STA 522, Spring 2021
Introduction to Theoretical Statistics II

Lecture 8

Department of Biostatistics
University at Buffalo

22 March, 2021

AGENDA

- Wrap up discussion on Cramér-Rao Lower Bound
- Rao-Blackwell Theorem
- Lehmann–Scheffé THeorem
- Cramér-Rao Lower Bound

Review: UMVUE & Cramér-Rao lower bound

- An estimator W^* is a **uniform minimum variance unbiased estimator** (UMVUE) of $\tau(\theta)$ if (a) W^* is unbiased, and (b) among all unbiased estimators, the variance (or MSE) of W^* is a minimum.
- **CRLB:** Let $\underline{X} = (X_1, X_2, \ldots, X_n)$ have pdf $f(\underline{x} \mid \theta)$, and let $W(\underline{X})$ be any estimator satisfying
  (a) $\displaystyle \frac{d}{d\theta} \mathsf{E}_\theta \left[W(\underline{X})\right] = \int_{\mathcal{X}} \frac{\partial}{\partial\theta} \left[W(\underline{x}) f(\underline{x} \mid \theta)\right] d\underline{x}$; and
  (b) $\mathsf{Var}_\theta \left[W(\underline{X})\right] < \infty$.

  Then $$\mathsf{Var}_\theta(W(\underline{X})) \geq \frac{\left[\frac{d}{d\theta} \mathsf{E}_\theta \left[W(\underline{X})\right]\right]^2}{\underbrace{\mathsf{E}_\theta \left[\left[\frac{\partial}{\partial\theta} \log f(\underline{X} \mid \theta)\right]^2\right]}_{\text{Fisher Information}}}.$$
- If an estimator satisfies the above two assumptions, and its variance attains the CRLB, then the estimator is UMVUE.
- There is no guarantee that the bound given in the Cramér-Rao Inequality is sharp. That is, our best unbiased estimator may not achieve the CRLB.

Example: Let $X_1, X_2, \ldots, X_n \sim$ iid Poisson(λ). (a) Compute the CRLB for an unbiased estimator for $\tau(\lambda) = e^{-\lambda} = P(X_1 = 0)$. (b) Consider $W = \frac{1}{n}\sum_{i=1}^{n} I(X_i = 0)$. Show that W is an unbiased estimator for $\tau(\lambda)$ whose variance is larger than the CRLB.

At the outset, note that Poisson distribution is a member of the (regular) expoential family, and therefore the two conditions in the CRLB hold.

First find Fisher Information (iid form). Here common log pmf is

$$\log f(x \mid \lambda) = -\lambda + x \log \lambda - \log(x!) \implies \frac{\partial \ \log f(x \mid \lambda)}{d\lambda} = -1 + \frac{x}{\lambda} = \frac{(x-\lambda)}{\lambda}$$

Therefore, Fisher information

$$\mathrm{E}_\lambda\left[\left[\frac{\partial}{\partial\lambda}\log f(\underline{X} \mid \lambda)\right]^2\right] = n\,\mathrm{E}_\lambda\left[\frac{(X_1-\lambda)^2}{\lambda^2}\right] = n\frac{\mathrm{Var}_\lambda(X_1)}{\lambda^2} = \frac{n}{\lambda}$$

Therefore, the CRLB for an unbiased estimator of $\tau(\lambda) = e^{-\lambda}$ is:

$$\mathrm{CRLB} = \frac{\left[\frac{d}{d\theta}\ \tau(\theta)\right]^2}{\mathrm{E}_\lambda\left[\left[\frac{\partial}{\partial\lambda}\log f(\underline{X} \mid \lambda)\right]^2\right]} = \frac{\left(-e^{-\lambda}\right)^2}{\frac{n}{\lambda}} = \frac{\lambda\ e^{-2\lambda}}{n}$$

