Statistical Inference and Multivariate Analysis (MA324)

LECTURE SLIDES

Lecture 15

Criteria to Compare Estimators

Indian Institute of Technology Guwahati

Jan-May 2023

Criteria to Compare Estimators:

- We have considered two different methods of estimation. Now, a natural question is to ask: Which method provide a **better estimator in a particular situation**? Or in other words, if we have multiple estimator for an unknown parameter, then **which one is "best"**?
- To find the best estimator, we need to **consider error that we may commit** if we use an estimator to estimate a parameter. We should choose an estimator with least error.
- As an estimator is a function of a RS, the **error will vary with realization of the RS**. Therefore, to have a **meaningful measure of an error**, we should consider **average of error over all possible realizations** of RS. There are different measures of error.
- We will discuss some of them (**Unbiasedness, Variance, and Mean Squared Error**) here along with some desirable properties of an estimator based on different measures of the error. Here, we will assume that $\tau : \Theta \to \mathbb{R}$ and we are interested to estimate $\tau(\boldsymbol{\theta})$.

Unbiased Estimator

Def: [**Unbiased Estimator**] A real valued estimator T is said to be an unbiased estimator (UE) of a parametric function $\tau(\boldsymbol{\theta})$ if $E_{\boldsymbol{\theta}}(T) = \tau(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$. Here it is assumed that $E_{\boldsymbol{\theta}}(T)$ exists. An estimator is called biased if it is not unbiased.

Note that $E_{\boldsymbol{\theta}}(T) = \tau(\boldsymbol{\theta})$ implies $E(T - \tau(\boldsymbol{\theta})) = 0$. Thus, unbiasedness tells us that on an average, there is no error. The average is taken over all possible realizations of the RS.

Def: [**Bias**] Bias of a real valued statistic T as an estimator of $\tau(\boldsymbol{\theta})$ is defined by

$$B_T(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(T) - \tau(\boldsymbol{\theta}) \quad \text{for all} \quad \theta \in \Theta.$$

Examples:

Example 1: Let $X_1, \ldots, X_n$ be a RS from a population with mean $\mu \in \mathbb{R}$. Then $\bar{X}$ is an unbiased estimator for μ. To see it, notice that, for all $\mu \in \mathbb{R}$,

$$E\left(\bar{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu.$$

Example 2: Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} U(0, \theta)$, $\theta > 0$. We saw that the MLE of θ is $X_{(n)}$. Now, we want to check if $X_{(n)}$ is unbiased or not.

Example 3: Let $X_1, \ldots, X_n$ be a RS from a population with mean $\mu \in \mathbb{R}$. Define $T_1 = X_1$, $T_2 = \frac{1}{2}\left(X_1 + X_2\right) \ldots, T_n = \bar{X}$. It is easy to verify that $E(T_i) = \mu$ for all $\mu \in \mathbb{R}$ and for all $i = 1, 2, \ldots, n$. Thus, T_i is an unbiased estimator of μ for all $i = 1, 2, \ldots, n$. This example shows that there may be **more than one unbiased estimator** for a parametric function. **Which one** should we **prefer**?

Mean Square Error

Def: [**Mean Square Error**] Mean square error (MSE) of a real valued statistic T as an estimator of $\tau(\boldsymbol{\theta})$ is defined by

$$MSE_T(\boldsymbol{\theta}) = E\left[(T - \tau(\boldsymbol{\theta}))^2\right],$$

provided the expectation exists.

Note that MSE gives us average square distance between the estimator and the true value of the parametric function. Hence, an estimator with smaller value of MSE is preferred.

Theorem

$MSE_T(\boldsymbol{\theta}) = Var_{\boldsymbol{\theta}}(T) + (B_T(\boldsymbol{\theta}))^2$.

Corollary

If T is an UE for θ, then $MSE_T(\boldsymbol{\theta}) = Var_{\boldsymbol{\theta}}(T)$.

