Optimal estimation theory: Consistency, unbiasedness, and efficiency

Dr. Qiuzhuang Sun

STAT3023

Consistency

An estimator $\hat{\theta}$ is consistent for θ if $\hat{\theta} \xrightarrow{P} \theta$ when $n \to \infty$; in other words, for any $\varepsilon > 0$, we have

$$\lim_{n\to\infty} P\left(|\hat{\theta} - \theta| \geq \varepsilon\right) = 0.$$

Example: Assume the data $X_1, \ldots, X_n$ are iid realizations of a random variable X with finite mean and variance. Then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a consistent estimator of $\mu = E[X]$.

Consistency and mean square error (MSE)

The MSE of an estimator $\hat{\theta}$ is $\mathrm{MSE}(\hat{\theta}) = E\left\{(\hat{\theta} - \theta)^2\right\}$.

By Markov's inequality,

$$P\left(|\hat{\theta} - \theta| \geq \varepsilon\right) = P\left((\hat{\theta} - \theta)^2 \geq \varepsilon^2\right) \leq \frac{E\left\{(\hat{\theta} - \theta)^2\right\}}{\varepsilon^2},$$

so a sufficient condition for θ being consistent is

$$\mathrm{MSE}(\hat{\theta}) = E\left\{(\hat{\theta} - \theta)^2\right\} \to 0$$

when $n \to \infty$.

In practice, we can use MSE as a criterion to compare estimators.

Bias–variance decomposition of mean square error

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + (\mathrm{Bias}(\hat{\theta}))^2$$

Unbiased estimator

An estimator $\hat{\theta}$ is unbiased for θ if and only if

$$\mathrm{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta = 0.$$

An unbiased estimator is consistent if

$$\mathsf{Var}(\hat{\theta}) \to 0, \text{ as } n \to \infty.$$

Efficiency

Given the two unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, the relative efficiency of $\hat{\theta}_1$ versus $\hat{\theta}_2$ is defined as

$$\mathrm{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\mathrm{Var}(\hat{\theta}_1)}{\mathrm{Var}(\hat{\theta}_2)}.$$

An estimator $\hat{\theta}$ is a minimum variance unbiased estimator (MVUE) for θ of a given distribution if and only if $E[\hat{\theta}] = \theta$ and $\mathrm{Var}(\hat{\theta}) \leq \mathrm{Var}(\hat{\theta}')$ for any other unbiased estimator $\hat{\theta}'$ of θ.

Cramer-Rao Lower Bound (CRLB)

Recall $\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + (\mathrm{Bias}(\hat{\theta}))^2$. For unbiased estimators, $\mathrm{Bias}(\hat{\theta}) = 0$ so $\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta})$. **Then, can we make the variance as low as possible?**

Let $W(X) = W(X_1, \ldots, X_n)$ be an estimator and $f(x;\theta)$ be the PDF of each iid sample. Assume

$$\frac{\partial}{\partial\theta} \int h(x) f(x;\theta) dx = \int h(x) \frac{\partial}{\partial\theta} f(x;\theta) dx,$$

for $h(x) = 1$ and $h(x) = W(x)$. Then

$$\mathrm{Var}(W(X)) \geq \frac{\left\{ \frac{\partial}{\partial\theta} E_\theta[W(X)] \right\}^2}{\mathrm{Var}\left(\frac{\partial}{\partial\theta} \log(f(X;\theta)) \right)}.$$

Here, $\partial \log(f(X;\theta)) / \partial\theta$ is called the score function.

Proof of CRLB

Proof of CRLB

Proof of CRLB

Example

Let $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$. Let $\hat{\lambda}$ be the maximum likelihood estimator of λ based on $X = (X_1, \ldots, X_n)$. What is the CRLB of $\hat{\lambda}$?

Example

Let $X_1, \ldots, X_n \sim$ Poisson(λ). Let $\hat{\lambda}$ be the maximum likelihood estimator of λ based on $X = (X_1, \ldots, X_n)$. What is the CRLB of $\hat{\lambda}$?

Example

Again let $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$. Let $W(X)$ be an unbiased estimator of λ^2 based on $X = (X_1, \ldots, X_n)$. What is the CRLB of $W(X)$?

Attainment of CRLB

We know the CRLB is attained if and only if $\mathrm{Cor}(U, V) = \pm 1$, for $U = W(X)$ and $V = \frac{\partial}{\partial\theta}\log(f(X;\theta))$.

This amounts to the existence of a constant C_θ (which may depend on θ), such that

$$V - E[V] = C_\theta(U - E[U]).$$

Recalling $E[V] = 0$, we have

$$\frac{\partial}{\partial\theta}\log(f(X;\theta)) = C_\theta(W(X) - E[W(X)]).$$

Example

Use the previous property to show that the maximum likelihood estimator $\hat{\lambda} = \hat{\lambda}(X_1, \ldots, X_n)$ for Poisson(λ) attains the CRLB.

