STAT-510 **2017 Fall**

Lecture 5: Best Unbiased Estimatior

Instructor: Xiaohui Chen Scribe: Yutong Dai Last Modified: 2018-02-13

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.*

As discussed in the previous lecture, a comparison based on MSE may not yield a clear favorite. For example, suppose the true $\theta = 1$, then the estimator $\hat{\theta}(X) = 17$ cannot be beaten. But it is a terrible estimator otherwise.

To cope with this issue, we need to limit it to a smaller class. A popular choice is to restrict estimators to be unbiased. Indeed, this choice can be extended to a more general situation.

5.1 Best Unbiased Estimatior

Consider the class of estimators

$$C_\tau = \{W : E_\theta W = \tau(\theta)\}.$$

The $\tau(\theta)$ is not necessarily equal to θ. For any $W_1, W_2 \in C_\tau$, $Bias_\theta W_1 = Bias_\theta W_2$, so

$$E_\theta(W_1 - \theta)^2 - E_\theta(W_2 - \theta)^2 = \mathbf{Var}_\theta(W_1) - \mathbf{Var}_\theta(W_2).$$

The MSE comparison can be based on the variance alone. Thus, although we speak of unbiased estimators, we really are comparing estimators that have the same expected value, $\tau(\theta)$.

Definition 5.1 An estimators W^* is a best unbiased estimator of $\tau(\theta)$ if it satisfies,

1. $E_\theta W^* = \tau(\theta)$ for all θ;

2. For any other estimator W with $E_\theta W = \tau(\theta)$, we have $\mathbf{Var}_\theta(W^*) \leq \mathbf{Var}_\theta(W)$ for all θ.

W^* is also called uniform minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$.

And the best unbiased estimator is unique.

Theorem 5.2 (uniqueness) If W is a best unbiased estimator of $\tau(\theta)$, then W is unique

Proof: Suppose W' is another best unbiased estimator of $\tau(\theta)$, and consider the estimator $W^* = \frac{1}{2}(W + W')$. Note that $E_\theta(W^*) = \tau(\theta)$ and

$$\begin{aligned}\mathbf{Var}_\theta W^* &= \frac{1}{4}\mathbf{Var}_\theta W + \frac{1}{4}\mathbf{Var}_\theta W' + \frac{1}{2}\mathbf{cov}_\theta(W, W') \\ &\leq \frac{1}{4}\mathbf{Var}_\theta W + \frac{1}{4}\mathbf{Var}_\theta W' + \frac{1}{2}\sqrt{\mathbf{Var}_\theta W \mathbf{Var}_\theta W'} \\ &= \mathbf{Var}_\theta W. \end{aligned} \tag{5.1}$$

As $\mathbf{Var}_\theta W$ is a best unbiased estimator, we know that $\mathbf{Var}_\theta W^* = \mathbf{Var}_\theta W$. The equality holds iff $W' = a(\theta)W + b(\theta)$. As $\mathbf{cov}(W, W') = a(\theta)\mathbf{Var}_\theta(W)$ and $\mathbf{cov}(W, W') = \mathbf{Var}_\theta(W)$ by the Cauchy-Schwarz inequality. Therefore, $a(\theta) = 1$. Also, due to the unbiasness of W^*, $b(\theta) = 0$. These mean that $W' = W$. ■

However, finding a best unbiased estimator is not a easy task. Situations as following might happen:

1. The calculation of finding the variance of a estimator is quite complicated;
2. The unbiased estimator doesn't exist;
3. There are too many unbiased estimators and it's hard to find a one with the minimum variance;
4. Some unbiased estimators lurk.

Example 5.3 (Non-existence) Let $X \sim Binomial(n, p), p \in (0, 1)$. Consider we want to estimate the parameter $1/p$. Suppose there exists an unbiased estimator $W(X)$ such that $E_\theta(W(X)) = 1/p$. Then we have

$$\sum_{x=0}^{n} C_n^x p^x (1-p)^{1-x} w(x) = 1/p, \quad \forall 0 < p < 1.$$

If $p \to 0$, $1/p \to +\infty$ and $\sum_{x=0}^{n} C_n^x p^x (1-p)^{1-x} w(x) \to 0$, hence no such a uniformly unbiased estimator exists.

