Lecture 07: Evaluation of Point Estimators

Mathematical Statistics II, MATH 60062/70062

Thursday February 10, 2022

Reference: Casella & Berger, 7.3.1-7.3.2

Mean squared error

The **mean squared error (MSE)** of a point estimator $W = W(\boldsymbol{X})$ of a parameter θ is

$$\begin{aligned}\mathrm{MSE}_\theta(W) &= E_\theta(W-\theta)^2\\ &= \mathrm{Var}_\theta(W-\theta) + (E_\theta(W-\theta))^2\\ &= \mathrm{Var}_\theta(W) + \mathrm{Bias}_\theta^2(W),\end{aligned}$$

where $\mathrm{Bias}_\theta(W) = E_\theta(W) - \theta$ is the **bias** of W, the difference between the expected value of W and θ. If $\mathrm{Bias}_\theta(W) = 0$ for all θ, then W is called an **unbiased** estimator.

MSE incorporates two components:

- **Precision**, measured by $\mathrm{Var}_\theta(W)$.
- **Accuracy**, measured by $\mathrm{Bias}_\theta(W)$.

Normal variance estimators

Suppose that $X_1, \dots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$, where $-\infty < \mu < \infty$ and $\sigma^2 > 0$. Set $\boldsymbol{\theta} = (\mu, \sigma^2)$.

- The sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

  is an unbiased estimator for σ^2, where $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, $E_{\boldsymbol{\theta}}(S^2) = \sigma^2$ for all σ^2 and $\mathrm{Var}_{\boldsymbol{\theta}}(S^2) = 2\sigma^4/(n-1)$.

- The MLE of σ^2

$$S_{\mathrm{B}}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{n-1}{n} S^2$$

  is a biased estimator for σ^2, $E_{\boldsymbol{\theta}}(S_{\mathrm{B}}^2) = \sigma^2 (n-1)/n$.

MSE of Normal variance estimators

- MSE of S^2

$$\mathrm{MSE}_{\boldsymbol{\theta}}(S^2) = \mathrm{Var}_{\boldsymbol{\theta}}(S^2) = \frac{2\sigma^4}{n-1}$$

- MSE of S^2_{B}

$$\begin{aligned}\mathrm{MSE}_{\boldsymbol{\theta}}(S^2_{\mathrm{B}}) &= \mathrm{Var}_{\boldsymbol{\theta}}(S^2_{\mathrm{B}}) + \mathrm{Bias}^2_{\boldsymbol{\theta}}(S^2_{\mathrm{B}}) \\ &= \left(\frac{n-1}{n}\right)^2 \mathrm{Var}_{\boldsymbol{\theta}}(S^2) + \left(\frac{-\sigma^2}{n}\right)^2 \\ &= \left(\frac{2n-1}{n^2}\right)\sigma^4\end{aligned}$$

The biased MLE S^2_{B} has smaller MSE than S^2, since $2/(n-1) > (2n-1)/n^2$ for all $n \geq 2$.

MSE of Bernoulli/Binomial estimators

Suppose that $X_1, \ldots, X_n$ are iid $\mathrm{Bern}(\theta)$, where the prior distribution on θ is $\mathrm{Beta}(a, b)$.

- The MLE of θ, $\hat{\theta} = \bar{X}$, is an unbiased estimator of θ, and its MSE is

$$\mathrm{MSE}_\theta(\hat{\theta}) = \mathrm{Var}_\theta(\bar{X}) = \frac{\theta(1-\theta)}{n}.$$

- The Bayes estimator is $\hat{\theta}_{\mathrm{B}} = (T + a)/(n + a + b)$, where $T = \sum_{i=1}^n X_i$. The MSE of $\hat{\theta}_{\mathrm{B}}$ is

$$\begin{aligned}\mathrm{MSE}_\theta(\hat{\theta}_{\mathrm{B}}) &= \mathrm{Var}_\theta\left(\frac{T+a}{n+a+b}\right) + \mathrm{Bias}_\theta^2\left(\frac{T+a}{n+a+b}\right) \\ &= \frac{n\theta(1-\theta)}{(n+a+b)^2} + \left(\frac{n\theta+a}{n+a+b} - \theta\right)^2.\end{aligned}$$

MSE of Bernoulli/Binomial estimators

Suppose that $X_1, \ldots, X_n$ are iid $\mathrm{Bern}(\theta)$, where the prior distribution on θ is $\mathrm{Beta}(a, b)$.

