Lecture 16: UMVUE: conditioning on sufficient and complete statistics

The 2nd method of deriving a UMVUE when a sufficient and complete statistic is available

1. Find an unbiased estimator of $\vartheta$, say $U(X)$.
2. Conditioning on a sufficient and complete statistic $T(X)$: $E[U(X)|T]$ is the UMVUE of $\vartheta$.
3. We need to derive an explicit form of $E[U(X)|T]$.
4. From the uniqueness of the UMVUE, it does not matter which $U(X)$ is used.
5. Thus, we should choose $U(X)$ so as to make the calculation of $E[U(X)|T]$ as easy as possible.
6. We do not need the distribution of $T$.
   But we need to work out the conditional expectation $E[U(X)|T]$.
7. Using the independence of some statistics (Basu’s theorem), we may avoid to work on conditional distributions.
Example 7.3.24 (binomial family)

Let $X_1, \ldots, X_n$ be iid from $binomial(k, \theta)$ with known $k$ and unknown $\theta \in (0, 1)$.

We want to estimate $g(\theta) = P_{\theta}(X_1 = 1) = k\theta(1 - \theta)^{k-1}$.

Note that $T = \sum_{i=1}^{n} X_i \sim binomial(kn, \theta)$ is the sufficient and complete statistic for $\theta$.

But no unbiased estimator based on it is immediately evident.

To apply conditioning, we take the simple unbiased estimator of $P_{\theta}(X_1 = 1)$, the indicator function $I(X_1 = 1)$.

By Theorem 7.3.23, the UMVUE of $g(\theta)$ is

$$\psi(T) = E[I(X_1 = 1) | T] = P(X_1 = 1 | T)$$

We need to simply $\psi(T)$ and obtain an explicit form.

When $T = 0$, $P(X_1 = 1 | T = 0) = 0$.

For $t = 1, \ldots, kn$, 

\[ \psi(t) = P(X_1 = 1 | T = t) \]
\[ = \frac{P_\theta(X_1 = 1, \sum_{i=1}^{n} X_i = t)}{P_\theta(\sum_{i=1}^{n} X_i = t)} \]
\[ = \frac{P_\theta(X_1 = 1, \sum_{i=2}^{n} X_i = t - 1)}{P_\theta(\sum_{i=1}^{n} X_i = t)} \]
\[ = \frac{P_\theta(X_1 = 1)P_\theta(\sum_{i=2}^{n} X_i = t - 1)}{P_\theta(\sum_{i=1}^{n} X_i = t)} \]
\[ = k\theta(1 - \theta)^{k-1} \left[ \binom{k(n-1)}{t-1} \theta^{t-1} (1 - \theta)^{k(n-1)-(t-1)} \right] \]
\[ = \frac{k^t \theta^t (1 - \theta)^{kn-t}}{\binom{kn}{t}} \]

Hence, the UMVUE of \( g(\theta) = k\theta(1 - \theta)^{k-1} \) is

\[ \psi(T) = \begin{cases} \frac{k^t \theta^t (1 - \theta)^{kn-t}}{\binom{kn}{t}} & T = 1, \ldots, kn \\ 0 & T = 0 \end{cases} \]
Example 3.3

Let $X_1, \ldots, X_n$ be i.i.d. from the exponential distribution $E(0, \theta)$. 

\[ F_\theta(x) = (1 - e^{-x/\theta})I_{(0, \infty)}(x). \]

Consider the estimation of $\vartheta = 1 - F_\theta(t)$. 

$\bar{X}$ is sufficient and complete for $\theta > 0$. 

$I_{(t, \infty)}(X_1)$ is unbiased for $\vartheta$, 

\[ E[I_{(t, \infty)}(X_1)] = P(X_1 > t) = \vartheta. \]

Hence 

\[ T(X) = E[I_{(t, \infty)}(X_1)|\bar{X}] = P(X_1 > t|\bar{X}) \]

is the UMVUE of $\vartheta$. 

If the conditional distribution of $X_1$ given $\bar{X}$ is available, then we can calculate $P(X_1 > t|\bar{X})$ directly. 

By Basu’s theorem (Theorem 2.4), $X_1/\bar{X}$ and $\bar{X}$ are independent. 