Example 5.4 (Too many unbiased estimator) Let $X_i \overset{iid}{\sim} Poission(\lambda)$. Then $E(S_n^2) = \sigma^2 = \lambda = E(\bar{x}_n)$. $\forall \alpha \in (0, 1), W(X) = \alpha \bar{x}_n + (1-\alpha)S_n^2$ is an unbiased estimator.

These examples suggest that we need a more comprehensive strategy to search for the best unbiased estimator. One workaround is that if we can specify the lower bound $B(\theta)$ of any unbiased estimator of $\tau(\theta)$, then we can try to find an unbiased estimator W^* with $\mathbf{Var}_\theta(W^*) = B(\theta)$.

5.2 Cramer-Rao Lower Bound

Theorem 5.5 (Cramer-Rao Lower Bound) Let $X_1, ..., X_n$ are iid with pdf $f(x; \theta)$, and $W(X)$ be any estimator satisfying

1. $E_\theta(W(X)) = \tau(\theta)$;
2. $\forall \theta \in \Theta, E_\theta[\frac{\partial}{\partial \theta} \log f(x; \theta)] = 0$;
3. $\forall \theta \in \Theta, \frac{\partial}{\partial \theta} E_\theta[W(X)] = \int \frac{\partial}{\partial \theta} W(x) f(x; \theta) dx$;
4. $\mathbf{Var}_\theta(W(X))$ is finite.

Then

$$\mathbf{Var}_\theta(W(X)) \geq \frac{(\tau'(\theta))^2}{I_X(\theta)}, \forall \theta \in \Theta$$

.

Proof: Let $Z = W(X), Y = \frac{d}{d\theta} \log f(X; \theta)$. Since $\mathbf{Var}_\theta(Y) = I_X(\theta) - E_\theta(Y)^2 = I_X(\theta)$,

$$\begin{aligned}
\mathbf{cov}_\theta(Y,Z) &= E_\theta[ZY - YE_\theta Z] = E_\theta[ZY] \\
&= E_\theta[W(X)\frac{d}{d\theta}\log f(X;\theta)] \\
&= \int W(x)f'(x;\theta)dx \\
&= \frac{d}{d\theta}\int W(x)f(x;\theta)dx \\
&= \frac{d}{d\theta}E_\theta(W(X)) \\
&= \tau'(\theta)
\end{aligned} \tag{5.2}$$

By Cauchy-Schwarz inequality, we have

$$\mathbf{Var}_\theta(W(X)) \geq \frac{(\tau'(\theta))^2}{I_X(\theta)}, \forall\theta\in\Theta.$$

■

Corollary 5.6 If the assumption of the Theorem 5.5 are satisfied and $\tau(\theta)=\theta$, then

$$\mathbf{Var}_\theta(W(X)) \geq \frac{1}{nI_{X_1}(\theta)}, \forall\theta\in\Theta.$$

Remark

1. Although the CRLB is stated for continuous random variables, it also applies to the discrete random variables. We need to require the summation and differentiation are interchangeable.

2. If there's an unbiased estimator $W(X)$ achieves the CRLB, then it is a UMVUE.

3. If $W(X)$ doesn't achieve the CRLB, it still could be a UMVUE. (Lehmann–Scheffé Theorem, next section)

4. The CRLB can be generalized to a biased estimators. Suppose $E_\theta(W(X)) = g(\theta) \neq \tau(\theta)$, then $bias(W(X)) = g(\theta) - \tau(\theta)$. If other assumptions in Theorem 5.5 holds, we have

$$\mathbf{Var}_\theta(W(X)) = \mathbf{Var}_\theta(W(X) - g(\theta) + \tau(\theta)) \geq \frac{(\tau'(\theta))^2}{I_X(\theta)}, \forall\theta\in\Theta$$

Example 5.7 $X_i \overset{iid}{\sim} Possion(\lambda)$, and $W(X) = \bar{X}_n, E(W(X)) = \lambda$. $\mathbf{Var}(W(X)) = \lambda/n$ and CRLB is $\frac{1}{nI_{X_1}(\lambda)} = \lambda/n$. Therefore, $\bar{x}_n$ is an UMVUE.

