STAT5010 Advanced Statistical Inference	**Fall 2023**

Lecture 3: UMVUE

Lecturer: Tony Sit *Scribe: Muhammad DANISH and Junyou LI*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

3.1 Minimal Sufficiency, Ancillarity and Completeness

Theorem 3.1 (TSH 4.3.1) *$(T_1, .., T_s)$ is complete for any s-dimensional full rank exponential family (See P.117 of TSH)*

Theorem 3.2 (Basu's Theorem) *If T is complete and sufficient for $P = \{P_\theta : \theta \in \Theta\}$ and V is ancillary, then T(X) $\perp$ V(X).*

Proof: Define $G_A(t) = P_\theta(V \in A|T = t)$ or $g_A(t) = P_\theta(V \in A|T)$ and $p_A = P_\theta(V \in A$. Then by sufficiency and acillarity, neither p_A nor g_A depends upon θ. By Smoothing (tower expectation), $p_A = P_\theta(V \in A) = E_\theta(P_\theta(V \in A)|T) = E_\theta(g_A(T))$, and by completeness, we have $g_A(T) = p_A$.

Again by smoothing: $P_\theta(T \in V, V \in A) = E_\theta(1_B(T), 1_A(V)) = E_\theta(E_\theta(1_B(T)1_A(V)|T)) = E_\theta(1_B(T)E_\theta(1_A(V)|T)) = E_\theta(1_B(T)g_A(T)) = E_\theta(1_B(T).p_A) = E_\theta(1_B(T)).p_A = P_\theta(T \in B)P_\theta(V \in A)$. Hence T and V are independent as A and B are arbitrary Borel sets.

Example: Let $X_1, .., X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$ where both μ, σ^2 are unknown. Then, $\bar{X}_n \perp n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ where $\bar{X}_n = n^{-1}\sum_{i=1}^N X_i$

Fix any $\sigma > 0$ and consider sub-model $P_\sigma = \{N(\mu, \sigma^2)\}$. In each sub-model, $\bar{X}_n$ is complete and sufficient, and $n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ is ancillary. By Basu's theorem, $\bar{X}_n \perp \sum_{i=1}^n (X_i - \bar{X}_n)^2$ under $N(\mu, \sigma^2)$ for any $\mu \in \mathbb{R}$. Since σ is arbitrary, the conclusion holds for the full model $P = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma > 0\}$.

3.2 Optimal Inference

Definition: A function $f : \Theta \to \mathbb{R}$ where Θ is a convex function space and $\gamma \in (0, 1)$ is called convex if $f(\gamma x + (1 - \gamma)y) \leq \gamma f(x) + (1 - \gamma)f(y)$. The function is strictly convex if the inequality holds strictly.

Example: For any $\theta \in \Theta$, the function $\tilde{f}(d) = |d - \theta|$ is convex but not strictly convex while $\tilde{f}(d) = (d - \theta)^2$ is strictly convex.

Theorem 3.3 (Jensen's Inequality) *If a function $f : \Omega \to \mathbb{R}$ is convex on an open set Ω and $P(X \in \Omega) = 1$ and $E(X)$ exists, then $f(E(X)) \leq E(f(X))$. If f is strictly convex then the inequality holds strictly.*

Theorem 3.4 (Rao-Blackwell Theorem) *Suppose T is sufficient for $P = \{P_\theta : \theta \in \Theta\}$ such that $\delta(X)$ is an estimator for $g(\theta)$ for which $E(\delta(X))$ exists and that $R(\theta, \delta) = E_\theta(L(\theta, \delta) < \infty$. If, in particular, $L(\theta, .)$ is convex, then $R(\theta, \eta) \leq R(\theta, \delta)$ for any $\eta(T(X)) = E(\delta(X)|T(X))$.*