By Proposition 1.10(vii), 

\[ P(X_1 > t|\bar{X} = \bar{x}) = P(X_1/\bar{X} > t/\bar{x}|\bar{X} = \bar{x}) \]

\[ = P(X_1/\bar{X} > t/\bar{x}) \]
To compute this unconditional probability, we need the distribution of

$$X_1 / \sum_{i=1}^{n} X_i = X_1 / \left( X_1 + \sum_{i=2}^{n} X_i \right).$$

Using the transformation technique discussed in §1.3.1 and the fact that $\sum_{i=2}^{n} X_i$ is independent of $X_1$ and has a gamma distribution, we obtain that $X_1 / \sum_{i=1}^{n} X_i$ has the Lebesgue p.d.f. $(n-1)(1-x)^{n-2}I_{(0,1)}(x)$.

Hence

$$P(X_1 > t|\bar{X} = \bar{x}) = (n-1) \int_{t/(n\bar{x})}^{1} (1-x)^{n-2} dx$$

$$= \left( 1 - \frac{t}{n\bar{x}} \right)^{n-1}$$

and the UMVUE of $\vartheta$ is

$$T(X) = \left( 1 - \frac{t}{n\bar{x}} \right)^{n-1}.$$
Example 3.4

Let $X_1, \ldots, X_n$ be i.i.d. from $N(\mu, \sigma^2)$ with unknown $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. From Example 2.18, $T = (\bar{X}, S^2)$ is sufficient and complete for $\theta = (\mu, \sigma^2)$.

$\bar{X}$ and $(n-1)S^2/\sigma^2$ are independent.

$\bar{X}$ has the $N(\mu, \sigma^2/n)$ distribution.

$S^2$ has the chi-square distribution $\chi^2_{n-1}$.

Using the method of solving for $h$ directly, we find that

- the UMVUE for $\mu$ is $\bar{X}$;
- the UMVUE of $\mu^2$ is $\bar{X}^2 - S^2/n$;
- the UMVUE for $\sigma^r$ with $r > 1 - n$ is $k_{n-1,r}S^r$, where
  \[ k_{n,r} = \frac{n^{r/2}\Gamma\left(\frac{n}{2}\right)}{2^{r/2}\Gamma\left(\frac{n+r}{2}\right)} \]
- the UMVUE of $\mu/\sigma$ is $k_{n-1,-1} \bar{X}/S$, if $n > 2$. 

UW-Madison (Statistics)  Stat 709 Lecture 16  2018  6 / 15
Example 3.4 (continued)

Suppose that $\vartheta$ satisfies $P(X_1 \leq \vartheta) = p$ with a fixed $p \in (0, 1)$. Let $\Phi$ be the c.d.f. of the standard normal distribution. Then

$$\vartheta = \mu + \sigma \Phi^{-1}(p)$$

and its UMVUE is

$$\bar{X} + k_{n-1.1} S \Phi^{-1}(p).$$

Let $c$ be a fixed constant and

$$\vartheta = P(X_1 \leq c) = \Phi \left( \frac{c - \mu}{\sigma} \right).$$

We can find the UMVUE of $\vartheta$ using the method of conditioning. Since $I_{(-\infty,c)}(X_1)$ is an unbiased estimator of $\vartheta$, the UMVUE of $\vartheta$ is

$$E[I_{(-\infty,c)}(X_1)|T] = P(X_1 \leq c|T).$$

By Basu’s theorem, the ancillary statistic $Z(X) = (X_1 - \bar{X})/S$ is independent of $T = (\bar{X}, S^2)$. 
Example 3.4 (continued)

Then, by Proposition 1.10(vii),

\[
P \left( X_1 \leq c \mid T = (\bar{x}, s^2) \right) = P \left( Z \leq \frac{c - \bar{x}}{S} \mid T = (\bar{x}, s^2) \right)
\]

\[
= P \left( Z \leq \frac{c - \bar{x}}{s} \right).
\]

It can be shown that \( Z \) has the Lebesgue p.d.f.