It's important to remember that a key assumption in the Cramer-Rao Theorem is the ability to differentiate under the integral sign, which, of course, is somewhat restrictive. But densities in the **exponential family** will satisfy assumptions. But in general, the story will change.

Example 5.8 (Condition Fails) $X_i \overset{iid}{\sim} U[0,\theta],\ 0 < \theta < +\infty$. Then $f(x;\theta) = \frac{1}{\theta}I(0 \le x \le \theta)$. Suppose $W(X)$ is an unbiased estimator of θ, and the CRLB holds, we will have

$$\mathbf{Var}_\theta(W(X)) \ge \frac{1}{nI_{X_1}(\theta)} = \frac{\theta^2}{n}.$$

Consider $Y = \max_i X_i$, then $f(y;\theta) = \frac{ny^{n-1}}{\theta^n}$ and $E_\theta(Y) = \frac{n\theta}{n+1}$. So the $\frac{n+1}{n}Y$ is another unbiased estimator of θ. However,

$$\mathbf{Var}_\theta(\frac{n+1}{n}Y) = \frac{1}{n(n+2)}\theta^2 \le \frac{1}{n}\theta^2.$$

The reason lies in the regularity condition $\forall\theta \in \Theta, E_\theta[\frac{\partial}{\partial\theta}\log f(x;\theta)] = 0$ fails, i.e. the integration and differentiation are not interchangeable.

Corollary 5.9 (Attainment) If X_i be *iid* $f(x;\theta)$, where $f(x;\theta)$ satisfies conditions of the CRLB theorem. Let $L(X;\theta) = \prod_i f(x_i;\theta)$ denote the likelihood function. If $W(X) = W(X_1, ..., X_n)$ is any unbiased estimator of $\tau(\theta)$, then $W(X)$ attains the CRLB if and only if

$$a(\theta)(W(X) - \tau(\theta)) = \frac{\partial}{\partial\theta}\log L(X;\theta)$$

for some function $a(\theta)$.

Remark This corollary suggests a way to find UMVUE:

1. Calculate $\frac{\partial}{\partial\theta}\log L(X;\theta)$;
2. Recognize $W(X)$ and $\tau(\theta)$;
3. Verify whether $E_\theta(W(X)) = \tau(\theta)$

If all are satisfied, then $W(X)$ is a UMVUE.

Example 5.10 $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ and μ is known.

$$\frac{\partial}{\partial\theta}\log L(X;\theta) = \frac{n}{2\sigma^4}[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - \sigma^2].$$

Set $W(X) = S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$ and $\tau(\theta) = \sigma^2$, then by Corollary 5.9, we know that S_n^2 is a UMVUE.

5.3 Sufficiency and Unbiasedness

The theory develpoed in the last section still some questions unanswered.

1. What can we do if $f(x;\theta)$ doesn't satisfy the Cramer-Rao Lower Bound assumption?
2. What if the Cramer-Rao Lower Bound is unattianable by allowable estimators?

In this section, we answer the question 2 by introducing the Rao-Blackwell theorem and Lehmann–Scheffé theorem. But before directly jumping into them, let's discuss some motivations.

When we given an unbiased estimator W for $\tau(\theta)$, one might ask how could we improve upon the W? This question motivates a the following theorem.

Theorem 5.11 If $E_\theta(W) = \tau(\theta), W$ is the UMVUE $\iff$ W is uncorrelated with all unbiased estimators of 0.