Proof: By Jensen's inequality, $E_\theta(L(g(\theta),\delta(X))|T) \geq L(g(\theta), E_\theta\delta(X)|T) = L(g(\theta),\eta(T))$. Taking another expectations, $E_\theta(L(g(\theta),\delta)) \geq E_\theta(L(g(\theta),\eta))$ which is equivalent to $R(g(\theta),\delta) \geq R(g(\theta),\eta)$. Note: Suppose $E_\theta(L(g(\theta),\delta)) = E_\theta(L(g(\theta),\eta))$, then $E_\theta(L(g(\theta),\delta)|T) = L(g(\theta),\eta)$ almost surely. Let $A = \{t : E(L(g(\theta),\delta(X))|T = t) = L(g(\theta),\eta(t))\}$, then $P_\theta(T \in A) = 1$. For $t \in A$, $\delta(X)$ is a constant random variable and so there exists $g(t)$ such that $P_\theta(\delta(X) = g(t)|T = t) = 1$ almost surely, which implies $P_\theta(\delta(X) = g(t)) = 1$.

3.3 Unbiased Estimation

Definition: An estimator is said to be unbiased if $E_\theta(\delta(X)) = g(\theta)$ for all $\theta \in \Theta$

We can find an unbiased estimator with uniformly minimum risk, when $L(\theta,d) = (\theta - d)^2$, then UMRUE becomes UMVUE because $E_\theta((g(\theta) - \delta(X))^2) = (E_\theta(\delta(X)) - g(\theta))^2 + E_\theta((\delta(X) - E_\theta(\delta(X)))^2)$ where the first term, the bias, is zero.

Definition: If an unbiased estimator exists, then $g(.)$ is called U-estimate.

Example: Suppose $X \sim U(0,\theta)$. Then δ is unbiased if $E_\theta(\delta(X)) = \int_0^\theta \delta(X)\theta^{-1}dx = g(\theta)$ or if $E_\theta(\delta(X)) = \int_0^\theta \delta(X)dx = \theta g(\theta)$ for any $\theta > 0$. So, g cannot be a U-estimate unless $\theta g(\theta) \to 0$ as $\theta \downarrow 0$. If g' exists, then by fundamental theorem of calculus, we have $\delta(X) = \partial/\partial X(xg(x)) = g(x) + xg'(x)$, then for $g(\theta) = \theta$, $\delta(X) = X + X(1) = 2X$.

Example: Suppose $X \sim Bin(n,\theta)$. If $g(\theta) = sin(\theta)$, then θ will be unbiased if $\sum_{k=0}^n \delta(k)\binom{n}{k}\theta^k(1-\theta)^{n-k} = sin(\theta)$, hence $g(\theta)$ is not U-estimable.

Definition: An unbiased estimator δ is UMVU if $Var_\theta(\delta) \leq Var(\delta')$ for all $\theta \in \Theta$ and δ' is any other competing unbiased estimator.

Theorem 3.5 (Lehmann-Scheffé Theorem) *If T is a complete and sufficient statistic, and $\mathbb{E}_\theta[h(T(X))] = g(\theta)$, i.e. $h(T(X))$ is unbiased for $g(\theta)$, then $h(T(X))$ is*

1) the only function of $T(X)$ that is unbiased for $g(\theta)$

2) an UMRUE under any convex loss function

3) the unique UMRUE (hence UMVUE), up to a $\mathcal{P}$-null set, under any strictly convex loss function.

Proof:
1) Suppose $\mathbb{E}_\theta\{\tilde{h}(T(X))\} = g(\theta)$, then

$$\mathbb{E}_\theta[\tilde{h}(T(X)) - h(T(X))] = 0, \quad \forall\theta \in \Theta$$

Thus, $\tilde{h}(T(X)) = h(T(X))$ a.s. for all $\theta \in \Theta$ by completeness.

2) Consider any unbiased $\delta(X)$, and let $\tilde{h}(T(X)) = \mathbb{E}_\theta[\delta(X) \mid T(X)]$. Then $\mathbb{E}_\theta[\tilde{h}(T(X))] = \mathbb{E}_\theta[\delta(X)] = g(\theta)$ by the tower property of conditional expectation. And by 1), $\tilde{h}(T(X)) = h(T(X))$ a.s., and by the Rao-Blackwell Theorem $R(\theta, h(T(X))) = R(\theta,\tilde{h}(T(X))) \leq R(\theta,\delta)$, for all $\theta \in \Theta$, if the loss function is convex. Therefore, $h(T(X))$ is an UMRUE under any convex loss function.

