

7.5: 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 \mathbf{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

$$\mathbf{X} = (X_1, X_2, \dots, X_n) \quad (7.5.1)$$

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

Suppose that θ is a real parameter of the distribution of \mathbf{X} , taking values in a parameter space Θ . Let f_θ denote the probability density function of \mathbf{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.

Definitions

Suppose now that $\lambda = \lambda(\theta)$ is a parameter of interest that is derived from θ . (Of course, λ might be θ itself, but more generally might be a function of 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. Mean square error is our measure of the quality of unbiased estimators, so the following definitions are natural.

Suppose that U and V are unbiased estimators of λ .

1. If $\text{var}_\theta(U) \leq \text{var}_\theta(V)$ for all $\theta \in \Theta$ then U is a *uniformly better* estimator than V .
2. If U is uniformly better than every other unbiased estimator of λ , then U is a *Uniformly Minimum Variance Unbiased Estimator (UMVUE)* of λ .

Given unbiased estimators U and V of λ , 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 , so that neither estimator is uniformly better than the other. Of course, a minimum variance unbiased estimator is the best we can hope for.

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 θ , then the estimator must be an UMVUE of λ . The derivative of the log likelihood function, sometimes called the *score*, will play a critical role in our analysis. A lesser, but still important role, is played by the negative of the second derivative of the log-likelihood function. Life will be much easier if we give these functions names.

For $\mathbf{x} \in S$ and $\theta \in \Theta$, define

$$L_1(\mathbf{x}, \theta) = \frac{d}{d\theta} \ln(f_\theta(\mathbf{x})) \quad (7.5.2)$$

$$L_2(\mathbf{x}, \theta) = -\frac{d}{d\theta} L_1(\mathbf{x}, \theta) = -\frac{d^2}{d\theta^2} \ln(f_\theta(\mathbf{x})) \quad (7.5.3)$$

In the rest of this subsection, we consider statistics $h(\mathbf{X})$ where $h : S \rightarrow \mathbb{R}$ (and so in particular, h does not depend on θ). We need a fundamental assumption:

We will consider only statistics $h(\mathbf{X})$ with $\mathbb{E}_\theta(h^2(\mathbf{X})) < \infty$ for $\theta \in \Theta$. We also assume that

$$\frac{d}{d\theta} \mathbb{E}_\theta(h(\mathbf{X})) = \mathbb{E}_\theta(h(\mathbf{X})L_1(\mathbf{X}, \theta)) \quad (7.5.4)$$

This is equivalent to the assumption that the derivative operator $d/d\theta$ can be interchanged with the expected value operator \mathbb{E}_θ .

Proof

Note first that

$$\frac{d}{d\theta} \mathbb{E}(h(\mathbf{X})) = \frac{d}{d\theta} \int_S h(\mathbf{x}) f_{\theta}(\mathbf{x}) d\mathbf{x} \quad (7.5.5)$$

On the other hand,

$$\mathbb{E}_{\theta}(h(\mathbf{X})L_1(\mathbf{X}, \theta)) = \mathbb{E}_{\theta}\left(h(\mathbf{X})\frac{d}{d\theta}\ln(f_{\theta}(\mathbf{X}))\right) = \int_S h(\mathbf{x})\frac{d}{d\theta}\ln(f_{\theta}(\mathbf{x}))f_{\theta}(\mathbf{x}) d\mathbf{x} \quad (7.5.6)$$

$$= \int_S h(\mathbf{x})\frac{\frac{d}{d\theta}f_{\theta}(\mathbf{x})}{f_{\theta}(\mathbf{x})}f_{\theta}(\mathbf{x}) d\mathbf{x} = \int_S h(\mathbf{x})\frac{d}{d\theta}f_{\theta}(\mathbf{x}) d\mathbf{x} = \int_S \frac{d}{d\theta}h(\mathbf{x})f_{\theta}(\mathbf{x}) d\mathbf{x} \quad (7.5.7)$$

Thus the two expressions are the same if and only if we can interchange the derivative and integral operators.

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

$\mathbb{E}_{\theta}(L_1(\mathbf{X}, \theta)) = 0$ for $\theta \in \Theta$.

Proof

This follows from the [fundamental assumption](#) by letting $h(\mathbf{x}) = 1$ for $\mathbf{x} \in S$.

