STAT 543 ☺

Lec 17, M, Mar 3

- Homework 4 posted, due M, Mar 10
- Exam 1 solutions, grading key, summary posted

Sufficiency and Point Estimation

Remarks on Completeness

1. If T is complete, then T is boundedly complete; the converse is false.

← connection between sufficiency & completeness

2. If T is sufficient and boundedly complete, then T is minimal sufficient.

   So by Remark 1 above, if T is sufficient and complete, then T is minimal sufficient.

3. Suppose T is complete and $h_1(T), h_2(T)$ are two estimators of $\gamma(\underset{\sim}{\theta})$

$$
\begin{aligned}
&\text{if } \mathrm{E}_{\underset{\sim}{\theta}} h_1(T) = \gamma(\underset{\sim}{\theta}) = \mathrm{E}_{\underset{\sim}{\theta}} h_2(T), \ \text{ for all} \underset{\sim}{\theta} \in \Theta \\
\Rightarrow\ & \mathrm{E}_{\underset{\sim}{\theta}} u(T) = 0, \ \text{ for all} \underset{\sim}{\theta} \in \Theta, \quad \text{where } u(T) = h_1(T) - h_2(T) \\
\Rightarrow\ & P_{\underset{\sim}{\theta}}\big(u(T) = 0\big) = 1, \ \text{ for all} \underset{\sim}{\theta} \in \Theta \\
\Rightarrow\ & P_{\underset{\sim}{\theta}}\big(h_1(T) = h_2(T)\big) = 1, \text{for all} \underset{\sim}{\theta} \in \Theta
\end{aligned}
$$

   Hence, there can be at most one (i.e., unique) UE of a parametric function $\gamma(\underset{\sim}{\theta})$ that is a function of a complete statistic.

4. Let $T \equiv h(X_1, \ldots, X_n)$ be an UE of $\gamma(\theta)$ & suppose $\underset{\sim}{S}$ is sufficient.

Recall:

$T \xrightarrow[\text{Rao-Blackwellize}]{\text{sufficient } \underset{\sim}{S}} T^* = E(T|\underset{\sim}{S})$ is U.E of $\gamma(\theta)$ & $Var_\theta(T^*) \le Var_\theta(T)$

$T \xrightarrow[\text{L-S Theorem (next)}]{\text{sufficient \& complete } \underset{\sim}{S}} T^* = E(T|\underset{\sim}{S})$ is UMVUE of $\gamma(\theta)$!

Sufficiency and Point Estimation

Lehmann-Scheffe Theorem: Completeness + Sufficiency + UE = UMVUE

Lehmann-Scheffe Theorem. Let $f(\underset{\sim}{x}|\underset{\sim}{\theta}) = f(x_1, \ldots, x_n|\underset{\sim}{\theta})$ be the *joint* pdf/pmf of $(X_1, \ldots, X_n)$, $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p) \in \Theta \subset \mathbb{R}^p$. Let $\underset{\sim}{S} = (S_1, S_2, \ldots, S_k)$ be a complete and sufficient statistic. If $T^* \equiv T(\underset{\sim}{S})$ is an UE of $\gamma(\underset{\sim}{\theta})$ and is a function of $\underset{\sim}{S}$, then T^* is the UMVUE of $\gamma(\underset{\sim}{\theta})$.

Proof. Let T be any UE of $\gamma(\underset{\sim}{\theta})$. We must show $\text{Var}_{\underset{\sim}{\theta}}(T^*) \le \text{Var}_{\underset{\sim}{\theta}}(T)$

Define $T_1 = E(T|\underset{\sim}{S})$. Since $\underset{\sim}{S}$ is sufficient, by the Rao-Blackwell theorem, we know

T_1 is a function of $\underset{\sim}{S}$ & U.E. of $\gamma(\theta)$ &

$\text{Var}_\theta(T_1) \le \text{Var}_\theta(T)$, $\forall\theta$

Now T_1 & T^* are functions of $\underset{\sim}{S}$ (complete) & both are U.E of $\gamma(\theta)$

Since $\underset{\sim}{S}$ is complete, we know

$P_\theta(T_1 = T^*) = 1$, $\forall\theta$

$\Rightarrow \text{Var}_\theta(T^*) = \text{Var}_\theta(T_1) \le \text{Var}_\theta(T)$, $\forall\theta$

$\Rightarrow T^* = h(\underset{\sim}{S})$ is UMVUE of $\gamma(\theta)$ [& so is T_1]

Remark. The R-B theorem & L-S theorem together suggest two methods for finding the UMVUE:

Method I: Given a parametric function $\gamma(\underset{\sim}{\theta})$, find an UE of $\gamma(\underset{\sim}{\theta})$ that is a function of a complete and sufficient statistic.

easier

$\underset{\sim}{S}$ sufficiency + complete
$T^* = h(\underset{\sim}{S})$
$E_\theta T^* = \gamma(\theta)$, $\forall\theta$
then T^* is UMVUE

Method II: Start with any UE T of $\gamma(\underset{\sim}{\theta})$. Then $T^* = E(T|\underset{\sim}{S})$ is the UMVUE of $\gamma(\underset{\sim}{\theta})$, if $\underset{\sim}{S}$ is complete and sufficient.

a little harder

Find $T^* = E(T|\underset{\sim}{S}) = h(\underset{\sim}{S})$

Sufficiency and Point Estimation

Lehmann-Scheffe Theorem: Illustrations

Example. Let $X_1, \ldots, X_n$ be iid Poisson(θ), $\theta > 0$. Find the UMVUE of θ.

