Biostatistics 602 - Statistical Inference
Lecture 13
Rao-Blackwell Theorem

Hyun Min Kang

February 26th, 2013

Recap - Using Leibnitz’s Rule

Leibnitz’s Rule

\[
\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} f(x|\theta) \, dx = f(b(\theta)|\theta) b'(\theta) - f(a(\theta)|\theta) a'(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial \theta} f(x|\theta) \, dx
\]

Applying to Uniform Distribution

\[
\frac{d}{d\theta} \int_{0}^{1} h(x) \left( \frac{1}{\theta} \right) \, dx = \frac{h(0)}{\theta} \frac{d}{d\theta} \left( \frac{1}{\theta} \right) + \int_{0}^{\theta} \frac{\partial}{\partial \theta} h(x) \left( \frac{1}{\theta} \right) \, dx
\]

\[
\neq \int_{0}^{\theta} \frac{\partial}{\partial \theta} h(x) \left( \frac{1}{\theta} \right) \, dx
\]

The interchangeability condition is not satisfied.

Recap - When is the Cramer-Rao Lower Bound Attainable?

It is possible that the value of Cramer-Rao bound may be strictly smaller than the variance of any unbiased estimator.

**Corollary 7.3.15 : Attainment of Cramer-Rao Bound**

Let \( X_1, \cdots, X_n \) be iid with pdf/pmf \( f_X(x|\theta) \), where \( f_X(x|\theta) \) satisfies the assumptions of the Cramer-Rao Theorem. Let \( L(\theta|X) = \prod_{i=1}^{n} f_X(x_i|\theta) \) denote the likelihood function. If \( W(X) \) is unbiased for \( \tau(\theta) \), then \( W(X) \) attains the Cramer-Rao lower bound if and only if

\[
\frac{\partial}{\partial \theta} \log L(\theta|X) = S_n(x|\theta) = a(\theta)[W(X) - t(\theta)]
\]

for some function \( a(\theta) \).
Recap - Attainability of C-R bound for $\sigma^2$ in $\mathcal{N}(\mu, \sigma^2)$

1. If $\mu$ is known, the best unbiased estimator for $\sigma^2$ is $\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 / n$, and it attains the Cramer-Rao lower bound, i.e.

$$\text{Var} \left[ \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 \right] = \frac{2\sigma^4}{n}$$

2. If $\mu$ is not known, the Cramer-Rao lower-bound cannot be attained. At this point, we do not know if $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ is the best unbiased estimator for $\sigma^2$ or not.

Fact for one-parameter exponential family

Let $X_1, \ldots, X_n$ be iid from the one parameter exponential family with pdf/pmf $f_X(x|\theta) = c(\theta)h(x) \exp \{w(\theta)t(x)\}$.

Assume that $E[t(X)] = \tau(\theta)$. Then $\frac{1}{n} \sum_{i=1}^{n} t(x_i)$, which is an unbiased estimator of $\tau(\theta)$, attains the Cramer-Rao lower-bound. That is,

$$\text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} t(X_i) \right) = \frac{[\tau'(\theta)]^2}{I_n(\theta)}$$

Proof

$$E \left[ \frac{1}{n} \sum_{i=1}^{n} t(X_i) \right] = E[t(X_1)] = \cdots = E[t(X_n)] = \tau(\theta)$$

So, $\frac{1}{n} \sum_{i=1}^{n} t(x_i)$ is an unbiased estimator of $\tau(\theta)$.

$$\log L(\theta|x) = \sum_{i=1}^{n} \log f_X(x_i|\theta)$$

$$= \sum_{i=1}^{n} \left[ \log c(\theta) + \log h(x) + w(\theta)t(x_i) \right]$$

Proof (cont'd)

$$\frac{\partial \log L(\theta|x)}{\partial \theta} = \sum_{i=1}^{n} \left[ \frac{c'(\theta)}{c(\theta)} + 0 + w'(\theta)t(x_i) \right]$$

$$= nw'(\theta) \left[ \frac{1}{n} \sum_{i=1}^{n} t(x_i) - \left\{ -\frac{c'(\theta)}{c(\theta)w'(\theta)} \right\} \right]$$