If $h(\mathbf{X})$ is a statistic then

$$\text{cov}_{\theta}(h(\mathbf{X}), L_1(\mathbf{X}, \theta)) = \frac{d}{d\theta} \mathbb{E}_{\theta}(h(\mathbf{X})) \quad (7.5.8)$$

Proof

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 [previous theorem](#). The result then follows from the [basic condition](#).

$\text{var}_{\theta}(L_1(\mathbf{X}, \theta)) = \mathbb{E}_{\theta}(L_1^2(\mathbf{X}, \theta))$

Proof

This follows since $L_1(\mathbf{X}, \theta)$ has mean 0 by the [theorem above](#).

The following theorem gives the general *Cramér-Rao lower bound* on the variance of a statistic. The lower bound is named for Harold Cramér and CR Rao:

If $h(\mathbf{X})$ is a statistic then

$$\text{var}_{\theta}(h(\mathbf{X})) \geq \frac{\left(\frac{d}{d\theta} \mathbb{E}_{\theta}(h(\mathbf{X}))\right)^2}{\mathbb{E}_{\theta}(L_1^2(\mathbf{X}, \theta))} \quad (7.5.9)$$

Proof

From the Cauchy-Scharwtz (correlation) inequality,

$$\text{cov}_{\theta}^2(h(\mathbf{X}), L_1(\mathbf{X}, \theta)) \leq \text{var}_{\theta}(h(\mathbf{X})) \text{var}_{\theta}(L_1(\mathbf{X}, \theta)) \quad (7.5.10)$$

The result now follows from the previous two theorems.

We can now give the first version of the Cramér-Rao lower bound for unbiased estimators of a parameter.

Suppose now that $\lambda(\theta)$ is a parameter of interest and $h(\mathbf{X})$ is an unbiased estimator of λ . Then

$$\text{var}_{\theta}(h(\mathbf{X})) \geq \frac{(d\lambda/d\theta)^2}{\mathbb{E}_{\theta}(L_1^2(\mathbf{X}, \theta))} \quad (7.5.11)$$

Proof

This follows immediately from the Cramér-Rao lower bound, since $\mathbb{E}_\theta(h(\mathbf{X})) = \lambda$ for $\theta \in \Theta$.

An estimator of λ that achieves the Cramér-Rao lower bound must be a uniformly minimum variance unbiased estimator (UMVUE) of λ .

Equality holds in the previous theorem, and hence $h(\mathbf{X})$ is an UMVUE, if and only if there exists a function $u(\theta)$ such that (with probability 1)

$$h(\mathbf{X}) = \lambda(\theta) + u(\theta)L_1(\mathbf{X}, \theta) \quad (7.5.12)$$

Proof

Equality holds in the Cauchy-Schwartz inequality if and only if the random variables are linear transformations of each other. Recall also that $L_1(\mathbf{X}, \theta)$ has mean 0.

The quantity $\mathbb{E}_\theta(L^2(\mathbf{X}, \theta))$ that occurs in the denominator of the lower bounds in the previous two theorems is called the *Fisher information number* of \mathbf{X} , named after Sir Ronald Fisher. The following theorem gives an alternate version of the Fisher information number that is usually computationally better.

If the appropriate derivatives exist and if the appropriate interchanges are permissible then

$$\mathbb{E}_\theta(L_1^2(\mathbf{X}, \theta)) = \mathbb{E}_\theta(L_2(\mathbf{X}, \theta)) \quad (7.5.13)$$

The following theorem gives the second version of the Cramér-Rao lower bound for unbiased estimators of a parameter.

If $\lambda(\theta)$ is a parameter of interest and $h(\mathbf{X})$ is an unbiased estimator of λ then

$$\text{var}_\theta(h(\mathbf{X})) \geq \frac{(d\lambda/d\theta)^2}{\mathbb{E}_\theta(L_2(\mathbf{X}, \theta))} \quad (7.5.14)$$

Proof

This follows from the results above.

Random Samples

Suppose now that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the distribution of a random variable X having probability density function g_θ and taking values in a set R . Thus $S = R^n$. We will use lower-case letters for the derivative of the log likelihood function of X and the negative of the second derivative of the log likelihood function of X .