3) If the loss function is strictly convex, $R(\theta, h(T(X))) < R(\theta,\delta)$ unless $\delta(X) = h(T(X))$ a.s. Thus, $h(T(X))$ is the unique UMRUE (resp. UMVUE if the loss function adopted is the squared loss function).

Some useful strategies ("educated guesses") for finding UMVUE
1. Rao-Blackwellisation.

2. Solve for the (unique) δ satisfying $\mathbb{E}_\theta[\delta(T(X))] = g(\theta)$ for all $\theta \in \Theta$.
3. Guess (the right form of unbiased function of $T(X)$)

Example (Rao-Blackwellisation) Suppose $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim}$ Bernoulli (θ). We know that $T(X) = \sum_{i=1}^n X_i$ is a complete sufficient statistic, we also know that $n^{-1}T(X)$ is an unbiased estimator for θ, i.e.

$$\mathbb{E}_\theta\left(\frac{T(X)}{n}\right) = \frac{1}{n}\mathbb{E}_\theta\left(\sum_{i=1}^n X_i\right) = \frac{n\theta}{n} = \theta.$$

So $n^{-1}T(X)$ is an UMRUE for θ under any convex loss function. Suppose we want to estimate θ^2, let's examine: $\delta(X) = \mathbb{I}\{X_1 = X_2 = 1\} = X_1 \cdot X_2$. Then we have

$$\mathbb{E}_\theta[\delta(X)] = \mathbb{E}_\theta[\mathbb{I}\{X_1 = X_2 = 1\}] = \mathbb{E}_\theta(X_1 \cdot X_2) \overset{\text{i.i.d.}}{=} [\mathbb{E}_\theta(X_1)]^2 = \theta^2.$$

By conditioning on $T(X)$, we can find the UMRUE via

$$\begin{aligned}
\mathbb{E}_\theta[\delta(X) \mid T(X) = t] &= \mathbb{P}_\theta(X_1 = X_2 = 1 \mid T(X) = t) \\
&= \frac{\mathbb{P}_\theta\left(X_1 = X_2 = 1, \sum_{i=1}^n X_i = t - 2\right)}{\mathbb{P}_\theta(T(X) = t)} \\
&= \frac{\theta^2 \begin{pmatrix} n-2 \\ t-2 \end{pmatrix} \theta^{t-2}(1-\theta)^{n-t}\mathbb{I}\{t \geq 2\}}{\begin{pmatrix} n \\ t \end{pmatrix} \theta^2(1-\theta)^{n-t}} \\
&= \frac{t(t-1)\mathbb{I}\{t \geq 2\}}{n(n-1)}.
\end{aligned}$$

Note that in this case $\mathbb{I}\{t \geq 2\}$ is redundant as for $t = 0$ or 1 , the term $t(t-1) = 0$. Hence, we conclude that the UMRUE for θ^2 is

$$\frac{T(X)(T(X)-1)}{n(n-1)}.$$

Example (Rao-Blackwellisation) Suppose $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} U(0, \theta)$. In this case, $T(X) = X_{(n)}$ is a complete and sufficient statistic, and $\delta(X) = 2X$ is an unbiased estimator for θ, i.e. $\mathbb{E}_\theta(2X) = 2 \times \frac{\theta}{2} = \theta$.

Given $X_{(n)}, X_1$ is equal to $X_{(n)}$ with probability $1/n$ and follows uniform $\left(0, X_{(n)}\right)$ with probability $1 - 1/n$. Hence,

$$\mathbb{P}_\theta(X_1 = x_1 \mid T(X)) = \frac{1}{n} \times \mathbb{I}\{T(X) = x_1\} + \left(1 - \frac{1}{n}\right) \times \frac{1}{T(X)}\mathbb{I}\{0 < x_1 < T(X)\}$$

To find the UMVUE, we calculate

$$\begin{aligned}
\mathbb{E}_\theta[\delta(X) \mid T(X)] = 2\mathbb{E}_\theta[X_n \mid T(X)] &= 2\left\{\frac{1}{n}T(X) + \left(1 - \frac{1}{n}\right)\int_0^{T(X)} \frac{x_1}{T(X)} dx_1\right\} \\
&= 2\left\{\frac{T(X)}{n} + \left(1 - \frac{1}{n}\right)\frac{T(X)}{2}\right\} \\
&= \frac{n+1}{n}T(X).
\end{aligned}$$

