8.2.1 Evaluating Estimators

We define three main desirable properties for point estimators. The first one is related to the estimator's *bias*. The bias of an estimator $\hat{\Theta}$ tells us on average how far $\hat{\Theta}$ is from the real value of θ.

Let $\hat{\Theta} = h(X_1, X_2, \cdots, X_n)$ be a point estimator for θ. The **bias** of point estimator $\hat{\Theta}$ is defined by

$$B(\hat{\Theta}) = E[\hat{\Theta}] - \theta.$$

In general, we would like to have a bias that is close to 0, indicating that on average, $\hat{\Theta}$ is close to θ. It is worth noting that $B(\hat{\Theta})$ might depend on the actual value of θ. In other words, you might have an estimator for which $B(\hat{\Theta})$ is small for some values of θ and large for some other values of θ. A desirable scenario is when $B(\hat{\Theta}) = 0$, i.e, $E[\hat{\Theta}] = \theta$, for all values of θ. In this case, we say that $\hat{\Theta}$ is an *unbiased* estimator of θ.

Let $\hat{\Theta} = h(X_1, X_2, \cdots, X_n)$ be a point estimator for a parameter θ. We say that $\hat{\Theta}$ is an **unbiased** of estimator of θ if

$$B(\hat{\Theta}) = 0, \qquad \text{for all possible values of } \theta.$$

Example 8.2

Let X_1, X_2, X_3, ..., X_n be a random sample. Show that the sample mean

$$\hat{\Theta} = \overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

is an unbiased estimator of $\theta = EX_i$.

Solution

We have

$$\begin{aligned} B(\hat{\Theta}) &= E[\hat{\Theta}] - \theta \\ &= E\left[\overline{X}\right] - \theta \\ &= EX_i - \theta \\ &= 0. \end{aligned}$$

Note that if an estimator is unbiased, it is not necessarily a good estimator. In the above example, if we choose $\hat{\Theta}_1 = X_1$, then $\hat{\Theta}_1$ is also an unbiased estimator of θ:

$$\begin{aligned} B(\hat{\Theta}_1) &= E[\hat{\Theta}_1] - \theta \\ &= EX_1 - \theta \\ &= 0. \end{aligned}$$

Nevertheless, we suspect that $\hat{\Theta}_1$ is probably not as good as the sample mean $\overline{X}$. Therefore, we need other measures to ensure that an estimator is a "good" estimator. A very common measure is the *mean squared error* defined by $E\big[(\hat{\Theta} - \theta)^2\big]$.

> The **mean squared error** (MSE) of a point estimator $\hat{\Theta}$, shown by $MSE(\hat{\Theta})$, is defined as
>
> $$MSE(\hat{\Theta}) = E\big[(\hat{\Theta} - \theta)^2\big].$$

Note that $\hat{\Theta} - \theta$ is the error that we make when we estimate θ by $\hat{\Theta}$. Thus, the MSE is a measure of the distance between $\hat{\Theta}$ and θ, and a smaller MSE is generally indicative of a better estimator.

Example 8.3

Let X_1, X_2, X_3, $\ldots$, X_n be a random sample from a distribution with mean $EX_i = \theta$, and variance $\mathrm{Var}(X_i) = \sigma^2$. Consider the following two estimators for θ:

1. $\hat{\Theta}_1 = X_1$.
2. $\hat{\Theta}_2 = \overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$.

Find $MSE(\hat{\Theta}_1)$ and $MSE(\hat{\Theta}_2)$ and show that for $n > 1$, we have

$$MSE(\hat{\Theta}_1) > MSE(\hat{\Theta}_2).$$

Solution

We have

$$\begin{aligned} MSE(\hat{\Theta}_1) &= E\big[(\hat{\Theta}_1 - \theta)^2\big] \\ &= E[(X_1 - EX_1)^2] \\ &= \mathrm{Var}(X_1) \\ &= \sigma^2. \end{aligned}$$

To find $MSE(\hat{\Theta}_2)$, we can write

$$\begin{aligned} MSE(\hat{\Theta}_2) &= E\big[(\hat{\Theta}_2 - \theta)^2\big] \\ &= E[(\overline{X} - \theta)^2] \\ &= \mathrm{Var}(\overline{X} - \theta) + \big(E[\overline{X} - \theta]\big)^2. \end{aligned}$$

