sigma^2)$, only now both α and σ^2 are unknown. Since the joint p.d.f.

$$f(x_1, \dots, x_n | \alpha, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{ -\frac{\sum x_i^2}{2\sigma^2} + \frac{\sum x_i \alpha}{\sigma^2} - \frac{n\alpha^2}{2\sigma^2} \right\}$$

is a function of

$$T_1 = \sum_{i=1}^n X_i \text{ and } T_2 = \sum_{i=1}^n X_i^2,$$

by Neyman-Fisher criterion (T_1, T_2) is jointly sufficient.

Example 2. Let us consider a uniform distribution $U[a,b]$ on the interval $[a,b]$ where both end points are unknown. The p.d.f. is

$$f(x|a,b) = \begin{cases} \frac{1}{b-a}, & x \in [a,b], \\ 0, & \text{otherwise.} \end{cases}$$

The joint p.d.f. is

$$\begin{aligned} f(x_1, \cdots, x_n | a, b) &= \frac{1}{(b-a)^n} I(x_1 \in [a,b]) \times \ldots \times I(x_n \in [a,b]) \\ &= \frac{1}{(b-a)^n} I(x_{\min} \in [a,b]) \times I(x_{\max} \in [a,b]). \end{aligned}$$

The indicator functions make the product equal to 0 if at least one of the points falls out of the range $[a,b]$ or, equivalently, if either the minimum $x_{\min} = \min(x_1, \ldots, x_n)$ or maximum $x_{\max} = \max(x_1, \ldots, x_n)$ falls out of $[a,b]$. Clearly, if we take

$$T_1 = \max(X_1, \ldots, X_n) \text{ and } T_2 = \min(X_1, \ldots, X_n)$$

then (T_1, T_2) is jointly sufficient by Neyman-Fisher factorization criterion.

Importance of Sufficient Statistics

- Gives a way of compressing information about underlying parameter θ.
- Gives a way of improving estimator using sufficient statistic (application in Rao-Blackwell Theorem)

Rao-Blackwellization & MVUE.

Where, CR inequality is applicable under some regularity conditions, ie under a no. of stringent (কঠোর) conditions, Rao-Blackwellization is applicable in a much more relaxed situation. Moreover, here we can have MVUE (if exists) directly from the theorem.

Rao-Blackwellization Theorem:

Suppose, $U = U(X_1, X_2, \dots, X_n)$ is an unbiased estimator of an estimable parametric function $\gamma(\theta)$, $\forall\ \theta \in \Theta$.

Suppose, $T = T(x_1, x_2, \dots, x_n)$ is a sufficient statistic for $\theta \in \Theta$. Then, the estimator $\phi(t) = E(U|T)$ is also an unbiased estimator of $\gamma(\theta)$, $\forall\ \theta \in \Theta$ & $V[\phi(T)] \leq V_\theta(U)$, $\forall\ \theta \in \Theta$.

③ Let, $X_1, X_2, \cdots, X_n \overset{iid}{\sim} P(\lambda)$. Then find an MVUE of $(1-e^{-\lambda})$.

Solution: Since, by question, $X_1, X_2, \cdots, X_n \overset{iid}{\sim} P(\lambda)$, so the common p.d.f. of X

$$f_\lambda(x) = \begin{cases} \dfrac{e^{-\lambda} \cdot \lambda^x}{x!}, & x = 0, 1, 2, \cdots; \ \lambda > 0. \\ 0, & \text{o.w.} \end{cases}$$

$\therefore P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda} = \gamma(\lambda).$

Now, let us define $U = \begin{cases} 1, & \text{if } x_1 \geq 1 \\ 0, & \text{o.w.} \end{cases}$

$\therefore$ $E(U) = 1 \times P(x_1 \geq 1) + 0 \times P(x_1 \ngeq 1) = \gamma(\lambda)$, $\forall\ \lambda > 0$.

$\therefore$ U is an unbiased estimator of $\gamma(\lambda)$.

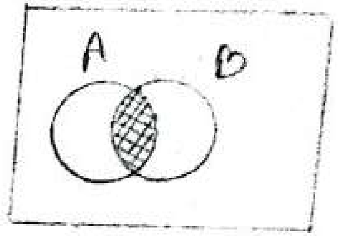
Now, we know that $T = \sum_{i=1}^{n} x_i$ is a sufficient statistic for λ.

$\therefore$ According to the Rao-Blackwellization theorem

$\phi(T) = E(U|T)$ is the MVUE of $\gamma(\lambda)$.

