Lec 7. Minimum variance unbiased estimator

Zhang W.P.

Contents

7.1 Rao-Blackwell theorem 2
7.2 Lehmann-Scheffé theorem 3
7.3 Unbiased estimators of zero 4
7.4 Cramér-Rao Inequality 7
7.5 Criticism of Unbiasedness 9

Learning objectives

After this section, you should be able to ...

- understand the Rao-Blackwell theorem
- understand the Lehmann-Scheffé theorem
- understand the Cramér-Rao Inequality
- find the MVUE

Uniform minimum variance unbiased estimator(UMVUE)

Let $X = (X_1, \ldots, X_n)$ be a sample from a parametric family $\{F_\theta, \theta \in \Theta\}$. An unbiased estimator $\hat{g}_n(X) = \hat{g}_n(X_1, \ldots, X_n)$ of $g(\theta)$ is the uniform minimum variance unbiased (UMVU) estimator of $g(\theta)$ if

Def.

$$Var_\theta(\hat{g}_n(X)) \leq Var_\theta(\hat{g}^*(X)) \text{ for all } \theta \in \Theta$$

where $\hat{g}^*(X)$ is any other unbiased estimator of $g(\theta)$.

7.1 Rao-Blackwell theorem

Theorem 1 (Rao-Blackwell)**.** *Suppose $S(X)$ is a sufficient statistic for $\{F_\theta, \theta \in \Theta\}$. If T is an estimate of $g(\theta)$ then:*

1. *$E(T \mid S)$ is a statistic.*
2. *$E(T \mid S)$ has the same bias as T; if T is unbiased so is $E(T \mid S)$.*
3. *$\mathrm{Var}_\theta(E(T \mid S)) \leq \mathrm{Var}_\theta(T), \theta \in \Theta$ and the inequality is strict unless T is a function of S.*

7.2 Lehmann-Scheffé theorem

Theorem 2. *If $T(X)$ is a complete sufficient statistic for $\{F_\theta, \theta \in \Theta\}$, and $\hat{g}_n(T)$ is an unbiased estimator of some parameter $g(\theta)$, then $\hat{g}_n(T)$ is, with probability 1, the unique UMVUE of $g(\theta)$.*

7.3 Unbiased estimators of zero

Note that if $\hat{g}_n(X)$ is an unbiased estimator of $g(\theta)$, then so is $\hat{g}_n(X)+aU(X)$, for any constant a and any unbiased estimator U of zero and that

$$Var_\theta[\hat{g}_n(X)+aU(X)] = Var_\theta[\hat{g}_n(X)]+a^2Var_\theta[U(X)]+2aCov_\theta(U(X),\hat{g}_n(X))$$

If $Cov_\theta(U(X),\hat{g}_n(X)) \neq 0$ for some θ, we shall show below that there exists a value of a for which

$$Var_\theta[\hat{g}_n(X)+aU(X)] < Var_\theta[\hat{g}_n(X)]$$

As a result, the covariance with unbiased estimators of zero is the key in characterizing the situations in which a UMVU estimator exists.

Theorem 3. *Let X be a sample from $F_\theta, \theta \in \Theta$, let $\delta(X)$ be an estimator in $\Delta = \{\delta : E_\theta \delta^2 < \infty$ for all $\theta \in \Theta\}$, and let $\mathcal{U}$ denote the set of all unbiased estimators of zero which are in Δ. Then, a necessary and sufficient condition for δ to be a UMVU estimator of its expectation $g(\theta)$ is that*

$$E_\theta(\delta U) = 0 \text{ for all } U \in \mathcal{U} \quad \text{and all } \theta \in \Theta \tag{7.1}$$

(Note: Since $E_\theta(U) = 0$ for all $U \in \mathcal{U}$, it follows that $E_\theta(\delta U) = \mathrm{cov}_\theta(\delta, U)$, so that (7.1) is equivalent to the condition that δ is uncorrelated with every $U \in \mathcal{U}$.)

Nonexistence of UMVU estimator

Theorem 3 can be used to check the nonexistence of UMVU estimator:

Let $X_1, \ldots, X_n$ be a sample from a discrete distribution which assigns probability $1/3$ to each of the points $\theta - 1, \theta, \theta + 1$, and let θ range over the integers. Then, no nonconstant function of θ has a UMVU estimator.

↑Exam ↓Exam

A continuous version of this example is provided by a sample from the uniform distribution $U(\theta - 1/2, \theta + 1/2)$;

↑Exam ↓Exam

7.4 Cramér-Rao Inequality

Theorem 4 (**Cramér-Rao lower bound**). *Suppose that $X_1, \ldots, X_n$ is a simple random sample from a population having a common density function $f(x \mid \theta)$ depending on a parameter θ, and let $\hat{g}_n(X)$ be an unbiased estimator of differentiable function $g(\theta)$ based on $X_1, \ldots, X_n$. If $f(x \mid \theta)$ is a smooth function of x and θ (C-R regular conditions), then*

$$\mathrm{Var}(\hat{g}) \geq \frac{[g'(\theta)]^2}{nI(\theta)}$$

where the Fisher information $I(\theta)$ is given by

$$I(\theta) := \mathbb{E}\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right)^2$$

for some random variable X having density $f(x \mid \theta)$.

Theorem 5. *Under the C-R regular conditions, and estimable function $g(\theta)$ has a continuous nonvanishing derivative on Θ, then a necessary and sufficient condition for its unbiased estimator $\hat{g}_n(X)$ achieving the C-R lower bound is that the sample distribution follows a one-parameter exponential family with $\hat{g}_n(X)$ being the sufficient statistic.*

7.5 Criticism of Unbiasedness

- UMVUE can be inadmissible for squared error loss meaning there is a (biased, of course) estimate whose MSE is smaller for every parameter value. An example is the UMVUE of $\phi = p(1-p)$ which is $\hat{\phi} = n\hat{p}(1-\hat{p})/(n-1)$. The MSE of

  $$\tilde{\phi} = \min(\hat{\phi}, 1/4)$$

  is smaller than that of $\hat{\phi}$.

- Unbiased estimation may be impossible.

- Binomial (n, p) log odds is $\phi = \log(p/(1-p))$. Since the expectation of any function of the data is a polynomial function of p and since ϕ is not a polynomial function of p there is no unbiased estimate of ϕ

- The UMVUE of σ is not the square root of the UMVUE of σ^2. This method of estimation does not have the parameterization equivariance that maximum likelihood does.

- Unbiasedness is irrelevant (unless you average together many estimators). The property is an average over possible values of the estimate in which positive errors are allowed to cancel negative errors. An exception to this criticism is that if you plan to average a number of estimators to get a single estimator then it is a problem if all the estimators have the same bias.

Efficiency

The efficiency of a unbiased estimator $\hat{g}_n(X)$ of $g(\theta)$ is defined as

$$e_{\hat{g}_n}(\theta) = \frac{[g'(\theta)]^2/(nI(\theta))}{Var_\theta(\hat{g}_n(X))}$$

Def.

- $0 < e_{\hat{g}_n}(\theta) \leq 1$
- **Effective estimation:** $e_{\hat{g}_n}(\theta) = 1$

Theorem 6. *Let $X = (X_1, \ldots, X_n)$ be a simple sample from population $\{f(x, \theta) : \theta \in$ open set $\Theta \subset R, \}$. If $g(\theta)$ has an efficient estimator $\hat{g}_n(X)$, then it must be the MLE of $g(\theta)$.*