\[
f(z) = \frac{\sqrt{n} \Gamma \left( \frac{n-1}{2} \right)}{\sqrt{\pi(n-1)} \Gamma \left( \frac{n-2}{2} \right)} \left[ 1 - \frac{nz^2}{(n-1)^2} \right]^{(n/2)-2} I_{(0,(n-1)/\sqrt{n})}(|z|)
\]

Hence the UMVUE of \( \vartheta \) is

\[
P(X_1 \leq c \mid T) = \int_{-(n-1)/\sqrt{n}}^{(c-\bar{x})/S} f(z)dz
\]
Example 3.4 (continued)

Suppose that we would like to estimate

$$\vartheta = \frac{1}{\sigma} \Phi' \left( \frac{c - \mu}{\sigma} \right),$$

the Lebesgue p.d.f. of $X_1$ evaluated at a fixed $c$, where $\Phi'$ is the first-order derivative of $\Phi$.

By the previous result, the conditional p.d.f. of $X_1$ given $\bar{X} = \bar{x}$ and $S^2 = s^2$ is $s^{-1} f \left( \frac{x - \bar{x}}{s} \right)$.

Let $f_T$ be the joint p.d.f. of $T = (\bar{X}, S^2)$.

Then

$$\vartheta = \int \int \frac{1}{s} f \left( \frac{c - \bar{X}}{s} \right) f_T(t) dt = E \left[ \frac{1}{S} f \left( \frac{c - \bar{X}}{S} \right) \right].$$

Hence the UMVUE of $\vartheta$ is

$$\frac{1}{S} f \left( \frac{c - \bar{X}}{S} \right).$$
Example

Let $X_1, \ldots, X_n$ be i.i.d. with Lebesgue p.d.f. $f_\theta(x) = \theta x^{-2} I_{(\theta, \infty)}(x)$, where $\theta > 0$ is unknown.

Suppose that $\vartheta = P(X_1 > t)$ for a constant $t > 0$.

The smallest order statistic $X_{(1)}$ is sufficient and complete for $\theta$.

Hence, the UMVUE of $\vartheta$ is

$$P(X_1 > t | X_{(1)}) = P(X_1 > t | X_{(1)} = x_{(1)})$$

$$= P \left( \frac{X_1}{X_{(1)}} > \frac{t}{x_{(1)}} | X_{(1)} = x_{(1)} \right)$$

$$= P \left( \frac{X_1}{X_{(1)}} > \frac{t}{x_{(1)}} | X_{(1)} = x_{(1)} \right)$$

$$= P \left( \frac{X_1}{X_{(1)}} > s \right)$$

(Basu’s theorem), where $s = t / x_{(1)}$.

If $s \leq 1$, this probability is 1.
Example (continued)

Consider $s > 1$ and assume $\theta = 1$ in the calculation:

\[
P \left( \frac{X_1}{X(1)} > s \right) = \sum_{i=1}^{n} P \left( \frac{X_1}{X(1)} > s, X(1) = X_i \right) \\
= \sum_{i=2}^{n} P \left( \frac{X_1}{X(1)} > s, X(1) = X_i \right) \\
= (n - 1) P \left( \frac{X_1}{X(1)} > s, X(1) = X_n \right) \\
= (n - 1) P (X_1 > sX_n, X_2 > X_n, \ldots, X_{n-1} > X_n) \\
= (n - 1) \int_{x_1 > sx_n, x_2 > x_n, \ldots, x_{n-1} > x_n} \prod_{i=1}^{n} \frac{1}{x_i^2} dx_1 \cdots dx_n \\
= (n - 1) \int_1^\infty \left[ \int_1^{sx_n} \prod_{i=2}^{n-1} \left( \int_x^\infty \frac{1}{x_i^2} dx_i \right) \frac{1}{x_i^2} dx_1 \right] \frac{1}{x_n^2} dx_n \\
= (n - 1) \int_1^\infty \frac{1}{sx_n^{n+1}} dx_n = \frac{(n - 1)X(1)}{nt}
Example (continued)

This shows that the UMVUE of $P(X_1 > t)$ is

$$h(X_{(1)}) = \begin{cases} \frac{(n-1)X_{(1)}}{nt} & X_{(1)} < t \\ 1 & X_{(1)} \geq t \end{cases}$$

Another solution

The UMVUE must be $h(X_{(1)})$

The Lebesgue p.d.f. of $X_{(1)}$ is

$$\frac{n\theta^n}{x^{n+1}}I_{(\theta, \infty)}(x).$$

Use the method of finding $h$

If $\theta \geq t$, then $P(X_1 > t) = 1$ and $P(t > X_{(1)}) = 0$.