Now consider $W = \frac{1}{n}\sum_{i=1}^{n} I(X_i = 0)$. To find its variance first define $U_i = I(X_i = 0)$. Then

$$U_i \sim \text{iid Bernoulli}\left(p = P(X_1 = 0) = e^{-\lambda}\right)$$

which implies $Z = \sum_{i=1}^{n} U_i = \sum_{i=1}^{n} I(X_i = 0) \sim \text{Binomial}\left(n, e^{-\lambda}\right)$ and

$$W = \frac{1}{n}\sum_{i=1}^{n} I(X_i = 0) = \frac{Z}{n} \implies \text{Var}_\lambda(W) = \frac{\text{Var}_\lambda(Z)}{n^2} = \frac{e^{-\lambda}\left(1 - e^{-\lambda}\right)}{n}$$

Therefore

$$\text{Var}_\lambda(W) - \text{CRLB} = \frac{e^{-\lambda}\left(1 - e^{-\lambda}\right)}{n} - \frac{\lambda\ e^{-2\lambda}}{n} = \frac{e^{-\lambda}(1 - (\lambda + 1)e^{-\lambda})}{n}$$

Now consider the function $g(\lambda) = 1 - (\lambda + 1)e^{-\lambda}$. Then

$$g'(\lambda) = -\left(e^{-\lambda} - (\lambda + 1)e^{-\lambda}\right) = \lambda e^{-\lambda} > 0 \ \text{ for all } \lambda > 0$$

which means that $g(\lambda)$ is increasing in λ for $\lambda > 0$, so that $g(\lambda) > g(0) = 0$ for all $\lambda > 0$, i.e., $g(\lambda) > 0$ for all $\lambda > 0$.

Hence, $\text{Var}_\lambda(W) - \text{CRLB} = \dfrac{e^{-\lambda}(1 - (\lambda + 1)e^{-\lambda})}{n} > 0$ for all $\lambda > 0$.

HW: Show that $\overline{X}$ is a UMVUE for λ.

Attainment of Cramér-Rao Inequality

Result (Corollary 7.3.15)

Let $X_1, X_2, \ldots, X_n$ be iid $f(x \mid \theta)$, where $f(x \mid \theta)$ satisfies the conditions of the Cramér-Rao Theorem.

Let $L(\theta \mid \underline{x}) = \prod_{i=1}^{n} f(x_i \mid \theta)$ denote the likelihood function.

If $W(\underline{X}) = W(X_1, X_2, \ldots, X_n)$ is any unbiased estimator of $\tau(\theta)$, then $W(\underline{X})$ attains the Cram'er-Rao Lower Bound if and only if

$$a(\theta)\,[W(\underline{x}) - \tau(\theta)] = \frac{\partial}{\partial\theta} \log L(\theta \mid \underline{x})$$

for some function $a(\theta)$.

Proof: (steps)

- Write the Cramér-Rao Inequality as

$$\left[\mathrm{Cov}_\theta\left[W(\underline{X}), \frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(x_i\,|\,\theta)\right]\right]^2 \leq \mathrm{Var}_\theta\left[W(\underline{X})\right]\mathrm{Var}_\theta\left[\frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(x_i\,|\,\theta)\right].$$

- Now use property of correlation coefficient: for any random variables X and Y, $|\rho_{XY}| = 1$ if and only if there exist numbers $a \neq 0$ and b such that $P(Y = aX + b) = 1$.
- $\mathrm{E}_\theta\left[W(\underline{X})\right] = \tau(\theta)$ for all θ due to unbiasedness.
- $\mathrm{E}_\theta\left[\frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(x_i\,|\,\theta)\right] = 0$ (proved in lecture 7).
- Thus, we can have equality if and only if $W(\underline{X}) - \tau(\theta)$ is proportional to $\frac{\partial}{\partial\theta}\log L(\theta\,|\,\underline{x})$

Rao-Blackwell Theorem

Recall: Tower Property. Let X and Y be any two random variables. Then, provided the expectations exist, we have

(a) $\mathrm{E}(X) = \mathrm{E}\left[\mathrm{E}(X \mid Y)\right]$

(b) $\mathrm{Var}(X) = \mathrm{Var}\left[\mathrm{E}(X \mid Y)\right] + \mathrm{E}\left[\mathrm{Var}(X \mid Y)\right]$

Theorem (7.3.17)

Let W be any unbiased estimator of $\tau(\theta)$, and let T be a sufficient statistic for θ. Define $\phi(T) = \mathrm{E}(W \mid T)$. Then

(a) $\mathrm{E}_\theta\left[\phi(T)\right] = \tau(\theta)$ for all θ; and

(b) $\mathrm{Var}_\theta\left[\phi(T)\right] \leq \mathrm{Var}_\theta(W)$ for all θ.

