Methods of Evaluating Estimators

Tiandong Wang

Department of Statistics
Texas A&M University

Mean Squared Error (MSE)

Definitions:

- The *bias* of a point estimator W of θ is $\text{bias}_\theta(W) = E_\theta W - \theta$
- An estimator whose bias is equal to 0 is called *unbiased.*
- The *mean squared error* (MSE) of an estimator W of θ is defined by
$$MSE(W) = E_\theta(W - \theta)^2$$
- the MSE is a function of θ, and has the interpretation
$$\text{MSE} = E_\theta(W - \theta)^2 = Var_\theta W + (\text{Bias}_\theta W)^2.$$

Mean Squared Error: Examples

Examples:

- Example1: Let $X_1, \cdots, X_n$ iid $N(\mu, \sigma^2)$. Show $\bar{\mathbf{X}}$ is unbiased for μ and S^2 is unbiased for σ^2, and compute their MSEs.
- Example2: Let $X_1, \cdots, X_n$ iid $N(\mu, \sigma^2)$. Show the estimator $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is biased for σ^2, but it has a smaller MSE than S^2
- Example: Let $X_1, \cdots, X_n$ be a random sample from some distribution, and $E(X) = \mu$. Show that $\bar{\mathbf{X}}$ is a better estimator than X_1 and $(X_1 + X_2)/2$ for μ in terms of MSE.

Remarks on MSE

- the MSE measures the average squared difference between W and θ
- the MSE incorporates two components, one measuring the variability of the estimator (precision) and the other measuring its bias (accuracy).
- Small value of MSE implies small combined variance and bias.
- Unbiased estimators do a good job of controlling bias.
- Smaller MSE implies smaller probability for W to be far from θ, because (Chebyshev Inequality)

$$Pr(|W - \theta| > \epsilon) \leq \frac{1}{\epsilon^2} E_\theta (W - \theta)^2 = \frac{1}{\epsilon^2} MSE_\theta(W)$$

Remarks on MSE

- In general, there will not be one best estimator. Often the MSE of two estimators cross each other, showing that each estimator is better in only a portion of the parameter space. Example: Let X_1, X_2 be iid from $Bin(1, p)$ with $0 < p < 1$. Compare three estimators with respect to their MSE.
  - $\hat{p}_1 = X_1$
  - $\hat{p}_2 = \dfrac{X_1 + X_2}{2}$
  - $\hat{p}_3 = .5$
- The reason that there is no one best MSE estimator is the class of all estimators is too large a class. For example, $\hat{\theta} = 17$ is, in general, a terrible estimator because $MSE = (\theta - 17)^2$, but best when $\theta = 17$.
- Next, we will restrict our search to the class of unbiased estimators, and find the best estimator within this class.

Best Unbiased Estimators (UMVUE)

- We mainly focus on one-parameter case in this section. If the estimator W is unbiased for $\tau(\theta)$, then its MSE is equal to $Var_\theta(W)$.
- Therefore, choosing a better unbiased estimator is equivalent to choosing the one with smaller variance.
- The best estimator among all the unbiased estimators should have the uniformly smallest variance.

Best Unbiased Estimators (UMVUE)

Definition:
An estimator W^* is a *best unbiased estimator* of $\tau(\theta)$ if it satisfies:

- (1) $E_\theta W^* = \tau(\theta)$ for all θ;
- (2) For any other estimator W with $E_\theta W = \tau(\theta)$, we have

  $$Var_\theta W^* \leq Var_\theta W \quad \text{for all} \quad \theta. \quad \text{(Uniformly Smaller Variance)}$$

  W^* is also called a uniform minimum variance unbiased estimator (UMVUE)

Example: $X_1, \cdots, X_n$ iid $Poi(\lambda)$. Both $\bar{X}$ and S^2 are unbiased for λ.
Recall that If $X \sim \text{Pois}(\lambda)$, then $E[\lambda g(X)] = E[Xg(X-1)]$.

How To Find The Best Unbiased Estimator?

Method 1: Find the lower variance bound.
If $B(\theta)$ is a lower bound on the variance of any unbiased estimators of $\tau(\theta)$, and if W^* is unbiased satisfies $Var_\theta W^* = B(\theta)$, then W^* is a UMVUE.

