Chapter 8

Point Estimation

Inferential statistics consists of two main branches – estimation and hypothesis testing. Point estimation selects a value of a sample statistic to estimate a population parameter. We call the value this statistic a *point estimate*. Since statistics are random variables, the point estimate is simply a realization of the random variable. For example, $\bar{X}$ and S^2 are used to find point estimates $\bar{x}$ and s^2 for the mean and variance of a population. Point estimates are numbers even though the statistics or *point estimators* themselves are random variables. We will use lower case to denote the point estimates and upper case to denote the statistic. There are also *interval estimators* and *interval estimates*, which we will discuss after this chapter. Since our point estimators or statistics are random variables, we look at their sampling distribution. However, the sampling distribution comes after selecting which point estimator to use. Point estimators have various properties we use to decide which estimator to use, including *unbiasedness*, *efficiency*, and *consistency*.

8.1 Unbiased Estimators

Definition A statistic $\hat{\theta}$ is an *unbiased estimator* of the population parameter θ if and only if

$$\mathbb{E}[\hat{\theta}] = \theta.$$

Unbiasedness means that across samples the estimator will be the parameter value on average. On average means over repeated samples. For instance, suppose we take M different samples of size n and estimate $\hat{\theta}$ M times where $\theta = 1.11$.

m	$\hat{\theta}$
1	1.10
2	1.05
3	1.12
.	.
.	.
.	.
M	1.09

If $\hat{\theta}$ is an unbiased estimator of θ, then the average of the M estimates should be close to θ.

Example We showed that $\mathbb{E}[\bar{X}] = \mu$ when X_i is *i.i.d.*, so the sample mean is an unbiased estimator for the population mean when the random sample is independently and identically distributed.

Example Suppose we have a random sample $X_1, X_2, ..., X_n$ from the population represented by the random variable x with a PDF

$$f(x) = e^{-(x-\delta)},\ x > \delta.$$

$$\begin{aligned}
\mathbb{E}[X] &= \int_{\delta}^{\infty} x \cdot e^{-(x-\delta)} \mathrm{d}x \\
&= xe^{-(x-\delta)}\Big|_{\delta}^{\infty} + \int_{\delta}^{\infty} e^{-(x-\delta)} \mathrm{d}x \\
&= \delta - e^{-(x-\delta)}\Big|_{\delta}^{\infty} \\
&= \delta - (0-1) \\
&= 1 + \delta
\end{aligned}$$

We know $\mathbb{E}[\bar{X}] = 1 + \delta$. So while the sample mean is an unbiased estimator for the mean, it is a *biased* estimator for the unknown parameter δ. An unbiased estimator for δ will simply be $\bar{X} - 1$ where -1 is the *bias correction factor.*

Definition The *finite sample bias* $b_n(\theta)$ is

$$b_n(\theta) = \mathbb{E}[\hat{\theta}] - \theta.$$

The bias we defined is a finite sample bias, because the sample size is finite. We can also define asymptotic bias, which is when the sample size $n \to \infty$.

Definition $\hat{\theta}$ is an *asymptotically unbiased estimator* of θ if and only if

$$\lim_{n\to\infty} b_n(\theta) = 0.$$

Similarly, the *asymptotic bias* for a estimator $\hat{\theta}$ is

$$\lim_{n\to\infty} b_n(\theta) = \lim_{n\to\infty} \left(\mathbb{E}(\hat{\theta}) - \theta\right).$$

Proposition 8.1.1 *S^2 is an unbiased estimator of σ^2 and $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an asymptotically unbiased estimator of σ^2.*

Proof

$$\begin{aligned}
\mathbb{E}[S^2] &= \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] \\
&= \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^{n}\left((X_i - \mu) - (\bar{X} - \mu)\right)^2\right] \\
&= \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}\mathbb{E}(X_i - \mu)^2 - n\mathbb{E}(\bar{X} - \mu)^2\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}\sigma^2 - n\frac{\sigma^2}{n}\right] \\
&= \frac{1}{n-1}\left[\sigma^2(n-1)\right] \\
&= \sigma^2
\end{aligned}$$

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}\big((X_i-\mu)-(\bar{X}-\mu)\big)^2\right]$$

$$\vdots$$

$$= \frac{(n-1)\sigma^2}{n}$$

$$\to \sigma^2 \text{ as } n\to\infty \quad \blacksquare$$

The naive sample variance estimator has a finite sample bias of $-\frac{\sigma^2}{n}$, but this bias disappears as the sample size grows. We have to multiply by the bias correction factor of $\frac{n-1}{n}$ to obtain an unbiased estimator from the naive sample variance. Note that asymptotic unbiasedness depends on the sample size but unbiasedness does not. In this sense, unbiasedness is a stronger property. Typically, obtaining unbiased estimators is rather difficult but obtaining large samples is easy, so we often opt for biased estimators even though we desire this property.

8.2 Efficiency

Efficiency is a relative property that trades off bias of an estimator with its precision. Precision is the inverse of the variance. More precise estimators have smaller variance. Many estimators yield precise but biased estimates, so efficiency provides a criterion to guide us in trading off precision and bias. The criterion often used is called the *mean-squared error*.

Definition The mean-squared error (MSE) for an estimator $\hat{\theta}$ is

$$MSE(\hat{\theta}) = \mathbb{E}[\hat{\theta}-\theta]$$

where $\mathbb{E}[\hat{\theta}]$ is not necessarily θ. We can express the MSE as

$$MSE(\hat{\theta}) = bias^2 + Var(\hat{\theta}),$$

which is the squared bias plus the variance.

Now we can define efficiency.

Definition A estimator $\hat{\theta}$ is *more efficient* than an estimator $\hat{\phi}$ if and only if

$$MSE(\hat{\theta}) < MSE(\hat{\phi}).$$

Thus for unbiased estimators $\tilde{\theta}$ and $\tilde{\phi}$, $\tilde{\theta}$ is more efficient than $\tilde{\phi}$ if and only if

$$Var(\tilde{\theta}) < Var(\tilde{\phi}).$$

Efficiency is a comparative property. An estimator can only be more or less efficiency – not efficient. Efficiency is a notion of reliability, because it weights both bias and precision. We can also define *asymptotic efficiency* which compares the mean-squared error as the sample size grows.

Example Consider a new estimator $Y = \frac{\bar{X}}{2}$. Which estimator for the mean is more efficient? Y or $\bar{X}$?

$$Var(Y) = \frac{\sigma^2}{4n} < \frac{\sigma^2}{n} = Var(\bar{X})$$

$$b_n(Y) = -\frac{\mu}{2} > b_n(\bar{X}) = 0$$

$$MSE(Y) = \frac{\mu^2}{4} + \frac{\sigma^2}{4n},\ MSE(\bar{X}) = \frac{\sigma^2}{2}$$

Then Y is more efficient than $\bar{X}$ if and only if $n\mu^2 < 3\sigma^2$. So when the mean is small compared to the variance, we can shrink the sample mean to obtain biased estimates of the population mean with a high precision gain compared to the bias introduced over the sample mean.