5. Best Unbiased Estimators

Basic Theory

Consider again the basic statistical model, in which we have a random experiment that results in an observable random variable $\boldsymbol{X}$ taking values in a set S. Once again, the experiment is typically to sample n objects from a population and record one or more measurements for each item. In this case, the observable random variable has the form

$$\boldsymbol{X} = (X_1, X_2, ..., X_n)$$

where X_i is the vector of measurements for the i^{th} item.

Suppose that θ is a real parameter of the distribution of $\boldsymbol{X}$, taking values in a parameter space $\Theta \subseteq \mathbb{R}$. Let f_θ denote the probability density function of $\boldsymbol{X}$ for $\theta \in \Theta$. Note that the expected value, variance, and covariance operators also depend on θ, although we will sometimes suppress this to keep the notation from becoming too unwieldy.

Suppose now that $\lambda = \lambda(\theta)$ is a parameter of interest that is derived from θ. In this section we will consider the general problem of finding the best estimator of λ among a given class of unbiased estimators. Recall that if U is an unbiased estimator of λ, then $\text{var}_\theta(U)$ is the mean square error. Thus, if U and V are unbiased estimators of λ and

$$\text{var}_\theta(U) \le \text{var}_\theta(V) \text{ for all } \theta \in \Theta$$

Then U is a **uniformly better** estimator than V. On the other hand, it may be the case that U has smaller variance for some values of θ while V has smaller variance for other values of θ. If U is uniformly better than any other unbiased estimator of λ, then U is a **Uniformly Minimum Variance Unbiased Estimator** (**UMVUE**) of λ.

The Cramér-Rao Lower Bound

We will show that under mild conditions, there is a lower bound on the variance of any unbiased estimator of the parameter λ. Thus, if we can find an estimator that achieves this lower bound for all $\theta \in \Theta$, then the estimator must be an UMVUE of λ.

The assumption that we must make is that if $h : S \to \mathbb{R}$ with $\mathbb{E}_\theta(|h(\boldsymbol{X})|) < \infty$ for $\theta \in \Theta$ then

$$\frac{d\mathbb{E}_\theta(h(\boldsymbol{X}))}{d\theta} = \mathbb{E}_\theta\left(h(\boldsymbol{X})\frac{d\ln(f_\theta(\boldsymbol{X}))}{d\theta}\right), \quad \theta \in \Theta$$

1. Show that this condition is equivalent to the assumption that the derivative operator $\frac{d}{d\theta}$ can be interchanged with the expected value operator $\mathbb{E}_\theta$.

Generally speaking, the fundamental assumption will be satisfied if $f_\theta(\boldsymbol{x})$ is differentiable as a function of θ, with a derivative that is jointly continuous in $\boldsymbol{x}$ and θ, and if the support set $\{\boldsymbol{x} \in S : f_\theta(\boldsymbol{x}) > 0\}$ does not depend on θ.

2. Show that $\mathbb{E}_\theta\left(\frac{d\ln(f_\theta(\boldsymbol{X}))}{d\theta}\right) = 0$ for $\theta \in \Theta$. *Hint*: Use the basic condition with $h(\boldsymbol{x}) = 1$ for $\boldsymbol{x} \in S$.

3. Show that

$$\operatorname{cov}_\theta\left(h(\boldsymbol{X}), \frac{d\ln(f_\theta(\boldsymbol{X}))}{d\theta}\right) = \frac{d\mathbb{E}_\theta(h(\boldsymbol{X}))}{d\theta}$$

a. First note that the covariance is simply the expected value of the product of the variables, since the second variable has mean 0 by the Exercise 2.
b. Use the basic condition.

4. Prove the following result. *Hint*: The variable has mean 0.

$$\operatorname{var}_\theta\left(\frac{d\ln(f_\theta(\boldsymbol{X}))}{d\theta}\right) = \mathbb{E}_\theta\left(\left(\frac{d\ln(f_\theta(\boldsymbol{X}))}{d\theta}\right)^2\right)$$

5. Finally, use the Cauchy-Scharwtz inequality to establish the **Cramér-Rao lower bound**, named for Harold Cramér and CR Rao:

$$\operatorname{var}_\theta(h(\boldsymbol{X})) \ge \frac{\left(\frac{d\mathbb{E}_\theta(h(\boldsymbol{X}))}{d\theta}\right)^2}{\mathbb{E}_\theta\left(\left(\frac{d\ln(f_\theta(\boldsymbol{X}))}{d\theta}\right)^2\right)}$$

6. Suppose now that $\lambda(\theta)$ is a parameter of interest and $h(\boldsymbol{X})$ is an unbiased estimator of $\lambda(\theta)$. Use the Cramér-Rao lower bound to show that

$$\operatorname{var}_\theta(h(\boldsymbol{X})) \ge \frac{\left(\frac{d\lambda(\theta)}{d\theta}\right)^2}{\mathbb{E}_\theta\left(\left(\frac{d\ln(f_\theta(\boldsymbol{X}))}{d\theta}\right)^2\right)}$$

7. Show that equality holds in Exercise 6, and hence $h(\boldsymbol{X})$ is an UMVUE, if and only if there exists a function $u(\theta)$ such that (with probability 1)

$$h(\boldsymbol{X}) = \lambda(\theta) + u(\theta)\frac{d\ln(f_\theta(\boldsymbol{X}))}{d\theta}$$

a. Equality holds in the Cauchy-Schwartz inequality if and only if the random variables are linear transformations of each other.
b. Recall also that $\frac{d\ln(f_\theta(\boldsymbol{X}))}{d\theta}$ has mean 0.

