Stochastic Processes

Methods of evaluating estimators and best unbiased estimators

Hamid R. Rabiee

Outline

- Methods of Evaluating Estimators
  - Mean Squared Error
  - Bias and Unbiasedness
- Best Unbiased Estimators
  - CR-Bound for variance
  - Sufficiency and UMVUE

Evaluating Estimators

- Which estimator?
- Task of choosing between estimators!
- This chapter Introduces some basic criteria for evaluating estimators.

Mean Squared Error

- Mean squared error

  Of estimator W

  Of parameter θ

$$E_\theta[\underbrace{(w-\theta)}^2]$$

Mean squared error

Mean Squared Error

- Why MSE?

  Any increasing function on $|W-\theta|$ could be good.

  But MSE:

  Analytically tractable

  $$E_\theta(W-\theta)^2 = Var_\theta W + (E_\theta W-\theta)^2 = Var_\theta W + (\mathrm{Bias}_\theta W)^2$$

  Two components of MSE variability of estimator (precision) and bias (accuracy).

Bias: difference between the expected value of W and θ.

MSE for an unbiased estimator

- Unbiased Estimator

  $$\text{Bias} \equiv 0 \quad \sim \quad E_\theta\,[W] = \theta, \forall\theta$$

- For an unbiased estimator:

  $$E_\theta[(W-\theta)^2] = Var_\theta W$$

Normal distribution estimators

Example: Let $X_1, X_2, \ldots, X_n$ be iid $N(\mu, \sigma^2)$.

$\bar{X}$: unbiased estimator for μ

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} \text{ and } \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$$

Are both estimator for σ^2

$\hat{\sigma}^2$ has lower MSE but S^2 is unbiased.

Best Unbiased Estimator

- Comparison of estimators based on MSE is not practically possible, Since the class of all estimators is too large.
- To limit the class of estimators, we only consider unbiased estimators.
- We have to find an unbiased estimator with minimum MSE.
- It equivalent to find an unbiased estimator with minimum variance. Why?

Best Unbiased Estimator

- For any unbiased estimator the MSE is equal to the variance.
- Definition (UMVUE):

  An estimator W* is a best unbiased estimator of $\tau(\theta)$ if it satisfies $E_\theta W^* = \tau(\theta)$ for all θ and for any other estimator of W with $E_\theta W = \tau(\theta)$ we have $Var(W^*) \leq Var(W)$ for all θ.

  It's also called MVUE

Best Unbiased Estimator

- Example:

  $x_1, x_2, \ldots, x_n$ are samples from iid poison(λ). Let $\bar{X}$ and S^2 be the sample mean and variance.

  We have to calculateV$ar(\bar{X})$ and $Var(S^2)$.

  Is it sufficient??

Best Unbiased Estimator

- Example:

  $x_1, x_2, \dots, x_n$ are samples from iid poison(λ). Let $\bar{X}$ and S^2 be the sample mean and variance.

  We have to calculateVar$(\bar{X})$ and $Var(S^2)$.

  How can we sure that there are not better (with less variance) estimator??

  We should find a lower band, $B(\theta)$ on the variance of any unbiased estimator.

  W^* is an UMVUE if $Var_\theta(W^*) = B(\theta)$

Cramer-Rao Theorem

- Theorem (Cramer-Rao):

  Let $X_1, \dots, X_n$ be a sample with pdf $f(x|\theta)$ and let $W(X) = W(X_1, \dots, X_n)$ be any estimator where $E_\theta(W(X))$ is a differentiable function of θ. Suppose the joint pdf $f(x_1, x_2, \dots, x_n|\theta)$ satisfies:

$$\frac{d}{d\theta}\int \dots \int h(x) f(x|\theta) dx_1 \dots dx_n = \int \dots \int h(x) \frac{\partial}{\partial\theta} f(x|\theta) dx_1 \dots dx_n \quad (1)$$

for any function h(x) with E(h(X)) < ∞. Then

$$Var\big(W(X)\big) \geq \frac{\left(\frac{d}{d\theta} E_\theta W(X)\right)^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right)}$$

Cramer-Rao bound for unbiased estimators

- If W is an unbiased estimator of θ then

$$Var\big(W(X)\big) \geq \frac{1}{E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right)}$$

Why?

Fisher information

- The Fisher information is a way of measuring the amount of information that an observable random variable X carries about an unknown parameter θ upon which the probability of X depends.
- The value

$$\mathrm{I}(\theta) = E_\theta\left((\frac{\partial}{\partial\theta} \log f(X|\theta))^2 \right)$$

is the fisher information and can also be derived by

$$\mathrm{I}(\theta) = -E_\theta\left(\frac{\partial^2}{\partial\theta^2} \log f(X|\theta) \right)$$

Note that: $0 \leq I(\theta) < \infty$

Fisher information and CR bound

- The variance of any unbiased estimator $\hat{\theta}$ of θ is bounded by the inverse of Fisher information:

$$Var(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

CR bound

- **Example:**
- Distribution $N(\mu, \sigma^2)$, μ is known, but σ^2 is unknown.

Show $T = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{n}$ is unbiased for σ^2.

What is variance of T?

Find UMVUE for σ^2.

CR bound, iid case

- Let $X_1, \ldots, X_n$ be a sample with pdf $f(x|\theta)$ and let $W(X) = W(X_1, \ldots, X_n)$ be any estimator where $E(W)$ is a differentiable function of θ. If the joint pdf satisfies the necessary condition for CR bound then:

$$Var\big(W(X)\big) \geq \frac{\left(\frac{d}{d\theta} E_\theta W(X)\right)^2}{nE_\theta((\frac{\partial}{\partial\theta} \log f(X|\theta))^2)}$$

- Why?