How To Find The Best Unbiased Estimator?

Theorem

Let $X_1, \cdots, X_n$ be a sample with the joint pdf $f(\boldsymbol{x}; \theta)$. Suppose $W(\boldsymbol{X})$ is an estimator satisfying (i) $E_\theta W(\boldsymbol{X}) = \tau(\theta)$ for any $\theta \in \Theta$; (ii) $Var_\theta W(\boldsymbol{X}) < \infty$. Inf the following equation (interchangeability)

$$\frac{d}{d\theta} \int_{\mathcal{X}} h(x) f(x; \theta) dx = \int_{\mathcal{X}} h(x) \frac{\partial}{\partial \theta} [f(\boldsymbol{X}; \theta)] dx$$

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

$$Var_\theta(W(\boldsymbol{x})) \geq \frac{[\tau'(\theta)]^2}{E_\theta([\frac{\partial}{\partial \theta} \log f(\boldsymbol{x}; \theta)]^2)}$$

(The right-sided term in the inequality is called ***Cramér-Rao Lower Bound****)*

Cramér-Rao Inequality, iid case

Theorem

Let $X_1, \cdots, X_n$ be iid observations of the random variable with pdf $f(x;\theta)$, and the assumptions in the theorem above all hold. Then

$$Var_\theta(W(\boldsymbol{x})) \geq \frac{[\tau'(\theta)]^2}{nE_\theta([\frac{\partial}{\partial\theta}\log f(X;\theta)]^2)}$$

(The right-sided term in the inequality is called ***Cramér-Rao Lower Bound****)*

Remarks:

- For iid case, Cramér-Rao lower bound for unbiased estimators of θ is

$$\frac{1}{nE_\theta([\frac{\partial}{\partial\theta}\log f(X;\theta)]^2)}$$

- Cramér-Rao Lower Bound depends only on $\tau(\theta)$ and $f(\mathbf{x};\theta)$, and is a uniform lower bound on the variance.

Score Function and Fisher Information

Assume X is a random variable with pdf $f(x;\theta)$. Then

- The partial derivative of the log-likelihood with respect to θ is
$$s(X,\theta) = \frac{\partial}{\partial\theta}\log f(X;\theta) = \frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta}f(X;\theta)$$
is called the score or score function for X.
- The quantity
$$I(\theta) = E[s^2(X,\theta)] = E_\theta([\frac{\partial}{\partial\theta}\log f(X;\theta)]^2$$
is called the Fisher information number , or information number that X contains about θ.

Score Function and Fisher Information

If $X_1, \cdots, X_n$ is a random sample with the pdf $f(x;\theta)$. Then

- The score for the entire sample $X_1, \cdots, X_n$ is

$$s_n(\mathbf{X}, \theta) = \sum_{i=1}^{n} s(X_i, \theta)$$

- The Fisher information for the entire sample $X_1, \cdots, X_n$ is

$$I_n(\theta) = nI(\theta)$$

Score Function and Fisher Information

Remarks:

About the score,

- In general The score function is not a statistic, since it involves θ
- If differentiation and integration are interchangeable, then we have

$$E(s(X,\theta)) = 0, \quad \forall \theta \in \Theta$$

About the Fisher Information Number:

- If X and Y are independent, then $I_{X,Y}(\theta) = I_X(\theta) + I_Y(\theta)$
- The bigger the information number, the more information we have about θ, the smaller bound on the variance of the best unbiased estimator.
- If differentiation and integration are interchangeable, then the Fisher information number can be computed as

$$I(\theta) = Var[s(X,\theta)] = E[s^2(X,\theta)]$$

The following lemma helps in computation of Fisher information number.

Lemma

If $f(x;\theta)$ satisfies (the interchangeable condition)

$$\frac{\partial}{\partial\theta}E_\theta(\frac{\partial}{\partial\theta}\log f(X;\theta)) = \int \frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) f(x;\theta)\right] \mathrm{d}x$$

(true for exponential family), then

$$E_\theta([\frac{\partial}{\partial\theta}\log f(X;\theta)]^2) = -E_\theta(\frac{\partial^2}{\partial\theta^2}\log f(X;\theta))$$

This implies that $I(\theta) = -E[\frac{\partial}{\partial\theta}s(X,\theta)]$.