Hence, if $X_{(1)} \geq t$, $h(X_{(1)})$ must be 1 a.s. $P_\theta$

The value of $h(X_{(1)})$ for $X_{(1)} < t$ is not specified.
If \( \theta < t \),

\[
E[h(X_1))] = \int_{\theta}^{\infty} h(x) \frac{n\theta^n}{x^{n+1}} \, dx \\
= \int_{\theta}^{t} h(x) \frac{n\theta^n}{x^{n+1}} \, dx + \int_{t}^{\infty} \frac{n\theta^n}{x^{n+1}} \, dx = \int_{\theta}^{t} h(x) \frac{n\theta^n}{x^{n+1}} \, dx + \frac{\theta^n}{t^{n}}
\]

Since \( P(X_1 > t) = \frac{\theta}{t} \), we have

\[
\frac{\theta}{t} = \int_{\theta}^{t} h(x) \frac{n\theta^n}{x^{n+1}} \, dx + \frac{\theta^n}{t^{n}}
\]

i.e.,

\[
\frac{1}{t\theta^{n-1}} = \int_{\theta}^{t} h(x) \frac{n}{x^{n+1}} \, dx + \frac{1}{t^{n}}
\]

Differentiating both sizes w.r.t. \( \theta \) leads to

\[
- \frac{n-1}{t\theta^{n}} = - h(\theta) \frac{n}{\theta^{n+1}}
\]

Hence, for any \( X_1 < t \),

\[
h(X_1) = \frac{(n-1)X_1}{nt}.
\]
Unbiased estimators of 0

If a sufficient and complete statistic is not available, then what should we do?

If \( W \) is unbiased for \( \vartheta \) and \( T \) is sufficient, then by Theorem 2.5 (Rao-Blackwell), \( E(W|T) \) is better than \( W \).

If we have another sufficient statistic \( S \), should we consider \( E[E(W|T)|S] \)?

If there is a function \( h \) such that \( S = h(T) \), then by the properties of conditional expectation,

\[
\]

That is, we should always conditioning on a simpler sufficient statistic, such as a minimal sufficient statistic.

To see when an unbiased estimator is best unbiased, we might ask how could we improve upon a given unbiased estimator?

Suppose that \( T(X) \) is unbiased for \( g(\theta) \) and \( U(X) \) is a statistic satisfying \( E_\theta(U) = 0 \) for all \( \theta \), i.e., \( U \) is unbiased for 0.

Then, for any constant \( a \),
\[ T(X) + aU(X) \]

is unbiased for \( g(\theta) \).

Can it be better than \( T(X) \)?

\[
\text{Var}_\theta(T + aU) = \text{Var}_\theta(T) + 2a\text{Cov}_\theta(T, U) + a^2\text{Var}_\theta(U)
\]

If for some \( \theta_0 \), \( \text{Cov}_{\theta_0}(T, U) < 0 \), then we can make

\[
2a\text{Cov}_{\theta_0}(T, U) + a^2\text{Var}_{\theta_0}(U) < 0
\]

by choosing \( 0 < a - 2\text{Cov}_{\theta_0}(T, U)/\text{Var}_{\theta_0}(U) \).

Hence, \( T(X) + aU(X) \) is better than \( T(X) \) at least when \( \theta = \theta_0 \) and \( T(X) \) cannot be UMVUE.

Similarly, if \( \text{Cov}_{\theta_0}(T, U) > 0 \) for some \( \theta_0 \), then \( T(X) \) cannot be UMVUE either.

Thus, \( \text{Cov}_\theta(T, U) = 0 \) is necessary for \( T(X) \) to be a UMVUE, for all unbiased estimators of 0.

It turns out that \( \text{Cov}_\theta(T, U) = 0 \) for all \( U(X) \) unbiased for 0 is also sufficient for \( T(X) \) being a UMVUE.