Example (Solve for unique δ) Let $X \sim \text{Poisson}(\theta) . X$ is a complete and sufficient statistic, X is also unbiased, and therefore, UMVU for θ. Suppose we are interested in estimating $g(\theta)=e^{-a\theta}$ for a $\in \mathbb{R}$, known instead. We need to find an estimator δ such that $\mathbb{E}_\theta[\delta(X)]=g(\theta)$ for all $\theta>0$

$$\begin{aligned} & \mathbb{E}_\theta[\delta(X)]=\textstyle\sum_{x=0}^{\infty} \delta(x) \frac{e^{-\theta} \theta^x}{x!}=e^{-a\theta}, && \forall \theta>0 \\ \Leftrightarrow\ & \textstyle\sum_{x=0}^{\infty} \frac{\delta(x) \theta^x}{x!}=e^{(1-a)\theta}=\sum_{x=0}^{\infty} \frac{(1-a)^x \theta^x}{x!}, && \forall \theta>0 \\ \Rightarrow\ & \delta(X)=(1-a)^X \text{ is the UMVUE for } g(\theta) \end{aligned}$$

Note that the estimator is not ideal in the sense that if $a=2$, the estimator $\delta(X)=(-1)^X$ will change sign according to X's "evenness" even though the estimand $e^{-a\theta}$ is non-negative. The estimator is hence inadmissible when $a>1$ and dominated by $\max\{\delta(X), 0\}$.

Example (Guess) Let $X_1, \cdots, X_n \stackrel{\text{i.i.d.}}{\sim} N\left(\mu, \sigma^2\right)$. Consider the case where $\theta=\left(\mu, \sigma^2\right)$ is unknown.
1) The UMVUE for σ^2 is $\frac{1}{n-1} \sum_{i=1}^n\left(X_i-\bar{X}_n\right)^2=S^2$
2) How about the UMVUE for σ?
3) What is the UMVUE for μ^2?
For 2), Observe that

$$X_i-\bar{X}_n \sim N\left(0, \frac{n-1}{n} \sigma^2\right) \quad \Rightarrow \quad \mathbb{E}\left\{\left|X_i-\bar{X}_n\right|\right\}=\sigma \sqrt{\frac{2}{\pi}} \times \sqrt{\frac{n-1}{n}}$$

This implies

$$\delta^{\prime}=\frac{\sqrt{\pi n}}{\sqrt{2(n-1)}}\left|X_i-\bar{X}_n\right|$$

is unbiased for σ. While the Rao-Blackwellisation is very tedious. Instead, we can observe another fact that

$$S_*^2=\sum_{i=1}^n\left(X_i-\bar{X}_n\right)^2=(n-1) S^2$$

And $S_*^2 \sim \sigma^2 \chi_{n-1}^2$. Hence, $\mathbb{E}\left(S_*\right)=\sigma \mathbb{E}\left(\chi_{n-1}\right)$, which implies that

$$\frac{\mathbb{E}\left(S_*\right)}{\mathbb{E}\left(\chi_{n-1}\right)}=\sigma$$

meaning that $\frac{S_*}{\mathbb{E}\left(\chi_{n-1}\right)}$ is unbiased for σ and hence UMVU.
For 3), Taking the expecatation of the UMVUE for μ and squaring it, we obtain

$$\mathbb{E}\left(\bar{X}_n^2\right)=\mu^2+\frac{\sigma^2}{n}$$

So

$$\delta_n(X)=\bar{X}_n^2-\frac{S_*^2}{n(n-1)}$$

is the UMVUE. However, $\delta_n(X)$ can be negative even though the estimand is a non-negative quantity. The estimator is in fact inadmissible and dominated by the biased estimator $\max\left(0, \delta_n(X)\right)$ since

$$\mathbb{E}\left[\max\left(0, \delta_n(X)\right)\right] \neq \max\{\mathbb{E}(\delta(X)), 0\}.$$