Biostat 602 Winter 2017

Lecture Set 10

Point Estimation

Rao Blackwell Theorem, Lehman-Scheffe Theorem

Reading: CB 7.3.3

Important Facts

Let X and Y be two random variables. Then

- $\mathrm{E}(X) = \mathrm{E}[\mathrm{E}(X|Y)]$ (Theorem 4.4.3)
- $\mathrm{Var}(X) = \mathrm{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}[\mathrm{E}(X|Y)]$ (Theorem 4.4.7)
- $\mathrm{E}[g(X)|Y] = \int_{x \in \mathcal{X}} g(x) f(x|Y) dx$ is a function of Y.
- If X and Y are independent, $\mathrm{E}[g(X)|Y] = \mathrm{E}[g(X)]$.

Searching for a better unbiased estimator

Suppose $W(\mathbf{X})$ is an unbiased estimator of $\tau(\theta)$, i.e. $\mathrm{E}[W(\mathbf{X})] = \tau(\theta)$. Further suppose $T(\mathbf{X})$ is any function of $\mathbf{X} = (X_1, \cdots, X_n)$.

Consider $\phi(T) = \mathrm{E}[W(\mathbf{X})|T]$.

$$
\begin{aligned}
\mathrm{E}[\phi(T)] &= \mathrm{E}[\mathrm{E}(W(\mathbf{X})|T)] = \mathrm{E}[W(\mathbf{X})] = \tau(\theta) \qquad \text{(unbiased for } \tau(\theta)\text{)} \\
\mathrm{Var}(\phi(T)) &= \mathrm{Var}[\mathrm{E}(W|T)] \\
&= \mathrm{Var}(W) - \mathrm{E}[\mathrm{Var}(W|T)] \\
&\leq \mathrm{Var}(W) \qquad \text{(equal or smaller variance than } W\text{)}
\end{aligned}
$$

Does this mean that $\phi(T)$ is a better estimator than $W(\mathbf{X})$?

1. If $\phi(T)$ is an estimator, then $\phi(T)$ is equal or better than $W(\mathbf{X})$.
2. $\phi(T) = \mathrm{E}[W|T] = \mathrm{E}[W|T, \theta]$.

$\phi(T)$ may depend on θ, which means that $\phi(T)$ may not be an estimator.

A Note about the notation $\mathbf{E}(\cdot)$, $\mathbf{E}_\theta(\cdot)$, and $\mathbf{E}(\cdot|\theta)$

To be explict that $\mathrm{E}[W|T]$ depends on θ, sometimes it is represented as $\mathrm{E}_\theta[W|T]$ as in the textbook. Note that most of $\mathrm{E}(\cdot)$ or $Var(\cdot)$ in this lecture note can be represented as $\mathrm{E}_\theta(\cdot) = \mathrm{E}(\cdot|\theta)$ or $Var_\theta(\cdot) = Var(\cdot|\theta)$

Example 1: Let $X_1, \cdots, X_n$ be an i.i.d. random sample from $\mathcal{N}(\theta, 1)$. Then $W(\mathbf{X}) = \frac{1}{2}(X_1 + X_2)$ is an unbiased estimator of θ. Consider conditioning it on $T(\mathbf{X}) = X_1$. Then

$$\begin{aligned}
\phi(T) &= \mathrm{E}[W|T] = \mathrm{E}\left[\frac{1}{2}(X_1 + X_2)|X_1\right] \\
&= \frac{1}{2}\mathrm{E}(X_1|X_1) + \frac{1}{2}\mathrm{E}(X_2|X_1) \\
&= \frac{1}{2}X_1 + \frac{1}{2}\mathrm{E}(X_2) \\
&= \frac{1}{2}X_1 + \frac{1}{2}\theta
\end{aligned}$$

- $\mathrm{E}[\phi(T)] = \frac{1}{2}\theta + \frac{1}{2}\theta = \theta$ (unbiased)
- $\mathrm{Var}[\phi(T)] = \frac{1}{4} < \mathrm{Var}(\frac{1}{2}(X_1 + X_2)) = \frac{1}{2}$
- But $\phi(T)$ is NOT an estimator.

