Lecture 7: Comparison of Estimators

Remark 7.1 Suppose a random sample $X_1, \ldots, X_n$ follows a distribution with a pdf $f(x;\theta)$. Our goal is to estimate the parameter θ or a function of the parameter $\eta = \eta(\theta)$. How do we determine if our estimator is good? One metric is to check how "close" it is to the true value.

Definition 7.1 The **mean squared error (MSE)** for an estimator $\hat{\eta}$ is

$$\mathrm{MSE}(\hat{\eta}, \theta) = E_\theta\Big[(\hat{\eta}(X_1, \ldots, X_n) - \eta(\theta))^2\Big].$$

Other metrics include *mean absolute error*, $E_\theta[|\hat{\eta}(X_1, \ldots, X_n) - \eta(\theta)|]$, or *median absolute error*, $\mathrm{Median}[|\hat{\eta}(X_1, \ldots, X_n) - \eta(\theta)|]$.

Example 7.1 Suppose $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Unif}[0, \theta]$ with $n \geq 2$. We consider the MLE and MME, which are given by

$$\hat{\theta}^{MLE} = X_{(n)} \quad \text{and} \quad \hat{\theta}^{MME} = 2\bar{X}.$$

The mean squared errors for these estimators are follows:

$$\mathrm{MSE}(\hat{\theta}^{MLE}, \theta) = E_\theta[(X_{(n)} - \theta)^2] = \frac{2}{(n+1)(n+2)}\theta^2$$
$$\mathrm{MSE}(\hat{\theta}^{MME}, \theta) = E_\theta[(2\bar{X} - \theta)^2] = \frac{1}{3n}\theta^2$$

Therefore, $\mathrm{MSE}(\hat{\theta}^{MLE}, \theta) \leq \mathrm{MSE}(\hat{\theta}^{MME}, \theta) \quad \forall \theta > 0$.

Example 7.2 Suppose $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\theta, 1)$ where $-\infty < \theta < \infty$. We compare the estimators

$$\hat{\theta}^0 = 0, \quad \hat{\theta}^{MLE} = \bar{X}$$

Calculate the mean squared errors (MSE) for these estimators as

$$MSE(\hat{\theta}^0, \theta) = E_\theta[(0 - \theta)^2] = \theta^2$$
$$MSE(\hat{\theta}^{MLE}, \theta) = E_\theta[(\bar{X} - \theta)^2] = \frac{1}{n}$$

Therefore, if $|\theta| < 1/\sqrt{n}$, the estimator $\hat{\theta}^0 = 0$, which does not use the sample observations, has a smaller MSE than the maximum likelihood estimator $\hat{\theta}^{MLE} = \bar{X}$. Thus, based on the MSE criterion, the maximum likelihood estimator $\bar{X}$ cannot *always* be considered better than $\hat{\theta}^0 = 0$. However, considering the rate at which the MSE approaches zero as the sample size increases, $\hat{\theta}^{MLE} = \bar{X}$ is a better estimator than $\hat{\theta}^0 = 0$.

Remark 7.2 As seen in Examples 7.1 and 7.2, there are many cases where it is difficult to compare two estimators based on just mean squared error (MSE). This is because an estimator's "superiority" depends on the parameter value. Therefore, we want to compare estimators based on criteria that do *not* depend on the value of the parameter. Among such comparison criteria, the most common are the maximum mean squared error (Maximum MSE).

Theorem 7.1 *Given that* $Bias(\hat{\eta}, \eta) = \mathbb{E}_\eta(\hat{\eta}) - \eta$*, it can be shown that*

$$MSE(\hat{\eta}, \eta) = Var(\hat{\eta}) + Bias(\hat{\eta}, \eta)^2$$

Proof. In-class exercise. □

Remark 7.3 When comparing estimators based on mean squared error (MSE), it is useful to restrict the comparison to estimators with certain properties. Particularly, an estimator $\hat{\eta}(X_1, \ldots, X_n)$ for η for all such that

$$\mathbb{E}_\eta[\hat{\eta}(X_1, \ldots, X_n)] = \eta.$$

We call $\hat{\eta}$ an **unbiased estimator**, often written at $\hat{\eta}^{UE}$. In this special case for unbiased estimators $\hat{\eta}^{UE}$, we know

$$\text{MSE}(\hat{\eta}^{UE}, \eta) = \mathbb{E}_\eta[(\hat{\eta}^{UE} - \eta)^2] = \text{Var}_\eta(\hat{\eta}^{UE}).$$

