

Definition A statistic T or its distribution $\{P_\theta, \theta \in \Theta\}$ is said to be (boundedly) complete if for any real valued (bounded) function $h(T)$ with $E(|h(T)|) < \infty$,

$$E_\theta h(T) = 0 \forall \theta \text{ implies } h(T) = 0$$

(with probability one under all θ).

Suppose T is discrete. The condition then simply means the family of p.m.f.'s $f_T(t|\theta)$ of T is rich enough that there is no non-zero $h(t)$ that is orthogonal to $f^T(t|\theta)$ for all θ in the sense $\sum_t h(t)f_T(t|\theta) = 0$ for all θ . In general, T is complete iff $h(T) \equiv 0$ is the only unbiased estimator of 0.

Complete implies boundedly complete.

Example. Let X_1, \dots, X_n be i.i.d Bernoulli(p), $0 < p < 1$. $E(X_1 - X_2) = 0$ for all p , so \mathbf{X} is not complete (but sufficient). $T = \sum_{i=1}^n X_i$ is minimal sufficient for p , and $T \sim \text{Binomial}(n, p)$.

Claim: T is complete.

Suppose $E_p h(T) = 0$ for all $p \in (0, 1)$. i.e.,

$$\begin{aligned} \sum_{t=0}^n h(t) \binom{n}{t} p^t (1-p)^{n-t} &= 0 \forall p \in (0, 1), \text{ or} \\ h(0)(1-p)^n + \sum_{t=1}^n h(t) \binom{n}{t} p^t (1-p)^{n-t} &= 0 \forall p \in (0, 1). \end{aligned}$$

As $p \rightarrow 0$, LHS $\rightarrow h(0)$ and RHS = 0. Therefore, $h(0) = 0$. Hence,

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

Thus we get $h(1) = 0 = h(2)$ and finally $h(n) = 0$.

Example. Let X_1, \dots, X_n be i.i.d $N(\theta, \theta^2)$, $\theta > 0$. Then

$$\begin{aligned} f(x_1, \dots, x_n | \theta) &= (2\pi)^{-n/2} \theta^{-n} \exp \left(-\frac{1}{2\theta^2} \sum_{i=1}^n (x_i - \theta)^2 \right) \\ &= (2\pi)^{-n/2} \theta^{-n} \exp \left(-\frac{1}{2\theta^2} \left[\sum_{i=1}^n x_i^2 - 2\theta \sum_{i=1}^n x_i + n\theta^2 \right] \right) \\ &= (2\pi)^{-n/2} \theta^{-n} \exp \left(\frac{1}{\theta} \sum_{i=1}^n x_i - \frac{1}{2\theta^2} \sum_{i=1}^n x_i^2 - \frac{n}{2} \right). \end{aligned}$$

Thus, we see that, $T(\mathbf{X}) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ is (minimal) sufficient for θ (even though θ is now one-dimensional).

Claim: $T = (T_1, T_2)$ is not complete.

Note that $h(t_1, t_2) = t_2 - \frac{2}{n+1}t_1^2$ is not the 0 function, but $E_\theta [T_2 - T_1^2] = 0$ for all θ . To see this, observe, $T_1 = \sum_{i=1}^n X_i \sim N(n\theta, n\theta^2)$, so

$$E_\theta(T_1^2) = E_\theta \left[\sum_{i=1}^n X_i \right]^2 = n\theta^2 + (n\theta)^2 = n(n+1)\theta^2.$$

Also,

$$E_\theta(T_2) = E_\theta \left[\sum_{i=1}^n X_i^2 \right] = \sum_{i=1}^n E_\theta(X_i^2) = n(\theta^2 + \theta^2) = 2n\theta^2,$$

and hence

$$E_\theta \left[\left(\frac{2}{n+1} \right) T_1^2 \right] = \frac{2n(n+1)}{n+1} \theta^2 = 2n\theta^2 = E_\theta(T_2).$$

Theorem. Let \mathbf{X} have distribution $P_\theta, \theta \in \Theta$ and let $T = T(\mathbf{X})$ be complete sufficient for θ (or $P_\theta, \theta \in \Theta$). Then every function $h(T)$ is the unique unbiased estimate of its own expected value. i.e., for any h , if $q(\theta) = E_\theta h(T)$, then $h(T)$ is the only unbiased estimate available for $q(\theta)$.

Proof. Suppose $h_1(T)$ and $h_2(T)$ are both unbiased estimates of a parametric function $\psi(\theta)$. Then

$$E_\theta [h_1(T) - h_2(T)] = 0 \quad \forall \theta \in \Theta.$$

i.e., if we let $h^*(T) = h_1(T) - h_2(T)$, then

$$E_\theta [h^*(T)] = 0 \quad \forall \theta \in \Theta.$$

Since T is complete, $h^*(T) \equiv 0$. i.e. $h_1(T) = h_2(T)$.