Example 2: Consider again a random sample $X_1, \cdots, X_n$ from $\mathcal{N}(\theta, 1)$. Then $W(\mathbf{X}) = X_1$ is an unbiased estimator of θ. Consider conditioning it on $\overline{X}$.

$$
\begin{aligned}
\phi(T) &= \mathrm{E}[W|T] = \mathrm{E}(X_1|\overline{X}) \\
&= \frac{\mathrm{E}(X_1|\overline{X}) + \mathrm{E}(X_2|\overline{X}) + \cdots + \mathrm{E}(X_n|\overline{X})}{n} \\
&= \frac{\mathrm{E}(X_1 + \cdots + X_n|\overline{X})}{n} \\
&= \frac{\mathrm{E}(n\overline{X}|\overline{X})}{n} = \frac{n\overline{X}}{n} = \overline{X}
\end{aligned}
$$

- $\mathrm{E}[\phi(T)] = \theta$ (unbiased)
- $\mathrm{Var}[\phi(T)] = \frac{\mathrm{Var}(X)}{n} = \frac{1}{n} < \mathrm{Var}(W) = 1$
- $\phi(T)$ is an estimator, thus a better unbiased estimator.

Question: Why is $\mathrm{E}(X_1|\overline{X})$ free of θ?

Rao-Blackwell Theorem

Theorem 7.3.17: Let $W(\mathbf{X})$ be any unbiased estimator of $\tau(\theta)$, and T be a sufficient statistic for θ. Define $\phi(T) = \mathrm{E}[W|T]$. Then the following hold.

1. $\mathrm{E}[\phi(T)|\theta] = \tau(\theta)$
2. $\mathrm{Var}[\phi(T)|\theta] \leq \mathrm{Var}(W|\theta)$ for all θ.

That is, $\phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$.

Proof: Using the properties of $E(\cdot)$ and $Var(\cdot)$,

1. $\mathrm{E}[\phi(T)] = \mathrm{E}[\mathrm{E}(W|T)] = \mathrm{E}(W) = \tau(\theta)$ (unbiased)
2. $\mathrm{Var}[\phi(T)] = \mathrm{Var}[\mathrm{E}(W|T)] = \mathrm{Var}(W) - \mathrm{E}[\mathrm{Var}(W|T)] \leq \mathrm{Var}(W)$ (better than W).
3. Need to show $\phi(T)$ is indeed an estimator.

$$\begin{aligned} \phi(T) &= \mathrm{E}(W|T) = \mathrm{E}[W(\mathbf{X})|T] \\ &= \int_{\mathbf{x}\in\mathcal{X}} W(\mathbf{x}) f(\mathbf{x}|T) d\mathbf{x} \end{aligned}$$

Because T is a sufficient statistic, $f(\mathbf{x}|T)$ does not depend on θ.

Therefore, $\phi(T) = \int_{\mathbf{x}\in\mathcal{X}} W(\mathbf{x}) f(\mathbf{x}|T) d\mathbf{x}$ does not depend on θ, and $\phi(T)$ is indeed an estimator of θ.

Uniqueness

Theorem 7.3.19 (Uniqueness of UMVUE) If W is a best unbiased estimator of $\tau(\theta)$, then W is unique.

Proof: Suppose W_1 and W_2 are two best unbiased estimators of $\tau(\theta)$. Since both are 'best', $Var(W_1) = Var(W_2)$. Consider estimator $W_3 = \frac{1}{2}(W_1 + W_2)$.

$$\begin{aligned}
\mathrm{E}(W_3) &= \mathrm{E}\left(\frac{1}{2}W_1 + \frac{1}{2}W_2\right) = \frac{1}{2}\tau(\theta) + \frac{1}{2}\tau(\theta) = \tau(\theta) \\
\mathrm{Var}(W_3) &= \mathrm{Var}\left(\tfrac{1}{2}W_1 + \tfrac{1}{2}W_2\right) \\
&= \frac{1}{4}\mathrm{Var}(W_1) + \frac{1}{4}\mathrm{Var}(W_2) + \frac{1}{2}\mathrm{Cov}(W_1, W_2) \\
&\leq \frac{1}{4}\mathrm{Var}(W_1) + \frac{1}{4}\mathrm{Var}(W_2) + \frac{1}{2}\sqrt{\mathrm{Var}(W_1)\mathrm{Var}(W_2)} \\
&= \mathrm{Var}(W_1) = \mathrm{Var}(W_2)
\end{aligned}$$

If strict inequality holds, W_3 is better than W_1 and W_2, which is contradictory to the assumption.