Thus, the estimator that minimizes the variance among unbiased estimators is optimal with respect to this criterion. If there exists an estimator that has the minimum variance among all unbiased estimators for all values of the parameter η, it is called the **uniformly minimum variance unbiased estimator** (UMVUE).

Definition 7.2 We say an estimator $\hat{\eta}^*$ is **UMVUE** of η if

(i) it is an unbiased estimator

(ii) for any unbiased estimator $\hat{\eta}^{UE}$, $\text{Var}_\eta(\hat{\eta}^*) \leq \text{Var}_\eta(\hat{\eta}^{UE})$.

Theorem 7.2 ***(Rao-Blackwell Theorem)*** *Let* $f(x;\theta)$, $\theta \in \Omega$, *be the probability density function of a random sample* $X_1, \ldots, X_n$ *from a population. For a sufficient statistic* $Y = u(X_1, \ldots, X_n)$ *and an estimator* $\hat{\eta}$ *for a parameter* $\eta(\theta)$*, define*

$$\hat{\eta}^{RB}(Y) = \mathbb{E}[\hat{\eta}(X_1, \ldots, X_n)|Y]$$

as the ***Rao-Blackwell estimator****. Then, the following inequality holds:*

$$\text{MSE}(\hat{\eta}^{RB}, \theta) = \mathbb{E}_\theta[(\hat{\eta}^{RB}(Y) - \eta(\theta))^2] \leq \mathbb{E}_\theta[(\hat{\eta}(X_1, \ldots, X_n) - \eta(\theta))^2] = \text{MSE}(\hat{\eta}, \theta), \quad \forall \theta \in \Omega.$$

Proof. Let $Y = u(X_1, \dots, X_n)$ be a sufficient statistic for $\theta \in \Omega$. Since Y is sufficient, the conditional distribution of $X_1, \dots, X_n$ given Y does not depend on θ. Therefore,

$$\mathbb{E}_\theta[\hat{\eta}(X_1, \dots, X_n)|Y] \text{ does not depend on } \theta \in \Omega.$$

Hence,

$$\hat{\eta}^{RB}(Y) = \mathbb{E}_\theta[\hat{\eta}(X_1, \dots, X_n)|Y] \overset{\theta \in \Omega}{=} \mathbb{E}[\hat{\eta}(X_1, \dots, X_n)|Y].$$

Applying the property of conditional expectation

$$\mathbb{E}[\mathbb{E}(V|W)] = \mathbb{E}(V), \quad \mathrm{Var}(V) = \mathbb{E}[\mathrm{Var}(V|W)] + \mathrm{Var}(\mathbb{E}[V|W])$$

to $V = \hat{\eta}$, and $W = Y$, we have $\mathbb{E}(\hat{\eta}) = \mathbb{E}(\mathbb{E}(\hat{\eta}|Y)) = \mathbb{E}(\hat{\eta}^{RB})$, i.e., the same expectation value, and

$$\mathrm{Var}(\hat{\eta}^{RB}) = \mathrm{Var}(\mathbb{E}[\hat{\eta}|Y]) \le \mathrm{Var}(\mathbb{E}[\hat{\eta}|Y]) + \mathbb{E}[\mathrm{Var}(\hat{\eta}|Y)] = \mathrm{Var}(\hat{\eta}).$$

Thus, we have

$$\begin{aligned} MSE(\hat{\eta}^{RB}, \theta) = E_\theta[(\hat{\eta}^{RB} - \eta(\theta))^2] &= \mathrm{Var}(\hat{\eta}^{RB}) + (E_\theta(\hat{\eta}^{RB}) - \eta(\theta))^2 \\ &\le \mathrm{Var}(\hat{\eta}) + (E_\theta(\hat{\eta}) - \eta(\theta))^2 \\ &= MSE(\hat{\eta}, \theta), \quad \forall \theta \in \Omega. \end{aligned}$$

□

Remark 7.4 From the Theorem 7.2, we can obtain an improved estimator by taking the conditional expectation given a sufficient statistic. In particular, when the given estimator is unbiased, and if the sufficient statistic has an additional property (namely, **completeness**), it is possible to find the UMVUE through this method.