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

Proof:

For (a) We have

$$\mathsf{E}_\theta\left[\phi(T)\right] = \mathsf{E}_\theta\left[\mathsf{E}(W \mid T)\right] = \mathsf{E}_\theta(W) = \tau(\theta) \text{ for all } \theta.$$

For (b) note that

$$\begin{aligned}\mathsf{Var}_\theta(W) &= \mathsf{Var}_\theta\left[\mathsf{E}(W \mid T)\right] + \mathsf{E}_\theta\left[\mathsf{Var}(W \mid T)\right] \\ &= \mathsf{Var}_\theta\left[\phi(T)\right] + \underbrace{\mathsf{E}_\theta\left[\mathsf{Var}(W \mid T)\right]}_{\geq 0} \geq \mathsf{Var}_\theta\left[\phi(T)\right]\end{aligned}$$

for all θ.

It remains to show that $\phi(T) = \mathsf{E}(W \mid T)$ is indeed an estimator, i.e., is a function only of the sample, and is free of θ.

This follows from sufficiency – W being a function of sample the conditional distribution of $W \mid T$ is free of θ.

Finding UMVUEs

(a) So conditioning any unbiased estimator on a sufficient statistic will result in a uniform improvement.

(b) Thus, to find UMVUEs, we need only consider statistics that are functions of a sufficient statistic.

(c) But if T is sufficient, how do we know that $\phi(T)$ is a best unbiased estimator (UMVUE)?

(d) If it attains the CRLB, then it is best unbiased (UMVUE).

(e) What if it does not? We need a few more results to answer this question.

Uniqueness of UMVUEs

Theorem 7.3.19

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

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

$$\begin{aligned}
\mathrm{Var}_\theta(W^*) &= \mathrm{Var}_\theta\left(\frac{1}{2}(W + W')\right) \\
&= \frac{1}{4}\mathrm{Var}_\theta(W) + \frac{1}{4}\mathrm{Var}_\theta(W') + \frac{1}{2}\mathrm{Cov}_\theta(W, W') \\
&\overset{\text{Cauchy-Schwarz}}{\leq} \frac{1}{4}\mathrm{Var}_\theta(W) + \frac{1}{4}\mathrm{Var}_\theta(W') + \frac{1}{2}\sqrt{\mathrm{Var}_\theta(W)\,\mathrm{Var}_\theta(W')} \\
&= \mathrm{Var}_\theta(W) \qquad (\mathrm{Var}_\theta(W) = \mathrm{Var}_\theta(W'))
\end{aligned}$$

Since both W and W^* are UMVUEs the above inequality cannot be strict for any θ, i.e., must have equality in the Cauchy-Schwarz inequality.

Equality in Cauchy-Schwarz inequality holds only if $W' = a(\theta)W + b(\theta)$ for some $a(\theta)$ and $b(\theta)$.

Necessary and Sufficient Conditions for UMVUE

Theorem (7.3.20)

If $\mathrm{E}_\theta(W) = \tau(\theta)$, then W is the best unbiased estimator of $\tau(\theta)$ if and only if W is uncorrelated with all unbiased estimators of 0.

NOTE: this theorem can sometimes be used to show that an unbiased estimator is not UMVUE, by showing that the estimator is correlated with an unbiased estimator of 0.

