

The fact that S^2 is an unbiased estimator of σ^2 for any population distribution is one of the most compelling reasons to use the $n - 1$ in the denominator of the definition of S^2 . This result does not imply, however, that $E[S] = \sigma$.

The previous two examples have illustrated unbiased estimates of parameters. It is helpful to quantify the bias in order to have a measure of the expected distance between a point estimator and its target value.

Definition 2.2 Let $\hat{\theta}$ denote a statistic that is calculated from the sample X_1, X_2, \dots, X_n . The *bias* associated with using $\hat{\theta}$ as an estimator of θ is

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

There is a subset of the biased estimators that is of interest. The classification is a bit of a consolation prize for biased estimators. Their redeeming feature is that although they are biased estimators for finite sample sizes n , they are unbiased as $n \rightarrow \infty$. These estimators are known as asymptotically unbiased estimators and are defined formally below.

Definition 2.3 Let $\hat{\theta}$ denote a statistic that is calculated from the sample X_1, X_2, \dots, X_n . If

$$\lim_{n \rightarrow \infty} B(\hat{\theta}, \theta) = 0,$$

then $\hat{\theta}$ is an *asymptotically unbiased estimator* of θ .

All unbiased estimators are necessarily asymptotically unbiased. But only some of the biased estimators are asymptotically unbiased. To this end, we subdivide the biased portion of the Venn diagram from Figure 2.12 to include asymptotically unbiased estimators in Figure 2.14.

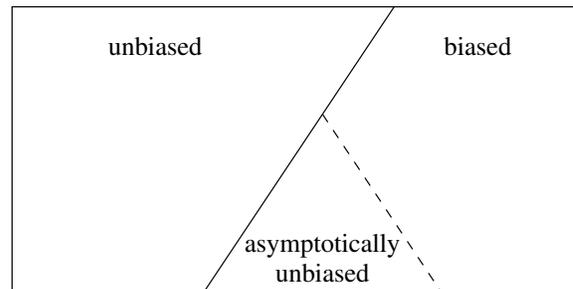


Figure 2.14: Venn diagram of unbiased, biased, and asymptotically unbiased point estimators.

Example 2.17 Let X_1, X_2, \dots, X_n denote a random sample from a $U(0, \theta)$ population, where θ is a positive unknown parameter. Classify the following point estimators of θ into the categories given in Figure 2.14 and select the best estimator:

- $2\bar{X}$,
- $3\bar{X}$,
- $X_{(n)}$,

- $(n+1)X_{(n)}/n$,
- $(n+1)X_{(1)}$,
- 17,

where $X_{(1)} = \min\{X_1, X_2, \dots, X_n\}$ and $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$.

When faced with a real data set, we oftentimes have to choose a point estimator from a set of potential point estimators such as this. The purpose of this example is to investigate the properties of these six point estimators.

- The first point estimator, $2\bar{X}$, is the method of moments estimator. The derivation was given in Example 2.2. Since the population mean of the $U(0, \theta)$ distribution is $\theta/2$ and

$$E[2\bar{X}] = 2E[\bar{X}] = 2 \cdot \frac{\theta}{2} = \theta,$$

via Example 2.16, the method of moments estimator is classified as an unbiased estimator.

- The point estimator $3\bar{X}$ is classified as a biased estimator because

$$E[3\bar{X}] = 3E[\bar{X}] = 3 \cdot \frac{\theta}{2} = \frac{3}{2}\theta.$$

This estimator overestimates the population parameter θ on average. The positive bias is

$$B(\hat{\theta}, \theta) = B(3\bar{X}, \theta) = E[3\bar{X}] - \theta = \frac{3}{2}\theta - \theta = \frac{\theta}{2}.$$

- The point estimator $X_{(n)}$ is the maximum likelihood estimator. The derivation, after some minor manipulation of the objective function or the support of the population distribution, was given in Example 2.9. Using an order statistic result, or the APPL code

```
X := UniformRV(0, theta);
Y := OrderStat(X, n, n);
Mean(Y);
```

the expected value of $X_{(n)}$ is

$$E[X_{(n)}] = \frac{n\theta}{n+1}.$$

The maximum likelihood estimator misses low, on average, because $E[X_{(n)}]$ is less than θ . Since the expected value is not equal to θ for finite values of n , this estimator is biased. The bias is

$$B(\hat{\theta}, \theta) = B(X_{(n)}, \theta) = E[X_{(n)}] - \theta = \frac{n\theta}{n+1} - \theta = -\frac{\theta}{n+1}.$$

This estimator should be classified as asymptotically unbiased, however, because

$$\lim_{n \rightarrow \infty} B(\hat{\theta}, \theta) = \lim_{n \rightarrow \infty} \left(-\frac{\theta}{n+1} \right) = 0.$$

- The point estimator $(n+1)X_{(n)}/n$ was presented in Example 2.9 as a modification of the maximum likelihood estimator that included an unbiasing constant. The expected value of $(n+1)X_{(n)}/n$ is

$$E\left[\frac{n+1}{n}X_{(n)}\right] = \frac{n+1}{n} \cdot \frac{n\theta}{n+1} = \theta,$$

so this point estimator is classified as an unbiased estimator.