Definition 7.3 Let $f(x;\theta), \theta \in \Omega$ be a pdf of a population. For a random sample $X_1, \dots, X_n$, a sufficient statistic $Y = u(X_1, \dots, X_n)$ is a **complete sufficient statistic** for $\theta \in \Omega$ if:

$$E_\theta[g(Y)] = 0 \implies g(Y) = 0 \quad \text{for all} \quad \theta \in \Omega.$$

Theorem 7.3 *Let $f(x;\theta)$, $\theta \in \Omega$, be a probability density function of a population. For a random sample $X_1, \dots, X_n$, suppose that the statistic $Y = u(X_1, \dots, X_n)$ is **complete and sufficient**.*

(a) If $\hat{\eta}^{UE} = \hat{\eta}^{UE}(X_1, \dots, X_n)$ is an unbiased estimator of $\eta(\theta)$, then the Rao-Blackwell estimator, $\hat{\eta}^{RB}(Y) = \mathbb{E}[\hat{\eta}^{UE}(X_1, \dots, X_n)|Y]$, is an unbiased estimator of $\eta(\theta)$ and is the UMVUE of $\eta(\theta)$.

(b) If there exists a function δ such that $\delta(Y) = \hat{\eta}(Y)$ is an unbiased estimator of $\eta(\theta)$, then $\delta(Y)$ is the UMVUE of $\eta(\theta)$.

Theorem 7.4 ***(Lehmann-Sheffe Theorem)***: *A complete sufficient statistic is unique, up to one-to-one transformation, and therefore the UMVUE is unique, up to one-to-one transformation (refers to Theorem 7.3(b)).*

Theorem 7.5 *Let $f(x;\boldsymbol{\eta})$ be a probability density function given by the form*

$$f(x;\boldsymbol{\eta}) = \exp\left(\sum_{j=1}^{k} \eta_j T_j(x) - A(\eta) + S(x)\right), \quad x \in \mathcal{X}, \quad \boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)^T \in \Omega$$

where the conditions of the exponential family hold. Suppose that the joint probability distribution of a random sample $X_1, \ldots, X_n$ belongs to this family. Then, the statistic

$$\sum_{i=1}^{n} \boldsymbol{T}(X_i) = \left(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i)\right)^T$$

is a ***complete sufficient statistic*** *for $\boldsymbol{\eta} \in \Omega$. In the case of applying the theorem more generally, the parameter vector is expressed as a function*

$$\boldsymbol{\eta} = g(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), \ldots, g_k(\boldsymbol{\theta}))^T, \quad \boldsymbol{\theta} \in \Theta$$

and the pdf is then given by:

$$f(x;\boldsymbol{\theta}) = \exp\left(\sum_{j=1}^{k} g_j(\boldsymbol{\theta}) T_j(x) - A(g(\boldsymbol{\theta})) + S(x)\right), \quad x \in \mathcal{X}, \quad \theta \in \Theta,$$

then $\sum_{i=1}^{n} \boldsymbol{T}(X_i)$ is a complete sufficient statistic for $\boldsymbol{\theta} \in \Theta$.

Example 7.3 Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where $-\infty < \mu < \infty$ and $\sigma^2 > 0$ with pdf

$$f(x;\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Find the UMVUE for μ and σ^2.

