Introduction to (Bayesian) Estimation
MAE 5020

Small-sample properties of estimators

Oklahoma State University

May 31, 2024

Core question:

How do we know whether the results from LSE or any estimator is good?

- Recall all estimators represent transformation of random data, which means that the estimate $\hat{\theta}(k)$ is random.
- We can study its properties from a statistical point of view.

Small-sample vs. large-sample properties

- *Sample*:
- Small sample:
- Large sample:

Relationship between small-sample and large-sample properties:

Properties

- Small-sample: unbiasedness and efficiency
- Large-sample: asymptotic unbiasedness, consistency, asymptotic normality, asymptotic efficiency

Unbiased

Property 1: Unbiased estimator Estimator $\hat{\theta}(k)$ is an unbiased estimator of deterministic θ if $E(\hat{\theta}(k)) = \theta$, or of random θ if $E(\hat{\theta}(k)) = E(\theta)$.

Example

Sample mean

Structure 1

Many estimators are of the form *Structure 1:*

$$\hat{\theta}(k) = F(k)Z(k)$$

When is it unbiased?

Fact: If $Z(k) = H(k)\theta + V(K)$, $E(V(K)) = 0$, and $H(k)$ and $F(k)$ are deterministic, then structure 1 is an unbiased estimator if $F(k)H(k) = I$, $\forall k$. Why?

Example: WLSE

Structure 2

Many recursive estimators are of the form *Structure 2:*

$$\hat{\theta}(k+1) = A(k+1)\hat{\theta}(k) + b(k+1)z(k+1).$$

Fact: If $z(k+1) = h^T(k+1)\theta + v(K+1)$, $E(v(k+1)) = 0$, $E(\hat{\theta}(k+1)) = E(\hat{\theta}(k))$ for any k, and $h(k+1)$ is deterministic, then structure 2 is an unbiased estimator if $A(k+1) = I - b(k+1)h^T(k+1)$, where $A(k+1)$ and $b(k+1)$ are deterministic.

Implications

- $A(k+1)$ and $b(k+1)$ are
- Another form of *Structure 2*

Efficiency

- The second order statistics of an estimator is important to understand the variation of the estimates produced by the estimator, provided that it is already unbiased.

Property 2: efficiency An unbiased estimator $\hat{\theta}(k)$ of a vector θ is said to be more efficient than other unbiased estimators $\hat{\hat{\theta}}(k)$ if

$$E\{\underbrace{[\theta - \hat{\theta}(k)]}_{\tilde{\theta}(k)}[\theta - \hat{\theta}(k)]^T\} \leq E\{[\theta - \hat{\hat{\theta}}(k)][\theta - \hat{\hat{\theta}}(k)]^T\}. \quad (1)$$

- When θ is a scalar,

- A more efficient unbiased estimator produces

Cramer-Rao lower bound

Let Z denote the dataset available to estimate θ and $\hat{\theta}(k)$ be any unbiased estimator of deterministic θ based on Z. Then

$$E\{\tilde{\theta}(k)\tilde{\theta}(k)\} \geq J^{-1}, \tag{2}$$

where J is the "Fisher information matrix"

$$J = E\{[\frac{\partial}{\partial\theta}\log p(Z)][\frac{\partial}{\partial\theta}\log p(Z)]^T\} = -E\{[\frac{\partial^2}{\partial\theta^2}\log p(Z)]\}. \tag{3}$$

The equality holds if and only if $\frac{\partial}{\partial\theta}\log p(Z) = c(\theta)\tilde{\theta}(k)$, where $C(\theta)$ is a matrix that does not depend on Z.

Comments

- The vector derivative must exist and the norm of $\partial p(Z)/\partial\theta$ must be absolutely integrable (needed in the proof).
- J^{-1} is called the Cramer Rao lower bound (CRLB). There are other lower bounds that can be tighter, e.g., Bhattacharyya bound, but they are even more difficult to compute than J^{-1}.
- For individual element of θ, we have $E\{\tilde{\theta}_i^2(k)\} \geq J_{ii}^{-1}$.

Comments

- Whether the CRLB is achieved or not depends on the estimator. The BLUE (best linear unbiased estimator) achieves the CRLB by design. Under certain conditions, a LSE is also efficient.
- $p(Z)$ is the probability density function of Z that depends on θ. It is characterized by a pdf $p_\theta(Z)$.
- Example of a Gaussian distribution

Example: computation of $J(\theta)$ in Gaussian setting

Small-sample properties of LSE

Recall when $H(k)$ is deterministic, WLSE is unbiased

When $H(k)$ is random

$\hat{\theta}_{WLS}(k) = [H(k)^T W(k) H(k)]^{-1} H^T(k) W(k) Z(k)$ is unbiased **if** $V(k)$ is zero mean and **if** $V(k)$ and $H(k)$ are statistically independent.

Proof:

Example: parameter estimation in a first-order system

Covariance of LSE

Covariance: If $V(k)$ is zero mean, if $V(k)$ and $H(k)$ are statistically independent and if $E(V(k)V(k)^T) = R(k)$ (i.e., the covariance of $V(k)$ is $R(k)$ since $V(k)$ is zero mean), then

$$cov(\tilde{\theta}_{WLS}(k)) =$$

When $H(k)$ is deterministic, $W(k) = I$ and $R(k) = \sigma_v^2 I$,

Proof

Efficiency of LSE

If $V(k)$ is zero mean, $H(k)$ is deterministic, and the components of $V(k)$ are independent and identically distributed with a constant covariance σ_v^2, i.e., $R(k) = \sigma_v^2 I$, then $\hat{\theta}_{LS}(k)$ is an efficient estimator within the class of linear estimators.

Proof

Because we don't know $p(Z)$, we cannot compute CRLB. We will show that $\hat{\theta}_{LS}(k)$ is more efficient than any other linear estimator.