Proof: (If part) Let W be the UMVUE and U be any unbiased estimator of 0. Define $\phi_a = W + aU$ for some arbitrary number a. Then ϕ_a is unbiased for $\tau(\theta)$ and

$$\mathrm{Var}_\theta(\phi_a) = \mathrm{Var}_\theta(W + aU) = \mathrm{Var}_\theta(W) + a^2\,\mathrm{Var}_\theta(U) + 2a\,\mathrm{Cov}_\theta(W, U)$$

If for some $\theta = \theta_0$, $\text{Cov}_{\theta_0}(W, U) < 0$ then for $a \in \left(0, -2\frac{\text{Cov}_{\theta_0}(W,U)}{\text{Var}_{\theta_0}(U)}\right)$ we have $a^2 \text{Var}_\theta(U) + 2a \text{Cov}_{\theta_0}(W, U) < 0$, i.e., $\text{Var}_\theta(\phi_a) < \text{Var}_\theta(W)$ for $\theta = \theta_0$ which contradicts to W being UMVUE.

Similar contradiction arises if $\text{Cov}_{\theta_0}(W, U) < 0$.

Hence $\text{Cov}_\theta(W, U)$ must be 0 for all θ and for all U.

(Only if part): Reading exercise. See p. 345 of the textbook.

Lehmann–Scheffé Theorem

Theorem

Let T be a complete sufficient statistic for a parameter θ, and let $\phi(T)$ be any estimator based only on T. Then $\phi(T)$ is the unique best unbiased estimator of its expected value.

Proof: Since T is complete, $X = 0$ is the only unbiased estimator of 0. Since $\phi(T)$ is uncorrelated with 0, and hence uncorrelated with all unbiased estimators of 0, we have that $\phi(T)$ is UMVUE of $\mathsf{E}_\theta[\phi(T)]$.

Remark: The Lehmann–Scheffé theorem and the Rao-Blackwell theorem together provide UMVUE for parametric functions from many standard probability distributions.

Suppose we want the UMVUE for $\tau(\theta)$. We have a complete sufficient statistic T for θ and we have an unbiased estimator W of $\tau(\theta)$. Then the Rao-Blackwell estimator $\phi(T) = \mathsf{E}[W \mid T]$ is UMVUE for $\tau(\theta)$.

Example: Let $X_1, X_2, \ldots, X_n \sim$ iid Poisson(λ). (a) Find the UMVUE of λ, if it exists. (b) Find the UMVUE of $\tau(\lambda) = e^{-\lambda} = P(X = 0)$, if it exists.

At the outset note that for Poisson (a member of the exponential family) $T = \sum_{i=1}^{n} X_i$ is complete sufficient for λ. Also, $T \sim$ Poisson($n\lambda$).

For part (a), start with T. We have $\mathrm{E}_\lambda(T) = n\lambda$ for all λ so that $E_\lambda(T/n) = \lambda$ for all λ. Hence $\phi(T) = T/n = \overline{X}$ is unbiased for λ. Since T is complete sufficient, therefore $\overline{X}$ is UMVUE for λ.

For part (b), consider the simple unbiased estimator $W = I(X_1 = 0)$ of $\tau(\lambda) = e^{-\lambda}$. Now obtain the Rao-Blackwell estimator

$$\begin{aligned}
\phi(t) &= \mathrm{E}\left[W \mid T = t\right] \\
&= \mathrm{E}(X_1 = 0 \mid T = t) \\
&= P(X_1 = 0 \mid T = t) \\
&= \frac{P(X_1 = 0, T = t)}{P(T = t)} \\
&= \frac{P(X_1 = 0, \sum_{i=2}^{n} X_i = t)}{P(T = t)} \\
&= \frac{P(X_1 = 0)\ P(\sum_{i=2}^{n} X_i = t)}{P(T = t)} \\
&= \frac{e^{-\lambda}\ e^{(n-1)\lambda}\ ((n-1)\lambda)^t/t!}{e^{n\lambda}\ (n\lambda)^t/t!} = \left(\frac{n-1}{n}\right)^t
\end{aligned}$$

Therefore, by Lehmann–Scheffé theorem $\left(\frac{n-1}{n}\right)^T = \left(\frac{n-1}{n}\right)^{\sum_{i=1}^{n} X_i}$ is UMVUE for $e^{-\lambda}$.

Example: Let $X_1, X_2, \ldots, X_n \sim$ iid binomial(k, θ).