Solution. We can rewrite the pdf as

$$f(x;\mu,\sigma^2) = \exp\left(\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log\big(2\pi\sigma^2\big)\right)$$

which is in the form of an exponential family. We check the conditions of Definition 6.2: (i) the support of the distribution does not depend on the parameters, (ii) the parameter space is

$$\left\{\left(\frac{\mu}{\sigma^2}, \frac{-1}{2\sigma^2}\right) : -\infty < \mu < +\infty,\ \sigma^2 > 0\right\} = (-\infty, +\infty) \times (-\infty, 0)$$

forms an open rectangle, and (iii) $T_1(x)$ and $T_2(x)$ are not constants and their linear combination is not constant. Therefore, by Theorem 7.5, the following statistic

$$\left(T_1 = \sum_{i=1}^n X_i,\ T_2 = \sum_{i=1}^n X_i^2\right)^T$$

is a complete sufficient statistic for $(\mu, \sigma^2) \in (-\infty, \infty) \times (0, \infty)$.

However, these estimators are not *unbiased* for either μ or σ^2. Therefore, we need to apply the Lehmann-Sheffe Theorem to find a one-to-one transformation of T_1 and T_2 that is unbiased for μ and σ^2, and then we know that such an estimator is the UMVUE.

First, recall the well-known unbiased estimators of μ and σ^2, $\frac{1}{n}\sum_{i=1}^n X_i$ and $\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$, respectively. If we can find a function g, such that $g(T_1, T_2) = (\frac{1}{n}\sum_{i=1}^n X_i,\ \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2)$, then we are done. In particular, we see that

$$g(T_1, T_2) = \left(\frac{1}{n}T_1,\ \frac{1}{n-1}\left\{T_2 - \frac{1}{n}T_1^2\right\}\right) = \left(\frac{1}{n}\sum_{i=1}^n X_i,\ \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2\right)$$

Therefore, the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ are the UMVUEs of μ and σ^2, respectively. ■

Example 7.4 Let $X_1, \ldots, X_n \sim \text{Unif}[0, \theta]$, where $\theta > 0$. Find the UMVUE directly.

Solution. We know that $Y_n = \max\{X_i\}_{1 \le i \le n}$ is sufficient for θ by the factorization theorem. The pdf for Y_n is given by

$$f(y;\theta) = \frac{n}{\theta}\left(\frac{y}{\theta}\right)^{n-1} \cdot \mathbf{1}\{y \in [0, \theta]\}$$

We need to now show it is complete. From the Fundamental Theorem of Calculus, we know that

$$\mathbb{E}_\theta[g(Y_n)] = 0 \implies G(\theta) = \int_0^\theta g(y)\frac{n}{\theta^n}y^{n-1}\,dy = 0 \implies G'(\theta) = g(\theta)\frac{n}{\theta^n}\theta^{n-1} = 0.$$

for all $\theta > 0$. Thus, $g(y) = 0$ for all y, hence $g(Y) = 0$. This proves that $Y_n = \max\{X_i\}$ is a complete sufficient statistic for θ. Next, we can compute:

$$\mathbb{E}_\theta(Y_n) = \int_0^\theta y\frac{n}{\theta^n}y^{n-1}\,dy = \theta \cdot \frac{n}{n+1}$$

Thus, $\hat{\theta} = \frac{n+1}{n}Y_n = \frac{n+1}{n}\max\{X_i\}$ is the UMVUE of θ. ■

Example 7.5 Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$, $\theta \in (0,1)$, and let $Y = \sum_{i=1}^n X_i$, which is complete and sufficient statistic for θ. Consider an unbiased estimator for $\eta = \theta(1-\theta)$

$$\hat{\eta}_0^{UE} = X_1(1 - X_2), \quad \eta = \mathbb{E}_\theta[X_1(1 - X_2)] = \theta(1-\theta)$$

Thus, $\hat{\eta}_0^{UE}$ is an unbiased estimator of $\eta = \theta(1-\theta)$, and the Rao-Blackwell estimator $\hat{\eta}_0^{RB}(Y)$ is

$$\hat{\eta}_0^{RB}(Y) = \mathbb{E}[X_1(1-X_2)|Y]$$

The specific form of $\hat{\eta}_0^{RB}(Y)$ is derived as follows. (Recall $\mathbb{E}(X) = \mathbb{E}(\mathbf{1}\{X=1\}) = P(X=1)$ for a Bernoulli random variable X.)