Example

Let $X_1, \ldots, X_n \sim N(\theta, 1)$ and $W(X) = W(X_1, \ldots, X_n)$ be an unbiased estimator of θ. What is the CRLB for $W(X)$? Show the CRLB can be attained.

Example

Let $X_1, \dots, X_n \sim N(\theta, 1)$ and $W(X) = W(X_1, \dots, X_n)$ be an unbiased estimator of θ. What is the CRLB for $W(X)$? Show the CRLB can be attained.

Example

Again assume $X_1, \ldots, X_n \sim N(\theta, 1)$, but this time let $W(X) = W(X_1, \ldots, X_n)$ be an unbiased estimator of θ^2. What is the CRLB for $W(X)$? Can the CRLB be attained?

Example

Again assume $X_1, \ldots, X_n \sim N(\theta, 1)$, but this time let $W(X) = W(X_1, \ldots, X_n)$ be an unbiased estimator of θ^2. What is the CRLB for $W(X)$? Can the CRLB be attained?

Attainment of CRLB

If the CRLB is attained for an unbiased estimator $W(X)$ of θ, then $\hat{\theta} = W(X)$ is MVUE of θ.

What can we do if the CRLB is not attained?

- In case the CRLB is not achieved, then we can combine unbiasedness with sufficiency to get a better result.
- This is called Rao-Blackwell Theorem.

Sufficiency and unbiasedness

(Rao-Blackwell Theorem) Let $\hat{\theta}_1$ be an unbiased estimator for θ, and let T be a sufficient statistic for θ. Define $\hat{\theta}_2 = E(\hat{\theta}_1 \mid T)$. Then $\hat{\theta}_2$ is unbiased for θ and is uniformly more efficient than $\hat{\theta}_1$.

Proof:

Sufficiency and unbiasedness

(Proof continued)

The Rao-Blackwell Theorem tells $\text{Var}(\hat{\theta}_2) \leq \text{Var}(\hat{\theta}_1)$. Then how much improvement can we achieve?

- It is NOT guaranteed to achieve the CRLB
- However, if X follows a full exponential family, then $\hat{\theta}_2$ is the best unbiased estimator, i.e., MVUE, we can get

Sufficiency and unbiasedness

If $X_1, \ldots, X_n$ follows an exponential family with PDF

$$\begin{aligned} f_X(x) &= h(x) \exp\left[\sum_{i=1}^{k} w_i(\theta) t_i(x) - A(\theta)\right] \\ &= h(x) \exp\left[\sum_{i=1}^{k} \eta_i t_i(x) - A^*(\eta)\right]. \end{aligned}$$

If the natural parameter space $\{(\eta_1, \ldots, \eta_k)\}$ contains an open subset of $\mathbb{R}^k$, then $T(X) = (\sum_{i=1}^n t_1(X_i), \ldots, \sum_{i=1}^n t_k(X_i))$ is a sufficient statistic. Any function of T is the best unbiased estimator for its expected value. (Proof omitted.)

The above result does NOT apply to the curved exponential family.

Example

Let $X_1, \ldots, X_n \sim$ Binomial(m, p) with known m. We want to estimate $\theta = P(X = 1) = \binom{m}{1} p^1 (1-p)^{m-1} = mp(1-p)^{m-1}$. Find the MVUE for θ.

Idea:

- Note that the binomial distribution belongs to the full exponential family
- Find a sufficient statistic T for θ
- Find a function of T that is unbiased for θ
- Use the Rao-Blackwell Theorem
  - Start with an unbiased estimator
  - Condition on the sufficient statistic

Example

Let $X_1, \ldots, X_n \sim \text{Binomial}(m, p)$ with known m. We want to estimate $\theta = P(X = 1) = \binom{m}{1} p^1 (1-p)^{m-1} = mp(1-p)^{m-1}$. Find the MVUE for θ.

Example

(Example continued)

Example

(Example continued)

Example

(Example continued)

Summary

We have a population distribution $f(x;\theta)$, where θ is the parameter

We estimate θ from the random sample $(X_1, \ldots, X_n)$

- Estimator: $W(X) = W(X_1, \ldots, X_n)$, not depending on θ
- Sufficiency: $X \to T(X) \to \theta$
- Consistency: $\hat{\theta} \xrightarrow{P} \theta$ as $n \to \infty$
- Unbiasedness: $E[\hat{\theta}] = \theta$
- Efficiency: minimum variance unbiased estimator (MVUE)
    - Cramer-Rao lower bound (CRLB): lower bound on the variance of the unbiased estimators
    - The CRLB achieved iff $\frac{\partial \log(f(X;\theta))}{\partial\theta} = C_\theta(W(X) - E[W(X)])$

$$\mathrm{Var}(W(X)) \geq \frac{\left\{\frac{\partial}{\partial\theta} E_\theta[W(X)]\right\}^2}{\mathrm{Var}\left(\frac{\partial}{\partial\theta}\log(f(X;\theta))\right)}$$