Let $\tau(\theta) = P_\theta(X = 1) = k\theta(1-\theta)^{k-1}$.

Find the UMVUE of $\tau(\theta)$, if it exists.

Reading exercise. Example 7.3.24 in the textbook.

Hypothesis Testing

Definition: A **hypothesis** is a statement about a population parameter.

The two complementary hypotheses in a hypothesis testing problem are called the **null hypothesis** and the **alternative hypothesis**, denoted H_0 and H_1 (or sometimes H_a), respectively.

For instance,

$$H_0 : \theta \in \Theta_0$$

$$H_1 : \theta \in \Theta_0^c,$$

where $\Theta = \Theta_0 \cup \Theta_0^c$ is the parameter space.

Definition: A **hypothesis testing procedure**, or **hypothesis test**, is a rule that specifies

(1) for which sample values we accept H_0 as true; and

(2) for which sample values we reject H_0 and accept H_1,

i.e., for what $\underline{x} \in \mathcal{X}$ do we accept or reject H_0.

The subset of $\mathcal{X}$ where we reject H_0 is called the **rejection region** (or **critical region**). The complement is sometimes called the **acceptance region**.

Likelihood Ratio Test

Definition: Recall the **likelihood function**,

$$L(\theta \,|\, \underline{x}) = f(\underline{x} \,|\, \theta) = \prod_{i=1}^{n} f(x_i \,|\, \theta).$$

The **likelihood ratio test (LRT) statistic** for testing

$$H_0 : \theta \in \Theta_0$$

$$H_1 : \theta \in \Theta_0^c$$

is

$$\lambda(\underline{x}) = \frac{\sup_{\Theta_0} L(\theta \,|\, \underline{x})}{\sup_{\Theta} L(\theta \,|\, \underline{x})}.$$

Some texts define the reciprocal as the LRT statistic. We shall follow the convention in the textbook and define the statistic as above.

Definition: A **likelihood ratio test (LRT)** is any test that has a rejection region of the form

$$\{\underline{x} : \lambda(\underline{x}) \leq c\},$$

where $c \in [0, 1]$.

Questions:

(a) How to choose c? Later...

(b) Note that the likelihood ratio test statistic can be viewed as

$$\lambda(\underline{x}) = \frac{L(\hat{\theta}_0 \,|\, \underline{x})}{L(\hat{\theta} \,|\, \underline{x})} = \frac{\text{restricted maximization}}{\text{unrestricted maximization}},$$

where $\hat{\theta}$ is the MLE obtained by maximizing $L(\theta \,|\, \underline{x})$ over the entire parameter space Θ, and $\hat{\theta}_0$ is the MLE obtained by maximizing over the restricted parameter space Θ_0.

Example: Let $X_1, X_2, \ldots, X_n \sim$ iid $\mathrm{N}(\theta, 1)$. We want to test

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0.$$

Find the LRT rejection region.

Under H_0, there is only one value of θ_0. So the restricted maximum in the numerator of LRT statistic $\lambda(\underline{x})$ is simply $L(\theta_0 \mid \underline{x})$.

The unrestricted MLE of θ is $\overline{X}$. So the denominator of $\lambda(\underline{x})$ is $L(\overline{x} \mid \underline{x})$.

So the LRT statistic is

$$\begin{aligned}
\lambda(\underline{x}) &= \frac{(2\pi)^{-n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta_0)^2\right]}{(2\pi)^{-n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - \overline{x})^2\right]} \\
&= \exp\left[-\frac{1}{2}\left\{\sum_{i=1}^{n}(x_i - \theta_0)^2 - \sum_{i=1}^{n}(x_i - \overline{x})^2\right\}\right] \\
&= \exp\left[-\frac{1}{2}\ n(\overline{x} - \theta_0)^2\right]
\end{aligned}$$

The LRT rejection region is $\{\underline{x} : \exp\left[-\frac{1}{2}\ n(\overline{x} - \theta_0)^2\right] < c\}$ for $0 < c < 1$.

Homework

- Method of evaluating estimators: Read p. 342 – 348.
- Hypothesis Tests: Read p. 373 – 376.
- Exercises: TBA.