Proof: "$\Rightarrow$"

Let V be an unbiased estimation of 0. Suppose $\mathbf{cov}_\theta(W, V) \neq 0$. Let $W^* = W + aV$, then $E_\theta(W^*) = \tau(\theta), \forall\theta$.

$$\mathbf{Var}_\theta(W^*) = \mathbf{Var}_\theta(W) + 2a\mathbf{cov}_\theta(W, V) + a^2\mathbf{Var}_\theta(V).$$

Now, if for some $\theta = \theta_0$, $\mathbf{cov}_{\theta_0}(W, V) < 0$, then we can make $2a\mathbf{cov}_{\theta_0}(W, V) + a^2\mathbf{Var}_{\theta_0}(V) < 0$ by setting $a \in (0, -2\mathbf{cov}_{\theta_0}(W, V)/\mathbf{Var}_{\theta_0}(V)$. Hence W^* will be better than W at θ_0. This is a contradiction. Situations are similar for $\mathbf{cov}_{\theta_0}(W, V) > 0$.Therefore,

$$\forall\theta, \mathbf{cov}_\theta(W, V) = 0.$$

"$\Leftarrow$"

Suppose we have another unbaised estimator W', then $V = W - W'$ is an unbiased estimator for 0. Then $\forall\theta, \mathbf{cov}_\theta(W, V) = 0$.

$$\mathbf{Var}_\theta(W') = \mathbf{Var}_\theta(W) - 2\mathbf{cov}_\theta(W, V) + \mathbf{Var}_\theta(V) = \mathbf{Var}_\theta(W) + \mathbf{Var}_\theta(V).$$

This means that $\mathbf{Var}_\theta(W') \geq \mathbf{Var}_\theta(W)$ and thus $\mathbf{Var}_\theta(W)$ is the UMVUE. ■

Remark

1. This theorem tells us that if an unbiased estimator W can be improved by adding a estimator of 0 (i.e. a random noise), then it cannot be the best.
2. This theorem is much more useful in appliction when determing an estimator is not the best unbiased one.

Example 5.12 $X \sim U(\theta, \theta + 1), \theta \in R$. $W = X - \frac{1}{2}$ is an unbiased estimator for θ. Let $h(x) = 2\pi \sin x$, then $V = E(h(X)) = 0$. But $\mathbf{cov}_\theta(W, V) \neq 0$ for some θ. Then W is not the UMVUE.

If the family of pdfs or pmfs has the property that there are no unbiased estimators of 0 (other than 0 itself), then our search would be ended. The property of completeness comes on the stage.

Theorem 5.13 (Rao-Blackwell) Let W be any unbiased estimator of $\tau(\theta)$, and let T be a sufficient statistic for θ. Define $\phi(T) = E(W|T)$. Then

1. $E_\theta[\phi(T)] = \tau(\theta)$;
2. $\mathbf{Var}_\theta \phi(T) \leq \mathbf{Var}_\theta W$ for all θ; that is $\phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$.

Proof: The proof can be completed using the fact

$$EX = E[E(X|Y)] \quad \mathbf{Var}(X) = Var(E(X|Y)) + E(Var(X|Y)).$$

$$E_\theta(\phi(T)) = E_\theta[E_\theta[W|T]] = E[W] = \tau(\theta).$$

$$\begin{aligned}
\mathbf{Var}_\theta(W) &= \mathbf{Var}_\theta[E[W|T]] + E_\theta[\mathbf{Var}_\theta[W|T]] \\
&= \mathbf{Var}_\theta[\phi(T)] + E_\theta[\mathbf{Var}_\theta[W|T]] \\
&\geq \mathbf{Var}_\theta[\phi(T)].
\end{aligned} \tag{5.3}$$

The only things left unfinished is to show that $\phi(T)$ is an estimator. By the sufficiency of T, we know the distribution of $W|T$ doesn't depend on θ. That is $\phi(T) = E_\theta[W|T]$ doesn't depend on θ. ■

Remark:

1. This theorem tells us that conditioning any unbiased estimator for θ on an sufficients statistic for θ will result in a uniform improvement, so we need consider only statistics that are functions of a sufficient statistic in our search for the best unbiased estimators.