- $\frac{1}{n} \sum_{i=1}^{n} t(x_i)$ is the best unbiased estimator of $-\frac{c'(\theta)}{c(\theta)w'(\theta)}$.
- And it attains the Cramer-Rao lower bound.
- Because $E\left[ \frac{\partial}{\partial \theta} \log L(\theta|x) \right] = 0$, $\tau(\theta) = -\frac{c'(\theta)}{c(\theta)w'(\theta)}$. 
Cramer-Rao Theorem on Exponential Family

**Fact**

\[
f_X(x|\theta) = c(\theta) h(x) \exp \left[ w(\theta) t(x) \right]
\]

If \( X_1, \ldots, X_n \) are iid samples from \( f_X(x|\theta) \), \( \frac{1}{n} \sum_{i=1}^{n} t(X_i) \) is the best unbiased estimator for its expected value. In other words,

\[
E[t(X)] = \tau(\theta)
\]

\[
\text{Var} \left[ \frac{1}{n} \sum_{i=1}^{n} t(X_i) \right] = \frac{[\tau'(\theta)]^2}{I_n(\theta)}
\]

**Proof**

\[
\frac{\partial}{\partial \theta} \log L(\theta|x) = nw'(\theta) \left[ \frac{1}{n} \sum_{i=1}^{n} t(X_i) - \tau(\theta) \right]
\]

\[
E \left[ \left( \frac{\partial}{\partial \theta} \log L(\theta|x) \right)^2 \right] = I_n(\theta) = E \left[ (nw'(\theta))^2 \left( \frac{1}{n} \sum_{i=1}^{n} t(X_i) - \tau(\theta) \right)^2 \right]
\]

\[
= \text{Var} \left[ nw'(\theta) \left( \frac{1}{n} \sum_{i=1}^{n} t(X_i) - \tau(\theta) \right) \right]
\]

\[
= n^2 \{ w'(\theta) \}^2 \text{Var} \left[ \frac{1}{n} \sum_{i=1}^{n} t(X_i) \right]
\]

\[
= n^2 \{ w'(\theta) \}^2 \frac{[\tau'(\theta)]^2}{I_n(\theta)}
\]

Obtaining \( I_n(\theta) \)

\[
E \left[ \left( \frac{\partial}{\partial \theta} \log L(\theta|x) \right)^2 \right] = I_n(\theta)
\]

\[
= n^2 \{ w'(\theta) \}^2 \frac{[\tau'(\theta)]^2}{I_n(\theta)}
\]

\[
[ nw'(\theta) ]^2 = \frac{I_n(\theta) \cdot I_n(\theta)}{[\tau'(\theta)]^2}
\]

\[
= \left( \frac{I_n(\theta)}{\tau'(\theta)} \right)^2
\]

\[
I_n(\theta) = \left| nw'(\theta) \tau'(\theta) \right|
\]
Summary

1. If "regularity conditions" are satisfied, then we have a Cramer-Rao bound for unbiased estimators of $\tau(\theta)$.
   - It helps to confirm an estimator is the best unbiased estimator of $\tau(\theta)$ if it happens to attain the CR-bound.
   - If an unbiased estimator of $\tau(\theta)$ has variance greater than the CR-bound, it does NOT mean that it is not the best unbiased estimator.

2. When "regularity conditions" are not satisfied, $\frac{[\tau'(\theta)]^2}{I_n(\theta)}$ is no longer a valid lower bound.
   - There may be unbiased estimators of $\tau(\theta)$ that have variance smaller than $\frac{[\tau'(\theta)]^2}{I_n(\theta)}$.

Important Facts

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

Methods for finding best unbiased estimator

1. Using Cramer-Rao bound
   - How do we find the best unbiased estimator?
2. Using Rao-Blackwell theorem
   - Use complete and sufficient statistic.
   - Find a 'better' unbiased estimator

Seeking for a better unbiased estimator

Suppose $W(X)$ is an unbiased estimator of $\tau(\theta)$. That is, $E[W(X)] = \tau(\theta)$. Suppose $T(X)$ is any function of $X = (X_1, \cdots, X_n)$. Consider

$$\phi(T) = E(W(X)|T)$$
$$E[\phi(T)] = E[E(W(X)|T)] = E[W(X)] = \tau(\theta) \quad \text{(unbiased for } \tau(\theta))$$
$$\text{Var}(\phi(T)) = \text{Var}[E(W|T)] = \text{Var}(W) - E[\text{Var}(W|T)] \leq \text{Var}(W) \quad \text{(smaller variance than } W)$$
A better unbiased estimator?