Therefore, the equality must hold, requiring

$$\frac{1}{2}\mathrm{Cov}(W_1, W_2) = \frac{1}{2}\sqrt{\mathrm{Var}(W_1)\mathrm{Var}(W_2)}$$

By Cauchy-Schwarz (correlation) inequality, this is true if and only if $W_2 = aW_1 + b$.

Since both W_1, W_2 are unbiased estimators for $\tau(\theta)$

$$\mathrm{E}(W_1) = \tau(\theta) = \mathrm{E}(W_2) = \mathrm{E}(aW_1 + b) = a\tau(\theta) + b$$

and so $a = 1, b = 0$ must hold, yielding $W_2 = W_1$. Therefore, the best unbiased estimator is unique.

Unbiased estimator of zero

Definition

If $U(\mathbf{X})$ satisfies $\mathrm{E}(U) = 0$ for all $\theta \in \Omega$, then we call U an unbiased estimator of 0.

Relationship with ancillary statistics

Let $S(\mathbf{X})$ is an ancillary statistic for θ. $U(\mathbf{X}) = S(\mathbf{X}) - \mathrm{E}[S(\mathbf{X})]$ is always an unbiased estimator of zero. However, in general, an unbiased estimator of zero simply requires the expectation to be zero and need not be ancillary.

Theorem 7.3.20: If $\mathrm{E}[W(\mathbf{X})] = \tau(\theta)$, W is the best unbiased estimator of $\tau(\theta)$ if an only if W is uncorrelated with all unbiased estimator of 0.

Does the Theorem share some similarity to Basu's Theorem?

Proof: Let W be an unbiased estimator of $\tau(\theta)$. Let $V = W + U$ and $U \in \mathcal{U}$, which is the class of unbiased estimators of 0.

By construction, V is an unbiased estimator of $\tau(\theta)$. Consider

$$\mathcal{V} = \{V_a = W + aU\}$$

where a is a constant.

$$
\begin{aligned}
\mathrm{E}(V_a) &= \mathrm{E}(W + aU) = \mathrm{E}(W) + a\mathrm{E}(U) \\
&= \tau(\theta) + a \cdot 0 = \tau(\theta) \\
\mathrm{Var}(V_a) &= \mathrm{Var}(W + aU) \\
&= a^2\mathrm{Var}(U) + 2a\mathrm{Cov}(W, U) + \mathrm{Var}(W)
\end{aligned}
$$

The variance is minimized when

$$
a = \frac{-2\mathrm{Cov}(W, U)}{2\mathrm{Var}(U)} = -\frac{\mathrm{Cov}(W, U)}{\mathrm{Var}(U)}
$$

The best unbiased estimator in this class is

$$
W - \frac{\mathrm{Cov}(W, U)}{\mathrm{Var}(U)} U
$$

W is the best unbiased estimator in this class if and only if $\mathrm{Cov}(W, U) = 0$.

Therefore W is the best among all unbiased estimators of $\tau(\theta)$ if and only if $\mathrm{Cov}(W, U) = 0$ for every $U \in \mathcal{U}$.

Example 3: Let X be an observation from a Uniform$(\theta, \theta+1)$ distribution.

1. Is $X - \frac{1}{2}$ an unbiased estimator of θ?
2. Find an example of an unbiased estimator of zero.
3. Is $X - \frac{1}{2}$ the best unbiased estimator of θ?

Bias of $X - \frac{1}{2}$

$$\mathrm{E}X = \int_{\theta}^{\theta+1} x dx = \theta + \frac{1}{2}$$

so $X - \frac{1}{2}$ is an unbiased estimator of θ.

Further note that $Var(X - \frac{1}{2}) = VarX = \frac{1}{12}$.

Finding unbiased estimators of zero

If $U(X)$ is an unbiased estimator of zero, then it has to satisfy

$$\int_{\theta}^{\theta+1} U(x)dx = 0, \qquad \forall \theta \in \mathbb{R}$$

then

$$\frac{d}{d\theta}\int_{\theta}^{\theta+1} U(x)dx = U(\theta+1) - U(\theta) = 0, \qquad \forall \theta \in \mathbb{R}$$

So a periodic function with frequency 1 qualifies for $U(X)$. For example,

$$U(X) = \sin(2\pi X)$$

is an unbiased estimator for zero.

Is $X - \frac{1}{2}$ the best unbiased estimator?

If we can identify an unbiased estimator of zero, say $U(X)$ such that $Cov(X - \frac{1}{2}, U(X)) \neq 0$, then $X - \frac{1}{2}$ is not the best estimator.