CR bound, iid case

- Let $X_1, \dots, X_n$ be a sample with pdf $f(x|\theta)$ and let $W(X) = W(X_1, \dots, X_n)$ be any estimator where $E(W)$ is a differentiable function of θ. If the joint pdf satisfies the necessary condition for CR bound then:

$$Var\big(W(X)\big) \geq \frac{\left(\frac{d}{d\theta} E_\theta W(X)\right)^2}{nE_\theta\left((\frac{\partial}{\partial\theta} \log f(X|\theta))^2 \right)}$$

$$E_\theta\left((\frac{\partial}{\partial\theta} \log f(X|\theta))^2 \right) = E_\theta\left((\frac{\partial}{\partial\theta} \log \prod f(X_i|\theta))^2 \right), \text{ independency}$$

$$\rightarrow\ E_\theta\left((\frac{\partial}{\partial\theta} \log f(X|\theta))^2 \right) = nE_\theta\left((\frac{\partial}{\partial\theta} \log f(X|\theta))^2 \right)$$

Best Unbiased Estimator (Cont.)

- Example: Let $X_1, \dots, X_n$ be iid Poisson(λ). Prove that $\bar{X}$ is a UMVUE of λ.
- Solution:
  - $\bar{X}$ is an unbiased estimator
  - $\bar{X}$ has variance equal to the Cramer-Rao lower bound

When Cramer-Rao does not apply

- **Example**: Let $X_1, \dots, X_n$ be iid with pdf

$$f(x|\theta) = \frac{1}{\theta}, \ 0<x<\theta.$$

- The Cramer-Rao bound:

$$VarW \geq \frac{\theta^2}{n}$$

- To find unbiased estimator there is a guess:

$$Y = \max(X_1, \dots, X_n)$$

- $E_\theta[Y] = \frac{n}{n+1}\theta \quad \Rightarrow \quad \frac{n+1}{n} Y$ is unbiased.

When Cramer-Rao does not apply

- $Var_\theta\left(\frac{n+1}{n}Y\right) = \frac{1}{n(n+2)}\theta^2$

This is uniformly smaller than $\frac{\theta^2}{n}$

- Cramer-Rao is not applicable
- Because: $\frac{d}{d\theta}\int_0^\theta h(x)f(x|\theta)dx = \frac{d}{d\theta}\int_0^\theta h(x)\frac{1}{\theta}dx$

$$= \frac{h(\theta)}{\theta} + \int_0^\theta h(x)\frac{\partial}{\partial\theta}\left(\frac{1}{\theta}\right)dx \neq \int_0^\theta h(x)\frac{\partial}{\partial\theta}f(x|\theta)dx$$

- Unless $\frac{h(\theta)}{\theta} = 0$ for all θ.

If the range of pdf, depends on the parameter, the theorem will not be applicable.

Sufficiency and Unbiasedness

- Theorem (Rao-Blackwell):

  Let W be any unbiased estimator of $\tau(\theta)$ and let T be a sufficient statistic for all θ. Define $\phi(T) = E(W|T)$. Then $E_\theta\phi(T) = \tau(\theta)$ and $Var(\phi(T)) \leq$ Var(W) for all θ, that is $\phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$.

Uniqeness of UMVUE

- Theorem:

  If W is the best unbiased estimator of $\tau(\theta)$, then W is unique.

  Proof sketch:

  Suppose W' is another UMVUE, and consider

  $$W^* = \frac{1}{2}(W + W') \rightarrow \; E(W^*) = \tau(\theta)$$

  Find the Variance of W^*.

How to find UMVUE

- Theorem:

  Let T be a complete sufficient statistic for a parameter θ, and let $\phi(T)$ be any estimator based on T. Then T is unique UMVUE of its expected value.

- What if there is no candidate estimator based on T ?

  If T is a CSS for a parameter θ and $h(x_1, x_2, \dots, x_n)$ is any unbiased estimator of $\tau(\theta)$. Then
  $\phi(T) = E(h(x_1, x_2, \dots, x_n)|\mathrm{T})$ is the UMVUE of $\tau(\theta)$.

Binomial best unbiased estimation

- **Example**: Let $X_1, \dots, X_n$ be iid $binomial(k, \theta)$.
- Estimate the probability of exactly one success.

$$\tau(\theta) = P_\theta(X = 1) = k\theta(1-\theta)^{k-1}$$

- Simple-minded estimator

$$h(X_1) = \begin{cases} 1 & if\ X_1 = 1 \\ 0 & otherwise \end{cases}$$

$$\phi\left(\sum_{i=1}^{n} X_i\right) = E\left(h(X_1) \middle| \sum_{i=1}^{n} X_i\right) = k\,\frac{\binom{k(n-1)}{\sum X_i - 1}}{\binom{kn}{\sum X_i}}$$

Consistency

- Definition:

 A sequence of estimators $W_n = W_n(X_1, \ldots, X_n)$ is a consistent sequence of estimators of the parameter θ if for every $\varepsilon > 0$ and every $\theta \in \Theta$

$$\lim_{n\to\infty} P_\theta(|W_n - W| < \varepsilon) = 1$$

- Theorem:

 If W_n is a sequence of estimators of a parameter θ satisfying $\lim_{n\to\infty} Var\, W_n = 0$ and $\lim_{n\to\infty} Bias\, W_n = 0$. Then W_n is a consistent sequence of estimators of θ.