For $x \in R$ and $\theta \in \Theta$ define

$$l(x, \theta) = \frac{d}{d\theta} \ln(g_\theta(x)) \quad (7.5.15)$$

$$l_2(x, \theta) = -\frac{d^2}{d\theta^2} \ln(g_\theta(x)) \quad (7.5.16)$$

L^2 can be written in terms of l^2 and L_2 can be written in terms of l_2 :

1. $\mathbb{E}_\theta(L^2(\mathbf{X}, \theta)) = n\mathbb{E}_\theta(l^2(X, \theta))$
2. $\mathbb{E}_\theta(L_2(\mathbf{X}, \theta)) = n\mathbb{E}_\theta(l_2(X, \theta))$

The following theorem gives the second version of the general Cramér-Rao lower bound on the variance of a statistic, specialized for random samples.

If $h(\mathbf{X})$ is a statistic then

$$\text{var}_{\theta}(h(\mathbf{X})) \geq \frac{\left(\frac{d}{d\theta} \mathbb{E}_{\theta}(h(\mathbf{X}))\right)^2}{n \mathbb{E}_{\theta}(l^2(X, \theta))} \quad (7.5.17)$$

The following theorem give the third version of the Cramér-Rao lower bound for unbiased estimators of a parameter, specialized for random samples.

Suppose now that $\lambda(\theta)$ is a parameter of interest and $h(\mathbf{X})$ is an unbiased estimator of λ . Then

$$\text{var}_{\theta}(h(\mathbf{X})) \geq \frac{(d\lambda/d\theta)^2}{n \mathbb{E}_{\theta}(l^2(X, \theta))} \quad (7.5.18)$$

Note that the Cramér-Rao lower bound varies inversely with the sample size n . The following version gives the fourth version of the Cramér-Rao lower bound for unbiased estimators of a parameter, again specialized for random samples.

If the appropriate derivatives exist and the appropriate interchanges are permissible) then

$$\text{var}_{\theta}(h(\mathbf{X})) \geq \frac{(d\lambda/d\theta)^2}{n \mathbb{E}_{\theta}(l_2(X, \theta))} \quad (7.5.19)$$

To summarize, we have four versions of the Cramér-Rao lower bound for the variance of an unbiased estimate of λ : [version 1](#) and [version 2](#) in the general case, and [version 1](#) and [version 2](#) in the special case that \mathbf{X} is a random sample from the distribution of X . If an unbiased estimator of λ achieves the lower bound, then the estimator is an UMVUE.

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 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the distribution of a real-valued random variable X with mean μ and variance σ^2 . The sample mean is

$$M = \frac{1}{n} \sum_{i=1}^n X_i \quad (7.5.20)$$

Recall that $\mathbb{E}(M) = \mu$ and $\text{var}(M) = \sigma^2/n$. The special version of the sample variance, when μ is known, and standard version of the sample variance are, respectively,

$$W^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad (7.5.21)$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - M)^2 \quad (7.5.22)$$

The Bernoulli Distribution

Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the Bernoulli distribution with unknown success parameter $p \in (0, 1)$. In the usual language of reliability, $X_i = 1$ means success on trial i and $X_i = 0$ means failure on trial i ; the distribution is named for Jacob Bernoulli. Recall that the Bernoulli distribution has probability density function

$$g_p(x) = p^x(1-p)^{1-x}, \quad x \in \{0, 1\} \quad (7.5.23)$$

The [basic assumption](#) is satisfied. Moreover, recall that the mean of the Bernoulli distribution is p , while the variance is $p(1-p)$.

$p(1-p)/n$ is the [Cramér-Rao lower bound](#) for the variance of unbiased estimators of p .

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 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the Poisson distribution with parameter $\theta \in (0, \infty)$. Recall that this distribution is often used to model the number of “random points” in a region of time or space and is studied in more detail in the chapter on the Poisson Process. The Poisson distribution is named for Simeon Poisson and has probability density function

$$g_\theta(x) = e^{-\theta} \frac{\theta^x}{x!}, \quad x \in \mathbb{N} \quad (7.5.24)$$

The [basic assumption](#) is satisfied. Recall also that the mean and variance of the distribution are both θ .

θ/n is the [Cramér-Rao lower bound](#) for the variance of unbiased estimators of θ .

The sample mean M attains the lower bound in the previous exercise and hence is an UMVUE of θ .