The quantity $\mathbb{E}_\theta\left(\left(\frac{d\ln(f_\theta(\boldsymbol{X}))}{d\theta}\right)^2\right)$ that occurs in the denominator of the lower bounds of Exercise 5 and Exercise 6 is called the **Fisher information number** of $\boldsymbol{X}$, named after Sir Ronald Fisher. The following exercises gives an alternate version for the expression in Exercise 6 that is usually computationally better.

8. Show that if the appropriate derivatives exist and if the appropriate interchanges are permissible then

$$\mathbb{E}_\theta\left(\left(\frac{d\ln(f_\theta(\boldsymbol{X}))}{d\theta}\right)^2\right) = -\mathbb{E}_\theta\left(\frac{d^2}{d\theta^2}\ln(f_\theta(\boldsymbol{X}))\right)$$

9. Combine Exercise 6 and Exercise 8 to show that if $\lambda(\theta)$ is a parameter of interest and $h(\boldsymbol{X})$ is an unbiased estimator of $\lambda(\theta)$ then

$$\operatorname{var}_\theta(h(\boldsymbol{X})) \ge \frac{\left(\frac{d\lambda(\theta)}{d\theta}\right)^2}{-\mathbb{E}_\theta\left(\frac{d^2}{d\theta^2}\ln(f_\theta(\boldsymbol{X}))\right)}$$

Random Samples

Suppose now that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size n from the distribution of a random variable X having probability density function g_θ.

10. Prove the following special case of the Cramér-Rao lower bound. *Hint*: The joint probability density function is the product of the marginal probability density functions.

$$\operatorname{var}_\theta(h(\boldsymbol{X})) \ge \frac{\left(\frac{d\mathbb{E}_\theta(h(\boldsymbol{X}))}{d\theta}\right)^2}{n\,\mathbb{E}_\theta\left(\left(\frac{d\ln(g_\theta(X))}{d\theta}\right)^2\right)}$$

11. Suppose now that $\lambda(\theta)$ is a parameter of interest and $h(\boldsymbol{X})$ is an unbiased estimator of $\lambda(\theta)$. Use Exercise 10 to show that

$$\operatorname{var}_\theta(h(\boldsymbol{X})) \ge \frac{\left(\frac{d\lambda(\theta)}{d\theta}\right)^2}{n\,\mathbb{E}_\theta\left(\left(\frac{d\ln(g_\theta(X))}{d\theta}\right)^2\right)}$$

From Exercise 11, note that the Cramér-Rao lower bound varies inversely with the sample size n.

12. In the setting of the previous exercise, show the following result (assume that the appropriate derivatives exist and the appropriate interchanges are permissible):

$$\operatorname{var}_\theta(h(\boldsymbol{X})) \ge \frac{\left(\frac{d\lambda(\theta)}{d\theta}\right)^2}{-n\,\mathbb{E}_\theta\left(\frac{d^2}{d\theta^2}\ln(g_\theta(X))\right)}$$

Examples and Special Cases

We will apply the results above to several parametric families of distributions. First we need to recall some standard notation. Suppose that $\boldsymbol{X} = (X_2, X_2, ..., X_n)$ is a random sample of size n from the distribution of a real-valued random variable X with mean μ. The sample mean is

$$M = \frac{1}{n} \sum_{i=1}^{n} X_i$$

The special and standard versions of the sample variance are, respectively,

$$W^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - M)^2$$

The Bernoulli Distribution

Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size n from the Bernoulli distribution with unknown success parameter $p \in (0, 1)$. The basic assumption is satisfied.

13. Show that $\frac{1}{n} p (1 - p)$ is the Cramér-Rao lower bound for the variance of unbiased estimators of p.

14. Show that the sample mean M (which is the proportion of successes) attains the lower bound in the previous exercise and hence is an UMVUE of p.

The Poisson Distribution

Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size n from the Poisson distribution with unknown parameter $a \in (0, \infty)$. The basic assumption is satisfied.

15. Show that $\frac{a}{n}$ is the Cramér-Rao lower bound for the variance of unbiased estimators of a.

16. Show that the sample mean M attains the lower bound in the previous exercise and hence is an UMVUE of a.

The Normal Distribution

Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size n from the normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in (0, \infty)$. The basic assumption is satisfied with respect to both of these parameters. Recall also that the fourth central moment is $\mathbb{E}\left((X - \mu)^4\right) = 3\,\sigma^4$.