Method → ① Now, $\phi(t) = E[U \mid T=t] = \dfrac{E[U, \sum_{i=1}^{n} X_i = t]}{P[\sum_{i=1}^{n} X_i = t]}$

$$= \frac{1 \times P(X_1 \geq 1, \sum_{i=1}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)} \quad \cancel{= 1 - P(X_1 = 0, \sum_{i=1}^{n} X_i =}$$



$$= \frac{P[\sum_{i=1}^{n} X_i = t] - P(X_1 = 0, \sum_{i=1}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)}, \quad \because P(A^c \cap B) = P(B) - P(A \cap B)$$

$$= 1 - \frac{P(X_1 = 0, \sum_{i=2}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)} = 1 - \frac{P(X_1 = 0) P(\sum_{i=2}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)}, \quad \because X_i\text{'s are iid}, \forall\, i = 1(1)n$$

$$= 1 - \frac{e^{-\lambda} \cdot e^{-(n-1)\lambda} \cdot \{(n-1)\lambda\}^t / t!}{e^{-n\lambda} \cdot (n\lambda)^t / t!} = 1 - \frac{(n-1)^t}{n^t}, \quad t = 1, 2, \ldots$$

$\therefore \phi(T) = 1 - \dfrac{(n-1)^T}{n^T}$ is the MVUE of $\gamma(\lambda) = 1 - e^{-\lambda}$.

—x—

Minimal Sufficient Statistic

- We are interested in finding a statistic that achieves the most data reduction while still retaining all the information about parameter θ.

Definition

- A sufficient statistic $T(X)$ is called a *minimal sufficient statistic* if, for any other sufficient statistic $T'(X)$, $T(x)$ is a function of $T'(x)$.

Sufficient statistic can be thought as partition of sample space $\mathcal{X}$. Let $\mathcal{T} = \{t : t = T(\mathbf{x}), \mathbf{x} \in \mathcal{X}\}$, then $T(\mathbf{x})$ partitions the sample space into sets A_t, $t \in \mathcal{T}$, defined by $A_t := \{\mathbf{x} \in \mathcal{X} : T(\mathbf{x}) = t\}$. If $\{B_{t'} : t' \in \mathcal{T}'\}$ are the partition sets for $T'(\mathbf{x})$ and $\{A_t : t \in \mathcal{T}\}$ are the partition sets for $T(\mathbf{x})$, then Definition 5.1 states that every $B_{t'}$ is a subset of A_t. Thus, the partition associated with a minimal sufficient statistic is the coarsest possible partition for a sufficient statistic.

Theorem 5.3 *Let $f(\mathbf{x}|\theta)$ be the pmf or pdf of a sample $\mathbf{X}$. Suppose there exists a function $T(\mathbf{x})$ such that for every two sample points $\mathbf{x}$ and $\mathbf{y}$, the ratio $f(\mathbf{x}|\theta)/f(\mathbf{y}|\theta)$ is a constant as a function of θ if and only if $T(\mathbf{x}) = T(\mathbf{y})$. Then, $T(\mathbf{X})$ is a minimal sufficient statistic for θ.*

Example 5.4 (Normal Minimal Sufficient Statisitc) Assume $X_1, \cdots, X_n$ are i.i.d. $N(\mu, \sigma^2$ with both parameters unknown. Let $\mathbf{x}$ and $\mathbf{y}$ be two sample points, and let $(\bar{x}, s_{\mathbf{x}}^2)$ and $(\bar{y}, s_{\mathbf{y}}^2$ be the sample means and variances corresponding to $\mathbf{x}$ and $\mathbf{y}$ samples, respectively. Then the ratio of densities is

$$\begin{aligned}\frac{f(\mathbf{x}|\mu,\sigma^2)}{f(\mathbf{y}|\mu,\sigma^2)} &= \frac{(2\pi\sigma^2)^{-n/2} exp(-[n(\bar{x}-\mu)^2 + (n-1)s_{\mathbf{x}}^2]/(2\sigma^2))}{(2\pi\sigma^2)^{-n/2} exp(-[n(\bar{y}-\mu)^2 + (n-1)s_{\mathbf{y}}^2]/(2\sigma^2))} \\ &= exp([-n(\bar{x}^2 - \bar{y}^2) + 2n\mu(\bar{x}-\bar{y}) - (n-1)(s_{\mathbf{x}}^2 - s_{\mathbf{y}}^2)]/(2\sigma^2))\end{aligned}$$

This ratio will be a constant as a function of μ and σ^2 if and only if $\bar{x} = \bar{y}$ and $s_{\mathbf{x}}^2 = s_{\mathbf{y}}^2$. Thus by Theorem 5.3, $(\bar{X}, S^2)$ is a minimal sufficient statistic for (μ, σ^2).