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

1. If $\phi(T)$ is an estimator, then $\phi(T)$ is equal or better than $W(X)$.

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

Example 2

Let $X_1, \cdots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$. $W(X) = \frac{1}{2}(X_1 + X_2)$ is an unbiased estimator of $\theta$.

Consider conditioning it on $T(X) = X_1$.

\[
\phi(T) = E[W|T] = E\left[\frac{1}{2}(X_1 + X_2)|X_1\right]
\]

\[
= \frac{1}{2}E(X_1|X_1) + \frac{1}{2}E(X_2|X_1)
\]

\[
= \frac{1}{2}X_1 + \frac{1}{2}E(X_2)
\]

\[
= \frac{1}{2}X_1 + \frac{1}{2}\theta
\]

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

Rao-Blackwell Theorem

\[\text{Theorem 7.3.17}\]

Let $W(X)$ be any unbiased estimator of $\tau(\theta)$, and $T$ be a sufficient statistic for $\theta$.

Define $\phi(T) = E[W|T]$. Then the followings hold.

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

That is, $\phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$. 
Proof of Rao-Blackwell Theorem

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

$$\phi(T) = E(W|T) = E[W(X)|T] = \int_{x \in \mathcal{X}} W(x)f(x|T)dx$$

Because $T$ is a sufficient statistic, $f(x|T)$ does not depend on $\theta$. Therefore, $\phi(T) = \int_{x \in \mathcal{X}} W(x)f(x|T)dx$ does not depend on $\theta$, and $\phi(T)$ is indeed an estimator of $\theta$.

Proof of Theorem 7.3.19 (cont’d)

$\text{Var}(W_3) \leq \text{Var}(W_1) = \text{Var}(W_2)$.

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

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

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

$$\text{Cov}(W_1, W_2) = \text{Cov}(W_1, aW_1 + b) = a\text{Var}(W_1)$$
$$= \text{Var}(W_1)\text{Var}(W_2) = \text{Var}(W_1)$$

$$E(W_2) = a\tau(\theta) + b$$
$$= \tau(\theta)$$

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

Uniqueness of UMVUE

**Theorem 7.3.19**

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 $\tau(\theta)$. Consider estimator $W_3 = \frac{1}{2}(W_1 + W_2)$.

$$E(W_3) = E\left(\frac{1}{2}W_1 + \frac{1}{2}W_2\right) = \frac{1}{2}\tau(\theta) + \frac{1}{2}\tau(\theta) = \tau(\theta)$$

$$\text{Var}(W_3) = \text{Var}\left(\frac{1}{2}W_1 + \frac{1}{2}W_2\right)$$
$$= \frac{1}{4}\text{Var}(W_1) + \frac{1}{4}\text{Var}(W_2) + \frac{1}{2}\text{Cov}(W_1, W_2)$$
$$\leq \frac{1}{4}\text{Var}(W_1) + \frac{1}{4}\text{Var}(W_2) + \frac{1}{2}\sqrt{\text{Var}(W_1)\text{Var}(W_2)}$$
$$= \text{Var}(W_1) = \text{Var}(W_2)$$

Unbiased estimator of zero

**Definition**

If $U(X)$ satisfies $E(U) = 0$. Then we call $U$ an unbiased estimator of 0.

**Theorem 7.3.20**

If $E[W(X)] = \tau(\theta)$. $W$ is the best unbiased estimator of $\tau(\theta)$ if an only if $W$ is uncorrelated with all unbiased estimator of 0.
### Proof of Theorem 7.3.20

Let $W$ be an unbiased estimator of $\tau(\theta)$. Let $V = W + U$ and $U \in U$, which is the class of unbiased estimators of 0. By construction, $V$ is an unbiased estimator of $\tau(\theta)$. Consider

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

where $a$ is a constant.

$$E(V_a) = E(W + aU) = E(W) + aE(U)$$
$$\tau(\theta) + a \cdot 0 = \tau(\theta)$$

$$\text{Var}(V_a) = \text{Var}(W + aU)$$
$$= a^2 \text{Var}(U) + 2a \text{Cov}(W, U) + \text{Var}(W)$$

The variance is minimized when

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

The best unbiased estimator in this class is

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

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

### Summary

- Cramer-Rao Theorem with single parameter exponential family.
- Rao-Blackwell Theorem

### Today

- Cramer-Rao Theorem with single parameter exponential family.
- Rao-Blackwell Theorem

### Next Lecture

- More Rao-Blackwell Theorem