Examples:

- Example: $X_1, \cdots, X_n$ iid $Poi(\lambda)$. Find the Fisher information number and a UMVUE for λ
- $X_1, \cdots, X_n$ iid $N(\mu, \sigma^2)$, μ unknown but σ^2 known. (1) Find a UMVUE for μ;(2) Find a UMVUE for μ^2
  Remark: Find a complete statistic T for μ and find a function $\phi(\cdot)$ such that $\phi(T) = \mu^2$. Then $\phi(T)$ is the UMVUE.

When can we interchange the order of differentiation and integration?

Question: When can we interchange the order of differentiation and integration?

- yes for exponential family.
- not always true for non-exponential family. We have to do a match check for $\frac{\partial}{\partial\theta}\int_{\mathcal{X}} h(x)f(x;\theta)dx$ and $\int_{\mathcal{X}} h(x)\frac{\partial}{\partial\theta}[f(x;\theta)]dx$
- Example: $X_1, \cdots, X_n$ iid from $Unif(0, \theta)$. (Cramér-Rao bound does not work here!)

When is the Cramér-Rao Lower Bound attainable?

- The Cramér-Rao bound inequality says, if W^* achieves the variance bound then it is an UMVUE.
- In the one-parameter exponential family case, we can find such an estimator.
- But there is no guarantee that this lower bound is sharp (attainable) in other situations. It is possible that the value of Cramér-Rao bound may be strictly smaller than the variance of any unbiased estimator.

Corollary

(Attainment of C-R Bound)

Let $X_1, \cdots, X_n$ be iid with pdf $f(x;\theta)$, where $f(x;\theta)$ satisfies the assumptions of the C-R bound theorem. Let $L(\theta|\boldsymbol{x}) = \prod_{i=1}^n f(x_i;\theta)$ denote the likelihood function. If $W(\boldsymbol{X})$ is unbiased for $\tau(\theta)$, then $W(\boldsymbol{X})$ attains the C-R Lower Bound if and only if

$$a(\theta)[W(\boldsymbol{x}) - \tau(\theta)] = \frac{\partial}{\partial\theta}\log L(\theta|X)(\equiv s(\boldsymbol{x},\theta))$$

for some function $a(\theta)$

When is the Cramér-Rao Lower Bound attainable?: Examples

Examples:

- Example: $X_1, \cdots, X_n$ iid $Bin(1, p)$. Find an UMVUE of p and show it attains the Lower Bound.
- $X_1, \cdots, X_n$ iid $N(\mu, \sigma^2)$, μ unknown but σ^2 known. Find an UMVUE of μ and show it attains the Lower Bound.
- $X_1, \cdots, X_n$ iid Poisson(λ) and $\tau(\lambda) = \lambda^2$. Find a UMVUE for $\tau(\lambda)$. Does it attain the CRLB?

Cramér-Rao Lower Bound: Exponential family

One-parameter full-rank exponential family

Theorem

Let $X_1, \cdots, X_n$ be iid from the one-parameter exponential family with the pdf $f(x;\theta) = c(\theta)h(x)\exp\{w(\theta)T(x)\}$. Assume that $E[T(X)] = \tau(\theta)$. Then $\frac{1}{n}\sum_{i=1}^{n} T(X_i)$, as an unbiased estimator of $\tau(\theta)$, attains the C-R Lower Bound, i.e.

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} T(X_i)\right) = \frac{[\tau'(\theta)]^2}{I_n(\theta)}$$

Example:

- $X_1, \cdots, X_n$ iid $N(\mu, \sigma^2)$, μ unknown but σ^2 known. Consider estimation of μ. What is the C-R Lower bound and is it attainable?
- $X_1, \cdots, X_n$ iid $N(\mu, \sigma^2)$, both μ and σ^2 unknown. Consider estimation of σ^2. What is the C-R Lower bound and is it attainable?