Example 4: [Continuation of Example 3] Let $X_1, \ldots, X_n$ be a RS from a population with mean $\mu \in \mathbb{R}$ and finite variance σ^2. Let $T_1 = X_1$ and $T_i = \frac{1}{i}\sum_{j=1}^{i} X_j$ for $i = 2, 3, \ldots, n$. Then T_i is an UE for μ for all $i = 1, 2, \ldots, n$. Which one should we prefer? Note that

$$MSE(T_i) = Var(T_i) = \frac{\sigma^2}{i}$$

for $i = 1, 2, \ldots, n$. Hence, T_n has smallest MSE among these estimators and we will prefer T_n over other estimators. Note that only T_n is based on all observations.

Best Unbiased Estimator

- We are interested to find the "**best**" estimator among all UEs of a parametric function.
- There are situations where a parametric function **does not have a UE**. In such situations, looking for best unbiased estimator makes no sense.
- Therefore, we will only consider U-estimable (unbiased) parametric functions. How should we **compare the performance** of two UEs?
- We will use **MSE to compare them**. Recall that MSE of an UE is same as the variance of the UE. Thus, we have following definition.

Uniformly Minimum Variance Unbiased Estimator (UMVUE)

Def: [**Uniformly Minimum Variance Unbiased Estimator**] Let the set of all UEs of a parametric function $\tau(\boldsymbol{\theta})$ be denoted by $\mathcal{C}$, which is assumed to be non-empty. An estimator $T \in \mathcal{C}$ is called a uniformly minimum variance unbiased estimator (UMVUE) of $\tau(\boldsymbol{\theta})$ if for all estimator $T^* \in \mathcal{C}$,

$$Var_{\boldsymbol{\theta}}(T) \leq Var_{\boldsymbol{\theta}}(T^*) \quad \text{for all } \boldsymbol{\theta} \in \Theta.$$

Theorem

Let $X_1, X_2, \ldots, X_n$ be a RS from common PMF/PDF $f(\cdot, \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$. Let T be a real valued estimator with $Var_{\boldsymbol{\theta}}(T) < \infty$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Also assume that $\mathcal{U}$ be the set of all unbiased estimators of zero such that $Var_{\boldsymbol{\theta}}(U) < \infty$ for all $U \in \mathcal{U}$ and all $\boldsymbol{\theta} \in \Theta$. Then, a necessary and sufficient condition for T to be a UMVUE of its expectation $\tau(\boldsymbol{\theta})$ is that

$$Cov_{\boldsymbol{\theta}}(T, U) = E_{\boldsymbol{\theta}}(TU) = 0 \quad \textit{for all } U \in \mathcal{U} \textit{ and for all } \boldsymbol{\theta} \in \Theta.$$

Uniformly Minimum Variance Unbiased Estimator

Here, we will discuss the **methods of finding UMVUE** of a parametric function. We will discuss mainly one of the two methods. First method is based on **Cramer-Rao inequality** (and the second one is based on Lehmann-Scheffe theorem).

We assume that $X_1, X_2, \ldots, X_n$ is a RS from a common PMF/PDF $f(\cdot, \theta)$, where $\theta \in \Theta \subset \mathbb{R}$.

Theorem (Cramer-Rao Inequality)

Suppose that T is an unbiased estimator of a real valued parametric function $\tau(\theta)$. Assume that $\frac{d}{d\theta}\tau(\theta)$, denoted by $\tau'(\theta)$, is finite for all $\theta \in \Theta$. Then, for all $\theta \in \Theta$, under the assumptions (regularity conditions) 1 and 2, we have

$$Var_\theta(T) \geq \frac{\left(\tau'(\theta)\right)^2}{n\mathcal{I}_{X_1}(\theta)}.$$

The expression on the right hand side of the inequality is call Cramer-Rao lower bound (CRLB).

Remark: The Cramer-Rao inequality provides a **lower bound for variance of an UE** of the parametric function $\tau(\theta)$. Thus, if one can find an UE T of $\tau(\theta)$ such that $Var(T)$ equals CRLB for all $\theta \in \Theta$, **then T is the UMVUE** of $\tau(\theta)$. However, note that if there is an UE T of $\tau(\theta)$ such that the variance of T is greater than CRLB, we cannot decide if T is UMVUE of $\tau(\theta)$.