The Normal Distribution

Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in (0, \infty)$. Recall that the normal distribution plays an especially important role in statistics, in part because of the central limit theorem. The normal distribution is widely used to model physical quantities subject to numerous small, random errors, and has probability density function

$$g_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad x \in \mathbb{R} \quad (7.5.25)$$

The [basic assumption](#) is satisfied with respect to both of these parameters. Recall also that the fourth central moment is $\mathbb{E}((X - \mu)^4) = 3\sigma^4$.

σ^2/n is the [Cramér-Rao lower bound](#) for the variance of unbiased estimators of μ .

The sample mean M attains the lower bound in the previous exercise and hence is an UMVUE of μ .

$\frac{2\sigma^4}{n}$ is the [Cramér-Rao lower bound](#) for the variance of unbiased estimators of σ^2 .

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.

If μ is known, then the special sample variance W^2 attains the lower bound [above](#) and hence is an UMVUE of σ^2 .

If μ is unknown, no unbiased estimator of σ^2 attains the Cramér-Rao lower bound [above](#).

Proof

This follows from the result [above](#) on equality in the Cramér-Rao inequality.

The Gamma Distribution

Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the gamma distribution with known shape parameter $k > 0$ and unknown scale parameter $b > 0$. The gamma distribution is often used to model random times and certain other types of positive random variables, and is studied in more detail in the chapter on Special Distributions. The probability density function is

$$g_b(x) = \frac{1}{\Gamma(k)b^k} x^{k-1} e^{-x/b}, \quad x \in (0, \infty) \quad (7.5.26)$$

The [basic assumption](#) is satisfied with respect to b . Moreover, the mean and variance of the gamma distribution are kb and kb^2 , respectively.

$\frac{b^2}{nk}$ is the [Cramér-Rao lower bound](#) for the variance of unbiased estimators of b .

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

The Beta Distribution

Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the beta distribution with left parameter $a > 0$ and right parameter $b = 1$. Beta distributions are widely used to model random proportions and other random variables that take values in bounded intervals, and are studied in more detail in the chapter on Special Distributions. In our specialized case, the probability density function of the sampling distribution is

$$g_a(x) = a x^{a-1}, \quad x \in (0, 1) \quad (7.5.27)$$

The **basic assumption** is satisfied with respect to a .

The mean and variance of the distribution are

$$\begin{aligned} 1. \mu &= \frac{a}{a+1} \\ 2. \sigma^2 &= \frac{a}{(a+1)^2(a+2)} \end{aligned}$$

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

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 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the uniform distribution on $[0, a]$ where $a > 0$ is the unknown parameter. Thus, the probability density function of the sampling distribution is

$$g_a(x) = \frac{1}{a}, \quad x \in [0, a] \quad (7.5.28)$$

The **basic assumption** is *not* satisfied.

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.

Recall that $V = \frac{n+1}{n} \max\{X_1, X_2, \dots, X_n\}$ is unbiased and has variance $\frac{a^2}{n(n+2)}$. This variance 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 $\{x \in \mathbb{R} : g_a(x) > 0\}$ 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 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a sequence of observable real-valued random variables that are uncorrelated and have the same unknown mean $\mu \in \mathbb{R}$, but possibly different standard deviations. Let $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_n)$ where $\sigma_i = \text{sd}(X_i)$ for $i \in \{1, 2, \dots, n\}$.

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 $\mathbf{c} = (c_1, c_2, \dots, c_n)$ is to be determined:

$$Y = \sum_{i=1}^n c_i X_i \quad (7.5.29)$$

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

The variance of Y is

$$\text{var}(Y) = \sum_{i=1}^n c_i^2 \sigma_i^2 \quad (7.5.30)$$

The variance is minimized, subject to the unbiased constraint, when

$$c_j = \frac{1/\sigma_j^2}{\sum_{i=1}^n 1/\sigma_i^2}, \quad j \in \{1, 2, \dots, n\} \quad (7.5.31)$$

Proof

Use the method of Lagrange multipliers (named after Joseph-Louis Lagrange).

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 $i \in \{1, 2, \dots, n\}$ 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 .

In this case the variance is minimized when $c_i = 1/n$ 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.

This page titled [7.5: Best Unbiased Estimators](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.