2. T needs to be sufficient to guarantee that $\phi(T)$ is an estimator. Otherwise, this claim fails. For example, $X_1, X_2 \overset{iid}{\sim} N(\mu, 1)$. $W(X) = \frac{1}{2}(X_1 + X_2), T(X) = X_1$, then $\phi(T) = E(W|T) = \frac{1}{2}(X_1 + \mu)$ is not an estimator.

3. If T^* is a minimal sufficient statistics, then $\mathbf{Var}_\theta[\phi(T)] \geq \mathbf{Var}_\theta[\phi(T^*)]$. Intuitively, this is as a result of T^* achieving maximum data reduction.

Now, we only look at unbiased estimator based on a sufficient statistic.

Theorem 5.14 (Lehmann–Scheffé) Let T be a complete and sufficient statistic for θ. Let $\phi(T)$ be any estimator based only on T. Then $\phi(T)$ is the unique best unbiased estimator of $\tau(\theta) = E_\theta(\phi(T))$.

Proof: For any unbiased estimator W for $\tau(\theta)$, we apply the Rao-Blackwell theorem on it. Then we know that $\phi^*(T) = E[W|T]$ is an unbiased estimator for $\tau(\theta)$ and $\mathbf{Var}_\theta(\phi^*(T)) \leq \mathbf{Var}_\theta(W)$.

By the unbiasedness, for any θ, we have $E_\theta(\phi^*(T) - \phi(T)) = 0$. And by the completeness of T, we know $P_\theta(\phi^*(T) - \phi(T) = 0) = 1$.

Thereore,

$$\mathbf{Var}_\theta(\phi(T)) = \mathbf{Var}_\theta(\phi^*(T)) \leq \mathbf{Var}_\theta(W).$$

By the arbitrariness of W, $\phi(T)$ is the UMVUE. ■

Remark

1. Lehmann–Scheffé theorem tells us a systematic way of finding the UMVUE for targeted $\tau(\theta)$.

- Construct $h(X)$ such that $E_\theta[h(X)] = \tau(\theta)$;
- Find T that is complete and sufficient for θ;
- Obtain the UMVUE: $\phi(T) = E_\theta[h(X)|T]$.

  The choice of $h(X)$ is quite arbitrary, as long as it is an unbiased estiamator. So the rule of thumb is to choose one that can faciliate the computation of $E_\theta[h(X)|T]$.

2. Lehmann–Scheffé theorem tells us that an unbiased estimator who cannot attatin the CRLB still can be a UMVUE. For example, consider $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, where both μ and σ^2 are unknown. Notice that the sample variance $S_n^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x}_n)^2$ is an unbiased, sufficient and complete estimator for σ^2, we can let $\phi(S_n^2) = S_n^2$. Then by Lehmann–Scheffé theorem, we know that S^2 is a UMVUE. But $\mathbf{Var}_\theta(S_n^2) = \frac{2\sigma^4}{n-1} > \frac{2\sigma^2}{n} =$ CRLB, that shows even if an unbiased estimator who cannot attatin the CRLB still can be a UMVUE. (This computaion involves the general Cramer-Rao theorem for the multiple parameters case.)

An unbiased estimator conditioning on a complete and sufficient statistics will produce a UMVUE. A natural question is that if condition on a complete and **mininmal** sufficient statistics, what will happen? Since the UMVUE is unique, so we shall obtain the same UMVUE. But it seems that the completeness eliminate the gap between **mininmal** sufficient statistics and sufficient statistics. Indeed the following theorem tells us that if $T(x)$ is a sufficient and complete, then it is the minimal sufficient statistics.

Theorem 5.15 (Bahadur's Theorem) Suppose that $T(X)$ taking values in R^k is sufficient for θ and complete. Then $T(X)$ is minimal sufficient.

Proof: See here. ■