Define $U(X) = \sin(2\pi X)$,

$$\begin{aligned}
\mathrm{Cov}\left(X - \frac{1}{2}, \sin(2\pi X)\right) &= \mathrm{Cov}\left(X, \sin(2\pi X)\right) \\
&= \int_{\theta}^{\theta+1} x \sin(2\pi x) dx \\
&= \left[-\frac{x\cos(2\pi x)}{2\pi}\right]_{\theta}^{\theta+1} + \int_{\theta}^{\theta+1} \frac{\cos(2\pi x)}{2\pi} dx \\
&= -\frac{\cos(2\pi\theta)}{2\pi}
\end{aligned}$$

Hence, $X - \frac{1}{2}$ is correlated with an unbiased estimator of zero, and cannot be the best unbiased estimator of θ.

In fact, $Var\left(X - \frac{1}{2} + \frac{1}{2}\sin(2\pi X)\right) = 0.071 < \frac{1}{12}$. This provides an example of an unbiased estimator for θ which has smaller variance than $X - \frac{1}{2}$.

Lehmann-Scheffé Theorem

In searching for best unbiased estimators, we explored two approaches

CRLB

CRLB provides a loose lower bound for the variance of the unbiased estimators. But its use is limited due to the fact that this bound is not attained too frequently even if there are best estimators.

Rao-Blackwell Theorem

Rao Blackwell Theorem is useful but calculating $\phi(T) = \mathrm{E}[W|T]$ is not usually an easy task. Further, $\phi(T)$ is a 'better' unbiased estimator, but may not be the 'best'.

Question: Is there a way to easily obtain the 'best' unbiased estimator?

Theorem 7.3.23 - Lehmann-Scheffé: Let T be a complete sufficient statistic for parameter θ. Let $\phi(T)$ be any estimator based on T. Then $\phi(T)$ is the unique best unbiased estimator of its expected value.

Another version of Lehmann-Scheffé: Let T be a complete sufficient statistic for parameter θ, Then $\phi(T) = \mathrm{E}[W|T]$ is the unique best unbiased estimator of $\mathrm{E}(W)$.

Proof: Let $\phi(T)$ be any function of T such that $\mathrm{E}[\phi(T)] = \tau(\theta)$. Let W be any unbiased estimator of $\tau(\theta)$, i.e. $\mathrm{E}(W) = \tau(\theta)$.

1. Because T is sufficient, $\psi(T) = \mathrm{E}[W|T]$ is a legitimate estimator. Further it is unbiased for $\tau(\theta)$ since

$$\mathrm{E}[\psi(T)] = \mathrm{E}\left(\mathrm{E}[W|T]\right) = \mathrm{E}(W) = \tau(\theta).$$

2. By Rao-Blackwell Theorem, $Var[\psi(T)] \leq Var(W)$ for all θ.

3. Let $g(T) = \phi(T) - \psi(T)$, then because both $\phi(T)$ and $\psi(T)$ are unbiased estimators for $\tau(\theta)$,

$$\mathrm{E}[g(T)|\theta] \;=\; \mathrm{E}[\phi(T) - \psi(T)|\theta] = 0$$

   Because T is a complete statistic, the above equation always implies

$$\Pr(g(T) = 0|\theta) \;=\; \Pr(\phi(T) = \psi(T)|\theta) = 1$$

   for all θ, meaning that $\psi(T)$ and $\phi(T)$ are identical in distribution.

4. Because for any unbiased estimator of W,

$$Var(W) \geq Var[\psi(T)] = Var[\phi(T)],$$

   so $\phi(T)$ is always the unique UMVUE for its expected value $\tau(\theta)$.

Note: When T is complete and sufficient, the Lehmann-Scheé Theorem implies that there is at most one function of T thats unbiased for $\tau(\theta)$.

Application of Lehman-Scheffé

1. Find a complete sucient statistic T.
2. If we can nd an unbiased estimator $V(T)$, we have found the UMVUE.
3. Otherwise, nd any unbiased estimator $W(X)$ and then compute $\phi(T) = \mathrm{E}[W|T]$.

Remarks

- From Rao-Blackwell Theorem, we can always improve an unbiased estimator by conditioning it on a sufficient statistics.
  - $W(\mathbf{X})$: unbiased for $\tau(\theta)$.
  - $T^*(\mathbf{X})$: sufficient statistic for θ.