Theorem (Lehmann-Scheffe). Suppose $T(\mathbf{X})$ is complete sufficient (for $P_\theta, \theta \in \Theta$) and $S(\mathbf{X})$ is any unbiased estimate of $q(\theta)$. Then $T^*(\mathbf{X}) = E(S(\mathbf{X})|T)$ is the unique UMVUE of $q(\theta)$ if $Var_\theta(T^*(\mathbf{X})) < \infty$ for all θ .

Proof. Both S and T^* are unbiased, so that their MSE is the respective variance. By the Rao-Blackwell theorem, $Var_\theta(T^*) < Var_\theta(S)$ unless $S = T^*$. To show uniqueness, we show that T^* is the same, whichever S we start with, so that T^* cannot be improved upon further.

Let S_1 and S_2 be two unbiased estimators of $q(\theta)$, and let $g_1(T) = E(S_1(X)|T)$ and $g_2(T) = E(S_2(X)|T)$. But then,

$$E(g_1(T)) = E(S_1) = q(\theta) = E(S_2) = E(g_2(T)), \forall \theta \in \Theta.$$

Since T is complete, there can be only one unbiased estimate of $q(\theta)$ based on T . Therefore $g_1 \equiv g_2$.

Note. 1. Given any $S(X)$ unbiased for $q(\theta)$, UMVUE is found by obtaining $T^*(X) = E(S(X)|T)$ where T is complete sufficient.

2. If we already have $h(T)$ unbiased for $q(\theta)$ and T is complete sufficient, then $h(T)$ is UMVUE, since $T^* = E(h(T)|T) = h(T)$.

Remark. The idea behind the L-S method is that, conditioning on a sufficient statistic (possibly) improves the estimator (R-B), and conditioning on a complete sufficient statistic gives the most possible improvement.

Example. Let X_1, \dots, X_n be i.i.d Poisson(λ), $\lambda > 0$. Find the UMVUE of $q(\lambda) = 1 - \exp(-\lambda)$. Note that $q(\lambda) = 1 - \exp(-\lambda) = P_\lambda(X_1 > 0)$. Therefore,

$$S(X_1, \dots, X_n) = I(X_1 > 0) = \begin{cases} 1 & \text{if } X_1 \geq 1; \\ 0 & \text{if } X_1 = 0 \end{cases}$$

is an unbiased estimator of $q(\lambda)$. Also, $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is complete sufficient for λ . (Check this.) Therefore,

$$h(T) = E(S(\mathbf{X})|T(\mathbf{X})) = E(I(X_1 > 0) | \sum_{i=1}^n X_i) = P(X_1 > 0 | \sum_{i=1}^n X_i)$$

is the unique UMVUE. We need the conditional distribution of $X_1 | \sum_{i=1}^n X_i$. Note that $X_1 | (\sum_{i=1}^n X_i = t) \sim \text{Binomial}(t, \frac{1}{n})$. (This is from the fact that the conditional joint distribution of the X_i 's is a multinomial, as shown previously.) Therefore

$$h(t) = P(X_1 > 0 | \sum_{i=1}^n X_i = t) = 1 - \binom{t}{0} \left(\frac{1}{n}\right)^0 \left(\frac{n-1}{n}\right)^t = 1 - \left(\frac{n-1}{n}\right)^t.$$

Thus, $1 - \left(\frac{n-1}{n}\right)^T$ is the UMVUE. (How does one show directly that this is unbiased?)

Example. Let X_1, \dots, X_n be i.i.d Bernoulli(p), $0 < p < 1$. Find UMVUE of p .

(i) Consider $S(X_1, \dots, X_n) = X_1$. Then $E(X_1) = p$ for all $0 < p < 1$, so that

S is an unbiased estimate of p . Therefore, the UMVUE of p is $E(X_1 | \sum_{i=1}^n X_i)$ since $\sum_{i=1}^n X_i$ is complete sufficient. Now see that

$$E(X_1 | \sum_{i=1}^n X_i) = E(X_2 | \sum_{i=1}^n X_i) = \cdots = E(X_n | \sum_{i=1}^n X_i).$$

Therefore,

$$E(X_1 | \sum_{i=1}^n X_i) = \frac{1}{n} \sum_{j=1}^n E(X_j | \sum_{i=1}^n X_i) = \frac{1}{n} E \left(\sum_{j=1}^n X_j | \sum_{i=1}^n X_i \right) = \frac{1}{n} \sum_{i=1}^n X_i.$$

(ii) $\sum_{i=1}^n X_i$ is complete sufficient. Also, $E(\sum_{i=1}^n X_i) = np$, so that $h(\sum_{i=1}^n X_i) = \frac{1}{n} \sum_{i=1}^n X_i$ is an unbiased estimate of p depending on the complete sufficient statistics only; thus unique UMVUE.