Lecture 9  UMVUE via Sufficiency and Completeness

A different approach towards UMVUE is introduced based on the idea of reducing the variance of an estimator by taking the conditional expectation given a sufficient statistic.

**Theorem 1 (Rao-Blackwell)** Let $U = U(X)$ be a sufficient statistic for $\theta$, and $T \in U(q)$. Then $\tilde{T} \overset{\triangle}{=} E(T|U)$ improves $T$ in the sense $\tilde{T} \in U(q)$ and $Var_{\theta} \tilde{T} \leq Var_{\theta} T \forall \theta \in \Theta$. Moreover, the equality holds if and only if for every $\theta$ we have $P_{\theta}(\tilde{T} = T) = 1$.

**Proof:** First, $\mathbb{E}_{\theta} \tilde{T} = \mathbb{E}_{\theta}[E(T|U)] = \mathbb{E}_{\theta}T = q(\theta) \forall \theta$, and $E(T|U)$ does not depend on $\theta$ due to the sufficiency of $U$. Second, for every $\theta$,

$$E_{\theta}[T - q(\theta)]^2 = E_{\theta}[(\tilde{T} - q(\theta))^2] + 2 E_{\theta}((\tilde{T} - q(\theta))E_{\theta}[(T - \tilde{T})|U])$$

implies that

$$Var_{\theta}T = Var_{\theta} \tilde{T} + E_{\theta}(T - \tilde{T})^2 \geq Var_{\theta} \tilde{T}.$$

Finally, the equality holds $\iff E_{\theta}(T - \tilde{T})^2 = 0 \iff P_{\theta}(\tilde{T} = T) = 1$ (measure theory needed here).  

QED.

Here is a fact regarding uniqueness of UMVUE.

**Proposition 1** If a UMVUE exists, then it is unique.

**Proof:** Suppose $T, T'$ are both UMVUEs for $q(\theta)$. Then $\tilde{T} = \frac{1}{2}(T + T') \in U(q)$, and

$$Var_{\theta}T = \frac{1}{4}Var_{\theta}T + \frac{1}{4}Var_{\theta}T' + \frac{1}{2}Cov_{\theta}(T, T')$$

$$\leq \frac{1}{4}Var_{\theta}T + \frac{1}{4}Var_{\theta}T' + \frac{1}{2}(Var_{\theta}T Var_{\theta}T')^{1/2} = Var_{\theta} \tilde{T}.$$

Hence $Var_{\theta} \tilde{T} = Var_{\theta}T = Cov_{\theta}(T, T')$. For every $\theta$ with $P_{\theta}$-probability one, $T' = a(\theta) T + b(\theta)$ for some functions $a(\theta)$ and $b(\theta)$. This in turn implies that $Cov_{\theta}(T, T') = a(\theta) Var_{\theta}T$, and $a(\theta) = 1$, $b(\theta) = 0$. Therefore, $P_{\theta}(T' = T) = 1$.  

QED.

**Definition 1** A statistic $U = U(X)$ is said to be complete for the family $\mathcal{F} = \{ P_{\theta} : \theta \in \Theta \}$, or simply for $\theta$, if any function $g$ with $E_{\theta}[g(U)] = 0 \forall \theta \in \Theta$ will satisfy $P_{\theta}(g(U) = 0) = 1 \forall \theta \in \Theta$.  

1
**Note:** The completeness is a property for a family $\mathcal{F}$, not just for a specific distribution $P_\theta$ or a particular parameter value $\theta$. $g(U)$ is referred to as an estimator of zero in the book of Casella and Berger, not necessarily a good term.

**Theorem 2 (Lehmann-Scheffé)** Let $U$ be a complete sufficient statistic for $\theta$ and $T \in \mathcal{U}(q)$. Then $T^* \triangleq E(T|U)$ is the unique UMVUE for $q(\theta)$.

**Proof:** Starting from any $T_1, T_2 \in \mathcal{U}$, let $T_i^* = g_i(U) \triangleq E(T_i|U)$, $i = 1, 2$. Then

$$0 = E_\theta(T_1^* - T_2^*) = E_\theta[g_1(U) - g_2(U)] \quad \forall \theta \in \Theta,$$

which implies that $P_\theta(g_1(U) = g_2(U)) = P_\theta(T_1^* = T_2^*) = 1 \forall \theta \in \Theta$. With a complete sufficient statistic $U$, a single step of “Rao-Blackwellization” will give rise to the unique UMVUE. QED.

**A recipe for finding UMVUE**

**Step 1:** Check the completeness of a sufficient statistic $U$;

**Step 2:** Find a $T \in \mathcal{U}$;

**Step 3:** Compute $T^* = E(T|U)$.

It is an art to balance Step 2 and Step 3, i.e. a good guess on $T \in \mathcal{U}$ will make the computation in Step 3 easy.