  $\phi(T) = \mathrm{E}[W(\mathbf{X})|T^*(\mathbf{X})]$ is a better unbiased estimator of $\tau(\theta)$.

- Minimal sufficient statistics are more useful. In fact, we only need to consider functions of minimal sufficient statistics to find the best unbiased estimator.

  Let $T(\mathbf{X})$ be a minimal sufficient, and $T^*(\mathbf{X})$ be a sufficient statistic. Then by definition, there exists a function h that satisfies $T = h(T^*)$.

$$\begin{aligned} \mathrm{E}[\phi(T)|T^*] &= \mathrm{E}\left[\phi\left\{h(T^*)\right\}|T^*\right] \\ &= \phi\left\{h(T^*)\right\} = \phi(T) \end{aligned}$$

  Therefore $\phi(T)$ remains the same after conditioning on any sufficient statistic T^*.

- Complete sufficient statistics is a very useful ingredient to obtain a UMVUE.
  - We need to limit our search to the class of minimal sufficient statistics T to find the best unbiased estimator.
  - Let $\phi(T)$ be unbiased for $\mathrm{E}[\phi(T)] = \tau(\theta)$.
  - Consider $\{\phi(T) + U(T)|\ U(T) \in \mathcal{U}\}$, where $\mathcal{U}$ is unbiased estimators of zero among the functions of T.
  - By Theorem 7.3.20, $\phi(T)$ is UMVUE if and only if

$$\begin{aligned} \mathrm{Cov}(\phi(T), U(T)) &= \mathrm{E}[\phi(T)U(T)] - \mathrm{E}[\phi(T)]\mathrm{E}[U(T)] \\ &= \mathrm{E}[\phi(T)U(T)] = 0 \end{aligned}$$

  - If T is complete, $\phi(T)U(T) = 0$ almost surely, requiring $U(T) = 0$. Therefore, $\phi(T)$ is the best unbiased estimator of its expected value.

Example 4: Let $X_1, \cdots, X_n$ be i.i.d. from $\mathcal{N}(\mu, \sigma^2)$. Find the best unbiased estimator for (1) μ, (2) σ^2, (3) μ^2.

Solution:

- First, we need to find a complete and sufficient statistic for (μ, σ^2).
- We know that $\mathbf{T}(\mathbf{X}) = (\overline{X}, s^2_{\mathbf{X}})$ is complete, sufficient statistic for (μ, σ^2).
- Because $\mathrm{E}[\overline{X}] = \mu$, $\overline{X}$ is an unbiased estimator for μ, $\overline{X}$ is also a function of $\mathbf{T}(\mathbf{X})$.
- Therefore, $\overline{X}$ is the best unbiased estimator for μ.
- $\mathrm{E}(s^2_{\mathbf{X}}) = \sigma^2$
- $s^2_{\mathbf{X}}$ is a function of $\mathbf{T}$
- Therefore $s^2_{\mathbf{X}}$ is the best unbiased estimator of σ^2.

To obtain UMVUE for μ^2, we need a $\phi(\mathbf{T}) = \phi(\overline{X}, s^2_{\mathbf{X}})$ such that $\mathrm{E}[\phi(\mathbf{T})] = \mu^2$.

$$\begin{aligned}
\mathrm{E}(\overline{X}) &= \mu \\
\mathrm{E}((\overline{X})^2) &= \mathrm{Var}(\overline{X}) + \mathrm{E}[(\overline{X})]^2 = \frac{\sigma^2}{n} + \mu^2 \\
\mathrm{E}\left(\overline{X}^2 - \frac{\sigma^2}{n}\right) &= \mu^2 \\
\mathrm{E}\left(\overline{X}^2 - \frac{s^2_{\mathbf{X}}}{n}\right) &= \mu^2
\end{aligned}$$

- $\overline{X}^2 - s^2_{\mathbf{X}}/n$ is unbiased estimator for μ^2
- And it is a function of $(\overline{X}, s^2_{\mathbf{X}})$.
- Hence, $\overline{X}^2 - s^2_{\mathbf{X}}/n$ is the best unbiased estimator for μ^2.

Example 5: Let $X_1, \cdots, X_n$ be i.i.d. from $Bernoulli(p)$. Find the best unbiased estimator for (1) $p(1-p)$, (2) p^2.

Example 6: Let $X_1, \cdots, X_n$ be i.i.d. from $Poisson(\lambda)$. Find the best unbiased estimator for $\Pr(X_1 = 0) = \exp(-\lambda)$.