Optimality in MSE

Zhijian He

2021/4/14

Mean squared error (MSE)

A rule of measuring accuracy of the estimator $T = T(X_1, \ldots, X_n)$:

$$\mathrm{MSE}_\theta(T) = \mathbb{E}[(T - g(\theta))^2] = \mathrm{Var}[T] + (\mathbb{E}[T] - g(\theta))^2$$

- T_1 is **as good as** T_2 iff

$$\mathrm{MSE}_\theta(T_1) \le \mathrm{MSE}_\theta(T_2), \ \forall\theta \in \Theta. \tag{1}$$

- T_1 is **better than** T_2 iff (1) holds and for some $\theta_0 \in \Theta$,

$$\mathrm{MSE}_{\theta_0}(T_1) < \mathrm{MSE}_{\theta_0}(T_2).$$

Optimality

Definition: Let $\mathcal{T}$ be a set of estimators of $g(\theta)$. If there is an estimator $T^* \in \mathcal{T}$ that is as good as other estimator in $\mathcal{T}$, then T^* is said to be $\mathcal{T}$**-optimal**.

Example: Let $X_1, \dots, X_n$ be iid sample of the population X with mean μ and variance σ^2. Consider unbiased estimators of μ having the form

$$\hat{\mu} = \sum_{i=1}^{n} c_i X_i,$$

where c_i are constants. How to find the optimal one?

Existence:

- Case 1: $\mathcal{T}$ is the set of all estimators.
- Case 2: $\mathcal{T}$ is the set of all unbiased estimators.

UMVUE

Definition: An unbiased estimator $T = T(X_1, \ldots, X_n)$ of $g(\theta)$ is called the uniformly minimum variance unbiased estimator (UMVUE) iff

$$\mathrm{Var}[T] \leq \mathrm{Var}[T'], \ \forall \theta \in \Theta,$$

for any other unbiased estimator T' of $g(\theta)$.

Remarks:

- UMVUE is $\mathcal{T}$-optimal in MSE with $\mathcal{T}$ being the class of all unbiased estimators.
- Existence of UMVUE was studied by Blackwell, Rao, Lehmann, Scheffe etc.

BLS Theorem

Theorem (Blackwell-Lehmann-Scheff, BLS): Suppose that there exists **sufficient and complete** statistic $T(X_1, \ldots, X_n)$ for $\theta \in \Theta$. If $g(\theta)$ is estimable, then there is a unique unbiased estimator of $g(\theta)$ that is of the form $\psi(T)$ with a Borel function $\psi(\cdot)$. Furthermore, $\psi(T)$ is the unique UMVUE of $g(\theta)$.

Remarks:

- The key is to find a sufficient and complete statistic.
- Please refer to Sections 1.5 and 1.6 in my lecture notes for details.

Cramer-Rao (CR) Ineqaultity

Suppse that $X_1, \dots, X_n$ are iid sample of the population X with PDF $f(x;\theta)$, where $\theta \in \Theta = (a,b) \subset \mathbb{R}$. Let $\psi(X_1, \dots, X_n)$ be an unbiased estimator of $g(\theta)$. Assume that $g'(\theta)$ and $\frac{df(x;\theta)}{d\theta}$ exist, and

$$\frac{d}{d\theta}\int_{\mathbb{R}} f(x;\theta)dx = \int_{\mathbb{R}} \frac{d}{d\theta} f(x;\theta)dx = 0,$$

$$\frac{d}{d\theta}\int_{\mathbb{R}^n} \psi(x_{1:n}) \prod_{i=1}^{n} f(x_i;\theta)dx_{1:n} = \int_{\mathbb{R}^n} \psi(x_{1:n}) \frac{d}{d\theta} \prod_{i=1}^{n} f(x_i;\theta)dx_{1:n},$$

and $I(\theta) := \mathbb{E}[(\frac{d}{d\theta} \ln f(X;\theta))^2] > 0$, then

$$\mathrm{Var}[\psi(X_1, \dots, X_n)] \geq \frac{g'(\theta)^2}{nI(\theta)}.$$

Cramer-Rao (CR) Inequality

Corollary: If $g(\theta) = \theta$, then

$$\mathrm{Var}[\psi(X_1, \ldots, X_n)] \geq \frac{1}{nI(\theta)}.$$

Remarks:

- $I(\theta) = \mathbb{E}[(\frac{d}{d\theta} \ln f(X; \theta))^2]$ is called the Fisher information. Under a mild condition, it is found that

$$I(\theta) = -\mathbb{E}\left[\frac{d^2}{d\theta^2} \ln f(X; \theta)\right].$$

- More information means that it is possible to get more accurate unbiased estimator.
- CR inequality also holds for discrete distributions with PMF $f(x; \theta)$.
- CR inequality can be generalized for the cases of $\theta \in \mathbb{R}^m$.

Example

Let $X_1, \ldots, X_n$ be iid sample of $N(\mu, \sigma_0^2)$, where σ_0^2 is known.

1. Compute the Fisher information $I(\mu)$.
2. The sample mean $\bar{X}$ is an unbiased estimator of μ. Justify that its variance attains the lower bound of CR inequality.
3. It is easy to see that $T = (\bar{X})^2 - \sigma^2/n$ is an unbiased estimator of $g(\mu) = \mu^2$. Justify that its variance *does not* attains the lower bound of CR inequality. (**Homework**)