The last equality results from $EY^2 = \mathrm{Var}(Y) + (EY)^2$, where $Y = \overline{X} - \theta$. Now, note that

$$\mathrm{Var}(\overline{X} - \theta) = \mathrm{Var}(\overline{X})$$

since θ is a constant. Also, $E[\overline{X} - \theta] = 0$. Thus, we conclude

$$\begin{aligned} MSE(\hat{\Theta}_2) &= \mathrm{Var}(\overline{X}) \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

Thus, we conclude for $n > 1$,

$$MSE(\hat{\Theta}_1) > MSE(\hat{\Theta}_2).$$

From the above example, we conclude that although both $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased estimators of the mean, $\hat{\Theta}_2 = \overline{X}$ is probably a better estimator since it has a smaller MSE. In general, if $\hat{\Theta}$ is a point estimator for θ, we can write

$$\begin{aligned} MSE(\hat{\Theta}) &= E\big[(\hat{\Theta} - \theta)^2\big] \\ &= \mathrm{Var}(\hat{\Theta} - \theta) + \big(E[\hat{\Theta} - \theta]\big)^2 \\ &= \mathrm{Var}(\hat{\Theta}) + B(\hat{\Theta})^2. \end{aligned}$$

If $\hat{\Theta}$ is a point estimator for θ,

$$MSE(\hat{\Theta}) = \mathrm{Var}(\hat{\Theta}) + B(\hat{\Theta})^2,$$

where $B(\hat{\Theta}) = E[\hat{\Theta}] - \theta$ is the bias of $\hat{\Theta}$.

The last property that we discuss for point estimators is *consistency*. Loosely speaking, we say that an estimator is consistent if as the sample size n gets larger, $\hat{\Theta}$ converges to the real value of θ. More precisely, we have the following definition:

Let $\hat{\Theta}_1, \hat{\Theta}_2, \cdots, \hat{\Theta}_n, \cdots$, be a sequence of point estimators of θ. We say that $\hat{\Theta}_n$ is a **consistent** estimator of θ, if

$$\lim_{n \to \infty} P\big(|\hat{\Theta}_n - \theta| \geq \epsilon\big) = 0, \text{ for all } \epsilon > 0.$$

Example 8.4

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample with mean $EX_i = \theta$, and variance $\mathrm{Var}(X_i) = \sigma^2$. Show that $\hat{\Theta}_n = \overline{X}$ is a consistent estimator of θ.

Solution

We need to show that

$$\lim_{n \to \infty} P\big(|\overline{X} - \theta| \geq \epsilon\big) = 0, \qquad \text{for all } \epsilon > 0.$$

But this is true because of the weak law of large numbers. In particular, we can use Chebyshev's inequality to write

$$\begin{aligned} P(|\overline{X} - \theta| \geq \epsilon) &\leq \frac{\mathrm{Var}(\overline{X})}{\epsilon^2} \\ &= \frac{\sigma^2}{n\epsilon^2}, \end{aligned}$$

which goes to 0 as $n \to \infty$.

We could also show the consistency of $\hat{\Theta}_n = \overline{X}$ by looking at the MSE. As we found previously, the MSE of $\hat{\Theta}_n = \overline{X}$ is given by

$$MSE(\hat{\Theta}_n) = \frac{\sigma^2}{n}.$$

Thus, $MSE(\hat{\Theta}_n)$ goes to 0 as $n \rightarrow \infty$. From this, we can conclude that $\hat{\Theta}_n = \overline{X}$ is a consistent estimator for θ. In fact, we can state the following theorem:

Theorem 8.2
Let $\hat{\Theta}_1, \hat{\Theta}_2, \cdots$ be a sequence of point estimators of θ. If

$$\lim_{n \rightarrow \infty} MSE(\hat{\Theta}_n) = 0,$$

then $\hat{\Theta}_n$ is a consistent estimator of θ.

Proof

We can write

$$\begin{aligned} P(|\hat{\Theta}_n - \theta| \geq \epsilon) &= P(|\hat{\Theta}_n - \theta|^2 \geq \epsilon^2) \\ &\leq \frac{E[\hat{\Theta}_n - \theta]^2}{\epsilon^2} \qquad \text{(by Markov's inequality)} \\ &= \frac{MSE(\hat{\Theta}_n)}{\epsilon^2}, \end{aligned}$$

which goes to 0 as $n \rightarrow \infty$ by the assumption.