**Example 9.1** Let $X_1, \ldots, X_n$ be iid Poisson random variables with parameter $\theta$, and $q(\theta) = 1 - e^{-\theta} = P_\theta(X_1 > 0)$.

Step 1: To check the sufficient statistic $U = S_n = \sum_{i=1}^n X_i$ is complete, suppose

$$E_\theta[g(U)] = \sum_{u=0}^{\infty} g(u) \frac{e^{-n\theta} (n\theta)^u}{u!} = 0 \quad \forall \theta > 0.$$

Hence

$$\sum_{u=0}^{\infty} \left( g(u) \frac{n^u}{u!} \right) \theta^u = 0 \quad \forall \theta > 0,$$

which implies $g(u) \frac{n^u}{u!} = 0 \forall u = 0, 1, \ldots$ and $g(u) = 0 \forall u = 0, 1, \ldots$

Step 2: Propose $T = I\{X_1 > 0\} \in \mathcal{U}$.

Step 3: Note that

$$E(T|U = u) = 1 - P_\theta(X_1 = 0|S_n = u) = 1 - \frac{P_\theta(X_1 = 0) P_\theta(X_2 + \cdots + X_n = u)}{P_\theta(S_n = u)} = \left( \frac{n-1}{n} \right)^u.$$
Therefore, $T^* = 1 - \left(\frac{n-1}{n}\right)^{S_n}$ is the UMVUE. ($T^* \approx 1 - e^{-S_n/n}$ for large $n$?)

The next example shows that even for the simplest continuous model, measure theory is required for verification of completeness.

**Example 9.2** Let $X_1, ..., X_n$ be iid samples from $\mathcal{U}(0, \theta)$ and $q(\theta) = \theta$.

Step 1: To check $U = X_{(n)}$ is complete, note that

$$E_\theta[g(X_{(n)})] = \int_0^\theta g(x) \frac{nx^{n-1}}{\theta^n} \, dx = 0 \quad \forall \theta > 0$$

$$\Rightarrow \int_0^\theta g(x) \, x^{n-1} \, dx = 0 \quad \forall \theta > 0$$

$$\Rightarrow \int_A g(x) \, x^{n-1} \, dx = 0 \quad \forall \text{Borel set } A \subset (0, \infty)$$

$$\Rightarrow g(x) \, x^{n-1} = 0 \text{ and } g(x) = 0$$

for almost every $x > 0$ under the Lebesgue measure

$$\Rightarrow P_\theta(g(X_{(n)}) = 0) = 1 \quad \forall \theta > 0.$$ 

Step 2: Propose $T = \frac{n+1}{n} X_{(n)} \in \mathcal{U}$.

Step 3: $T$ is the UMVUE. (why?)

**Example 9.3** Let $X_1, ..., X_n$ be iid Cauchy random variables with location parameter $\theta$. We will show that the minimal sufficient statistic $U = (X_{(1)}, ..., X_{(n)})$ is not complete. Note that $E_\theta[\exp[-(X_{(n)} - X_{(1)})]] = E_\theta[\exp[-((X_{(n)} - \theta) - (X_{(1)} - \theta))] \overset{\triangle}{=} c_n$ which does not depend on $\theta$. Therefore, $g(U) \overset{\triangle}{=} \exp[-(X_{(n)} - X_{(1)})] - c_n$ satisfies $E_\theta[g(U)] = 0$ and $P_\theta(g(U) = 0) = 0 \quad \forall \theta \in \mathbb{R}$.

Note: It is stated in several books (e.g. Kiefer’s, Lehmann’s) that there are only two possible cases in any problem: either no complete statistics exist, or the set of complete sufficient statistics is the same as the set of minimal sufficient statistics. Hence complete statistics do not exist in Example 9.3.

**Example 9.4** Let $X_1, ..., X_n$ be iid samples from a minimal exponential family of rank $k$, with $X_1 \sim f(x|\theta) = \exp[\langle \theta, T(x) \rangle - \psi(\theta)]$. Then $T$ is a minimal sufficient statistic, and also a complete statistic. See the references for more details.

**Example 9.5** Let $X_1, ..., X_n$ be iid $\mathcal{N}(\mu, \sigma^2)$ random variables, and $\theta = (\mu, \sigma^2)$. Recall that $T = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ does not attain the C-R lower bound (Example 8.5). Nevertheless, $T$ is the UMVUE for $\sigma^2$ because $T$ is a function of the complete sufficient statistic $U = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$. (why?)