Example 5: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Poi(\lambda)$, where $\lambda > 0$ is unknown parameter. Let us consider $\tau(\lambda) = \lambda$. The Fisher information is $\mathcal{I}_{X_1}(\lambda) = \frac{1}{\lambda}$. Thus, CRLB is

$$\frac{\left(\tau'(\theta)\right)^2}{n\mathcal{I}_{X_1}(\theta)} = \frac{\lambda}{n}.$$

On the other hand, $\bar{X}$ is an UE of λ. Note that $Var(\bar{X}) = \frac{\lambda}{n}$. Thus, variance of $\bar{X}$ coincide with CRLB. Therefore, $\bar{X}$ is UMVUE of λ.

Example 6: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ is unknown parameter and $\sigma > 0$ is known. Consider $\tau(\mu) = \mu$. Then, $\bar{X}$ is an UE for μ. In this case Fisher information is $\mathcal{I}_{X_1}(\mu) = \frac{1}{\sigma^2}$. Therefore, CRLB is $\frac{\sigma^2}{n}$, which is same as variance of $\bar{X}$. Thus, $\bar{X}$ is the MUVUE of μ.

Consistent Estimator

Def: [**Consistent Estimator**] Let T_n be an estimator based on a RS of size n. The estimator T_n is said to be consistent for θ if the sequence of RVs $\{T_n : n \geq 1\}$ converges to θ in probability for all $\theta \in \Theta$, *i.e.*, if for all $\varepsilon > 0$ and all $\theta \in \Theta$,

$$\lim_{n\to\infty} P\left(|T_n - \theta| \leq \varepsilon\right) = 1.$$

Remark: Consistency says us that for a sample with reasonably large size, T_n is close to the true value of parameter with high probability.

Example 7: Let $X_1, X_2 \ldots, X_n$ be a RS from a population with mean $\mu \in \mathbb{R}$. Then, using WLLN, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a consistent estimator for μ.

Consistency of MLE:

Theorem (Consistency of MLE)

Let $X_1, X_2, \ldots, X_n$ be a RS from the population having PMF/PDF $f(x;\,\theta)$, where $\theta \in \Theta \subseteq \mathbb{R}$. Consider the following assumptions.

1. $\frac{\partial}{\partial\theta} \ln f(x;\,\theta)$, $\frac{\partial^2}{\partial\theta^2} \ln f(x;\,\theta)$, $\frac{\partial^3}{\partial\theta^3} \ln f(x;\,\theta)$ *are finite for all* $x \in \mathbb{R}$ *and for all* $\theta \in \Theta$.
2. $\displaystyle\int_{-\infty}^{+\infty} \frac{\partial}{\partial\theta} f(x;\,\theta)dx = 0$, $\displaystyle\int_{-\infty}^{+\infty} \frac{\partial^2}{\partial\theta^2} f(x;\,\theta)dx = 0$, *and* $\displaystyle\int_{-\infty}^{+\infty} \left\{\frac{\partial}{\partial\theta} f(x;\,\theta)\right\}^2 dx > 0$ *for all* $\theta \in \Theta$.
3. *For all* $\theta \in \Theta$, $\left|\dfrac{\partial^3}{\partial\theta^3} \ln f(x;\,\theta)\right| < a(x)$, *where* $E\left(a(X_1)\right) < b$ *for a constant* b *which is independent of* θ.

Under these three assumptions, the likelihood equation has a solution denoted by $\widehat{\theta}_n(\boldsymbol{x})$, such that $\widehat{\theta}_n(\boldsymbol{X})$ is consistent estimator of θ.

Consistency of MLE:

Theorem (Asymptotic Normality of MLE)

Under the three assumptions of the previous Theorem,

$$\sqrt{n\mathcal{I}_{X_1}(\theta)}\left(\widehat{\theta}_n(\boldsymbol{X}) - \theta\right) \to Z$$

in distribution, where $Z \sim N(0,\, 1)$, and $\mathcal{I}_{X_1}(\theta)$ is Fisher information based on a RS of size one.

Example 8: Let $X_1,\, X_2,\, \ldots \overset{i.i.d.}{\sim} Bernoulli(p)$. The MLE of p based on a sample of size n is $\widehat{p}_n = \bar{X}_n$ and $\mathcal{I}_{X_1}(p) = \frac{1}{p(1-p)}$. Using above theorems, $\widehat{p}_n$ is consistent for p and $\sqrt{n}(\widehat{p}_n - p) \to N(0,\, p(1-p))$ in distribution.