17. Show that $\frac{\sigma^2}{n}$ is the Cramér-Rao lower bound for the variance of unbiased estimators of μ.

18. Show that the sample mean M attains the lower bound in the previous exercise and hence is an

UMVUE of μ.

19. Show that $\frac{2\sigma^4}{n}$ is the Cramér-Rao lower bound for the variance of unbiased estimators of σ^2.

20. Show (or recall) that the sample variance S^2 has variance $\frac{2\sigma^4}{n-1}$ and hence does not attain the lower bound in the previous exercise.

21. Show that if μ is known, then the special sample variance W^2 attains the lower bound in Exercise 19 and hence is an UMVUE of σ^2.

22. Show that if μ is unknown, no unbiased estimator of σ^2 attains the Cramér-Rao lower bound in Exercise 19. *Hint:* Use the result in Exercise 7.

The Gamma Distribution

Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size n from the gamma distribution known shape parameter k and unknown scale parameter $b \in (0, \infty)$. The basic assumption is satisfied with respect to b.

23. Show that $\frac{b^2}{n\,k}$ is the Cramér-Rao lower bound for the variance of unbiased estimators of b.

24. Show that $\frac{M}{k}$ attains the lower bound in the previous exercise and hence is an UMVUE of b.

The Beta Distribution

Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size n from the beta distribution with left parameter $a > 0$ and right parameter $b = 1$. The basic assumption is satisfied with respect to a.

25. Show or recall that the mean and variance of the distribution are

a. $\mu = \frac{a}{a+1}$
b. $\sigma^2 = \frac{a}{(a+1)^2\,(a+2)}$

26. Show that the Cramér-Rao lower bound for the variance of unbiased estimators of μ is $\frac{a^2}{n\,(a+1)^4}$.

27. Show that the sample mean M does not achieve the Cramér-Rao lower bound in the previous exercise, and hence is not an UMVUE of μ.

The Uniform Distribution

Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size n from the uniform distribution on $[0, a]$ where $a > 0$ is the unknown parameter.

28. Show that the basic assumption is *not* satisfied.

29. Show that the Cramér-Rao lower bound for the variance of unbiased estimators of a is $\frac{a^2}{n}$. Of course, the Cramér-Rao Theorem does not apply, by the previous exercise.

30. Show (or recall) that $V = \frac{n+1}{n} \max\{X_1, X_2, ..., X_n\}$ is unbiased and has variance $\frac{a^2}{n\,(n+2)}$, which is smaller than the Cramér-Rao bound in the previous exercise.

The reason that the basic assumption is not satisfied is that the support set $\{x \in \mathbb{R} : f_a(x) > 0\}$ depends on the parameter a.

Best Linear Unbiased Estimators

We now consider a somewhat specialized problem, but one that fits the general theme of this section. Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a sequence of observable real-valued random variables that are uncorrelated and have the same unknown mean μ, but possibly different standard deviations. Let

$$\boldsymbol{\sigma} = (\sigma_1, \sigma_2, ..., \sigma_n) \text{ where } \sigma_i = \text{sd}(X_i) \text{ for } i \in \{1, 2, ..., n\}.$$

We will consider estimators of μ that are linear functions of the outcome variables. Specifically, we will consider estimators of the following form, where the vector of coefficients $\boldsymbol{c} = (c_1, c_2, ..., c_n)$ is to be determined:

$$Y = \sum_{i=1}^n c_i\, X_i$$

31. Show that Y is unbiased if and only if $\sum_{i=1}^n c_i = 1$.

32. Compute the variance of Y in terms of $\boldsymbol{c}$ and $\boldsymbol{\sigma}$.

33. Use the method of Lagrange multipliers (named after Joseph-Louis Lagrange) to show that the variance is minimized, subject to the unbiased constraint, when

$$c_j = \frac{1/\sigma_j{}^2}{\sum_{i=1}^n 1/\sigma_i{}^2}, \quad j \in \{1, 2, ..., n\}$$

This exercise shows how to construct the **Best Linear Unbiased Estimator** (**BLUE**) of μ, assuming that the vector of standard deviations σ is known.

Suppose now that $\sigma_i = \sigma$ for $i \in \{1, 2, ..., n\}$ so that the outcome variables have the same standard deviation. In particular, this would be the case if the outcome variables form a random sample of size n from a distribution with mean μ and standard deviation σ.

34. Show that in this case the variance is minimized when $c_i = \frac{1}{n}$ for each i and hence $Y = M$, the sample mean.

This exercise shows that the sample mean M is the best linear unbiased estimator of μ when the standard deviations are the same, and that moreover, we do not need to know the value of the standard deviation.

Virtual Laboratories > 7. Point Estimation > 1 2 3 4 5 6
Contents | Applets | Data Sets | Biographies | External Resources | Keywords | Feedback | ©