$$\mathbb{E}[X_1(1-X_2) \mid Y_n = y] = \mathbb{E}[\mathbf{1}\{X=1, X_2=0\} \mid Y_n = y] = P(X_1 = 1, X_2 = 0 \mid Y_n = y)$$

This can be computed logically using combinatorics. What is the probability $X_1 = 1$ and $X_2 = 0$ given my sum $X_1 + \ldots X_n = y$. Well, there are n total random variables and y of them need to be 1. Therefore, the denominator is $\binom{n}{y}$. Now, how many possible ways can I assign the random variables $X_3, \ldots, X_n$ with the knowledge that $X_1 = 1$ and $X_2 = 0$? The answer is $\binom{n-2}{y-1}$. Therefore,

$$P(X_1 = 1, X_2 = 0 \mid Y_n = y) = \frac{\binom{n-2}{y-1}}{\binom{n}{y}} = \frac{y(n-y)}{n(n-1)}$$

Thus, the Rao-Blackwell estimator is

$$\hat{\eta}_0^{RB}(Y_n) = \frac{Y_n(n-Y_n)}{n(n-1)}$$

and by Theorem 7.3 is the UMVUE for $\theta(1-\theta)$.

Theorem 7.6 ***(Cramér-Rao Lower Bound):*** *Let $X_1, \ldots, X_n \overset{iid}{\sim} f(x \mid \theta)$, $\theta \in \Omega$. Assume the regularity conditions (R0) – (R4) from Remark 5.9 (Lecture 5) hold. Let $Y = u(X_1, \ldots, X_n)$ be a statistic with mean $\mathbb{E}(Y) = \mathbb{E}[u(X_1, \ldots, X_n)] = k(\theta)$. Then,*

$$\mathrm{Var}(Y) \geq \frac{[k'(\theta)]^2}{n \cdot I(\theta)}$$

where $I(\theta) = \mathrm{Var}(S(\theta)) = \mathbb{E}_\theta\left[-\frac{\partial^2 \log f(X|\theta)}{\partial \theta^2}\right]$ is the Fisher Information.

Definition 7.4 Let Y be an unbiased estimator of a parameter θ. The statistic Y is called an **efficient estimator** of θ if and only if the variance of Y attains the Cramér-Rao lower bound (CRLB).

Example 7.6 Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Poisson}(\theta)$, $\theta > 0$, with pmf given by

$$f_X(x) = \frac{\theta^x \exp(-\theta)}{x!}, \quad \mathbb{E}(X) = \mathrm{Var}(X) = \theta$$

Suppose our estimator, Y, is defined as $Y = X_1$. Determine whether or not it is an efficient estimator of θ.

Solution. First, let's compute the CRLB. First, we see $\mathbb{E}(Y) = \mathbb{E}(X_1) = \theta$, therefore $k'(\theta) = 1$. Next, we need to compute the Fisher Information:

$$\begin{aligned}
I(\theta) = \mathrm{Var}(S(\theta)) &= \mathbb{E}_\theta\left[-\frac{\partial^2 \log f(X \mid \theta)}{\partial \theta^2}\right] \\
&= \mathbb{E}_\theta\left[-\frac{\partial^2}{\partial \theta^2}(x\log(\theta) - \theta - \log(x!))\right] \\
&= \mathbb{E}_\theta\left[-\frac{\partial}{\partial \theta}\left(\frac{x}{\theta} - 1\right)\right] \\
&= \mathbb{E}_\theta\left[\frac{x}{\theta^2}\right] \\
&= \frac{1}{\theta}
\end{aligned}$$

The CRLB is $\frac{\theta}{n}$. In turn, $\mathrm{Var}(Y) = \mathrm{Var}(X_1) = \theta$ means X_1 is **not** and efficient estimator of θ. ■