(could here use CRLB to find UMVUE)

Solution: Check $S = \sum_{i=1}^n X_i$ is sufficient (check by Factorization theorem) & is also complete (later)

Use $\bar{X}_n = \frac{S}{n} \Rightarrow$ check $E_\theta(\bar{X}_n) = E_\theta(X_1) = \theta, \forall \theta$

So $\bar{X}_n$ is UE of θ & a function of complete/sufficient $S \Rightarrow \bar{X}_n$ is UMVUE of θ.

$\tilde{S}^2 = \frac{n}{n-1}\left[\frac{\sum x_i^2}{n} - (\bar{X}_n)^2\right]$

Note: $\tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ has $E_\theta \tilde{S}^2 = Var_\theta(X_1) = \theta, \forall \theta > 0$

So $\tilde{S}^2$ is UE of θ. So, $E(\tilde{S}^2 | S) = \bar{X}_n$ by L-S theorem

Example. Let $X_1, \ldots, X_n$ be iid Bernoulli(θ), $0 < \theta < 1$. Find the UMVUE of $\gamma(\theta) = \theta^r(1-\theta)^{n-r}$, for a fixed (known) integer $1 \le r \le n$.

Solution: Check $S \equiv \sum_{i=1}^n X_i$ is sufficient & also complete (later)

Note $S \sim \text{Binomial}(n, \theta)$, $P_\theta(S = s) = \binom{n}{s}\theta^s(1-\theta)^{n-s}$, $0 \le s \le n$

Define $T^* = \begin{cases} \frac{1}{\binom{n}{r}} & \text{if } S = r \\ 0 & \text{o.w} \end{cases} = \frac{I[S=r]}{\binom{n}{r}}$

which is a function of S &

$E_\theta(T^*) = \frac{1}{\binom{n}{r}} E_\theta(I[S=r]) = \frac{P_\theta(S=r)}{\binom{n}{r}} = \gamma(\theta) = \theta^r(1-\theta)^{n-r}, \forall \theta$

So, T^* is UMVUE of $\tau(\theta)$ by L-S theorem.

Sufficiency and Point Estimation

Exponential Families (for Checking Sufficiency/Completeness)

Definition: A family of pdf/pmf $\{f(\underset{\sim}{x}|\underset{\sim}{\theta}) : \underset{\sim}{\theta} \in \Theta\}$, $\Theta \subset \mathbb{R}^p$, is called an **exponential family** if it can be written in the form

$$f(\underset{\sim}{x}|\underset{\sim}{\theta}) = \begin{cases} c(\underset{\sim}{\theta})h(\underset{\sim}{x})\exp\left[\sum_{i=1}^{k} q_i(\underset{\sim}{\theta})t_i(\underset{\sim}{x})\right] & \underset{\sim}{x} \in A \\ 0 & \text{otherwise} \end{cases}$$

where

$A \equiv \{\underset{\sim}{x} : f(\underset{\sim}{x}|\underset{\sim}{\theta}) > 0\}$ does NOT depend on $\underset{\sim}{\theta}$,

$c(\underset{\sim}{\theta}) > 0$ and $h(\underset{\sim}{x}) > 0$ are positive-valued functions,

and $q_i(\underset{\sim}{\theta})$, $t_i(\underset{\sim}{x})$ are real-valued functions for $i = 1, \ldots, k$.

Theorem: Let $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ be a (possibly vector-valued) random sample from $f(\underset{\sim}{x}|\underset{\sim}{\theta})$, where $\{f(\underset{\sim}{x}|\underset{\sim}{\theta}) : \underset{\sim}{\theta} \in \Theta\}$ is an exponential family admitting a representation as above. If

$$\left\{\left[q_1(\underset{\sim}{\theta}), \ldots, q_k(\underset{\sim}{\theta})\right] : \underset{\sim}{\theta} \in \Theta\right\} \supset (a_1, b_1) \times \cdots \times (a_k, b_k)$$

for some $a_i < b_i$, $i = 1, \ldots, k$, then

$$\underset{\sim}{S} = \left(\sum_{j=1}^{n} t_1(\underset{\sim}{X}_j), \ldots, \sum_{j=1}^{n} t_k(\underset{\sim}{X}_j)\right)$$

is complete and sufficient.