- The MLE of θ, $\hat{\theta} = \bar{X}$, is an unbiased estimator of θ, and its MSE is

$$\mathrm{MSE}_\theta(\hat{\theta}) = \mathrm{Var}_\theta(\bar{X}) = \frac{\theta(1-\theta)}{n}.$$

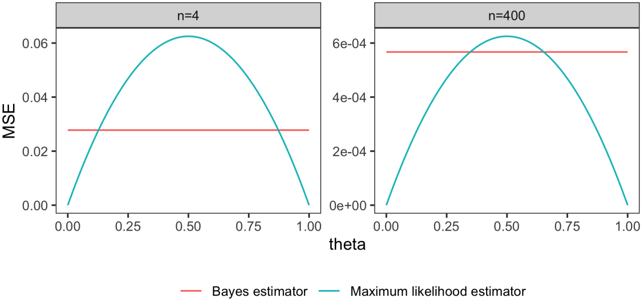
- The Bayes estimator is $\hat{\theta}_\mathrm{B} = (T+a)/(n+a+b)$, where $T = \sum_{i=1}^n X_i$. The MSE of $\hat{\theta}_\mathrm{B}$ is

$$\begin{aligned}\mathrm{MSE}_\theta(\hat{\theta}_\mathrm{B}) &= \mathrm{Var}_\theta\left(\frac{T+a}{n+a+b}\right) + \mathrm{Bias}^2_\theta\left(\frac{T+a}{n+a+b}\right) \\ &= \frac{n\theta(1-\theta)}{(n+a+b)^2} + \left(\frac{n\theta+a}{n+a+b} - \theta\right)^2.\end{aligned}$$

In the absence of good prior information, we choose $a = b = \sqrt{n/4}$ to make $\mathrm{MSE}_\theta(\hat{\theta}_\mathrm{B})$ constant,

$$\hat{\theta}_\mathrm{B} = \frac{T + \sqrt{n/4}}{n + \sqrt{n}}, \quad \mathrm{MSE}_\theta(\hat{\theta}_\mathrm{B}) = \frac{n}{4(n+\sqrt{n})^2}.$$

MSE of Bernoulli/Binomial estimators



- For small n, the Bayes estimator $\hat{\theta}_{\mathrm{B}}$ is a better choice (unless there is a strong belief that θ is near 0 or 1).
- For large n, the MLE $\hat{\theta}$ is a better choice (unless there is a strong belief that θ is close to $1/2$).

Best unbiased estimators

In most cases, there is no one "best MSE" estimator. This is because the class of all estimators is too large a class.

Consider a nonsense estimator $\hat{\theta}_{\mathrm{N}} = 0.21$.

- If $\theta = 0.21$, then $\hat{\theta}_{\mathrm{N}}$ is the best in terms of MSE, i.e., $\mathrm{MSE}_\theta(\hat{\theta}_{\mathrm{N}}) = 0$.
- If $\theta \neq 0.21$, then $\hat{\theta}_{\mathrm{N}}$ can be a terrible estimator.

Best unbiased estimators

In most cases, there is no one "best MSE" estimator. This is because the class of all estimators is too large a class.

Consider a nonsense estimator $\hat{\theta}_{\mathrm{N}} = 0.21$.

- If $\theta = 0.21$, then $\hat{\theta}_{\mathrm{N}}$ is the best in terms of MSE, i.e., $\mathrm{MSE}_{\theta}(\hat{\theta}_{\mathrm{N}}) = 0$.
- If $\theta \neq 0.21$, then $\hat{\theta}_{\mathrm{N}}$ can be a terrible estimator.

One way to make the problem of finding a "best" estimator tractable is to limit the class of estimators (e.g., to consider only unbiased estimators).

A *best unbiased estimator* in terms of MSE is essentially an unbiased estimator whose variance is smaller than any other unbiased estimator.

Uniformly minimum-variance unbiased estimator (UMVUE)

An estimator $W^* = W^*(\boldsymbol{X})$ is a **uniformly minimum-variance unbiased estimator (UMVUE)** of $\tau(\theta)$ if

1. $E_\theta(W^*) = \tau(\theta)$ for all $\theta \in \Theta$.
2. $\mathrm{Var}_\theta(W^*) \leq \mathrm{Var}_\theta(W)$ for all $\theta \in \Theta$, where W is any other unbiased estimator of $\tau(\theta)$.

Uniformly minimum-variance unbiased estimator (UMVUE)

An estimator $W^* = W^*(\boldsymbol{X})$ is a **uniformly minimum-variance unbiased estimator (UMVUE)** of $\tau(\theta)$ if

1. $E_\theta(W^*) = \tau(\theta)$ for all $\theta \in \Theta$.
2. $\mathrm{Var}_\theta(W^*) \leq \mathrm{Var}_\theta(W)$ for all $\theta \in \Theta$, where W is any other unbiased estimator of $\tau(\theta)$.

Note: UMVUEs may not exist. As we will show later, if a UMVUE does exist, then it's *unique*.

Finding UMVUE

Using the definition of UMVUE to find one is difficult. As we have to compare a candidate W^* to all unbiased estimators (there may be infinite of them!).

We will discuss two approaches to find UMVUEs:

1. (**Cramér-Rao Inequality**) Determine a **lower bound** on the variance of *any* unbiased estimator of $\tau(\theta)$. If we can find an unbiased estimator whose variance attains this lower bound, we have found the UMVUE.
2. (**Rao-Blackwell Theorem**) Relate the property of UMVUEs with the notation of **sufficiency** and **completeness**.