- The point estimator $(n+1)X_{(1)}$ is also an unbiased estimator. This can be seen by invoking an order statistic result and computing the expected value, or by the APPL code

```
X := UniformRV(0, theta);
Y := OrderStat(X, n, 1);
Mean(Y);
```

- The point estimator $\hat{\theta} = 17$ is quite bizarre. The statistician simply ignores the data values X_1, X_2, \dots, X_n and pulls 17 out of thin air as the estimate of θ . The expected value of $\hat{\theta}$ is $E[\hat{\theta}] = E[17] = 17$, which is not θ (unless θ just happens to be 17), so this estimator is classified as a biased estimator.

We now know that three of the six suggested point estimators are unbiased. The results of our analysis are summarized in Figure 2.15.

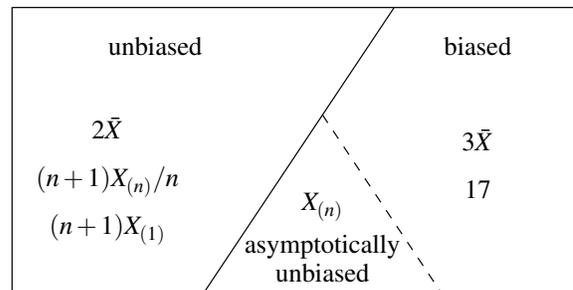


Figure 2.15: Venn diagram of several point estimates for θ for a $U(0, \theta)$ population.

Now to the more difficult question: which is the best of the six estimates? This is a purposefully vague question at this point, so the question will be addressed from several different angles. The choice between the point estimators boils down to which point estimator will perform best for a higher fraction of data sets than the other point estimators. This does not imply, of course, that the estimator selected will be the best for every data set. We begin by plotting the sampling distributions of the three unbiased estimators to gain some additional insight. This can only be done for specific values of n and θ , so let's arbitrarily choose $n = 5$ and $\theta = 10$. For this choice, the probability density functions of $2\bar{X}$, $6X_{(5)}/5$, and $6X_{(1)}$ are plotted in Figure 2.16. APPL was used to calculate the probability density functions. The sampling distributions of $2\bar{X}$, $6X_{(5)}/5$, and $6X_{(1)}$ reveal vastly different shapes. The probability density function of $2\bar{X}$ is bell shaped (via the central limit theorem) and symmetric about $\theta = 10$; the probability density functions of $6X_{(5)}/5$ and $6X_{(1)}$ are skewed distributions. Since the support of

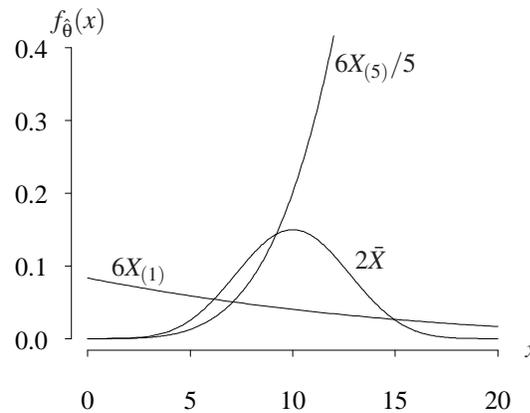


Figure 2.16: Sampling distributions of $2\bar{X}$, $6X_{(5)}/5$, and $6X_{(1)}$ when $n = 5$ and $\theta = 10$.

the population is $(0, 10)$, the support of $2\bar{X}$ is $(0, 20)$, the support of $6X_{(5)}/5$ is $(0, 12)$, and the support of $6X_{(1)}$ is $(0, 60)$. Figure 2.16 reveals that $6X_{(1)}$ has a significantly larger variance than the other two unbiased estimators, so it is probably the weakest candidate of the three unbiased estimators.

Since the variance of the estimators played a role in analyzing the sampling distributions of the three unbiased estimators, perhaps it is worthwhile calculating the population means and population variances of all six of the estimators. The values are summarized in Table 2.5. Notice that the three unbiased estimators, $2\bar{X}$, $(n+1)X_{(n)}/n$, and $(n+1)X_{(1)}$ all collapse to the same estimator when $n = 1$; the point estimator is just double the single observation. Choosing the point estimator with the smallest variance is not appropriate here because this would result in choosing the strange point estimate

Point estimate $\hat{\theta}$	$E[\hat{\theta}]$	$V[\hat{\theta}]$	Categorization
$2\bar{X}$	θ	$\frac{\theta^2}{3n}$	unbiased
$3\bar{X}$	$\frac{3\theta}{2}$	$\frac{3\theta^2}{4n}$	biased
$X_{(n)}$	$\frac{n\theta}{n+1}$	$\frac{n\theta^2}{(n+2)(n+1)^2}$	asymptotically unbiased
$\frac{(n+1)X_{(n)}}{n}$	θ	$\frac{\theta^2}{n(n+2)}$	unbiased
$(n+1)X_{(1)}$	θ	$\frac{n\theta^2}{n+2}$	unbiased
17	17	0	biased

Table 2.5: Population means and variances of the six point estimators for θ .

$\hat{\theta} = 17$. Instead, it is advantageous to choose the unbiased estimator with the smallest variance. Using this criteria, $(n+1)X_{(n)}/n$ has the smallest variance of the three for samples of $n = 2$ or more observations.

But the unbiased estimator with the smallest variance is not the only criteria that can be used to select the preferred estimator. The R code below simulates 10,000 random samples of size $n = 5$ from a $U(0, \theta)$ population when $\theta = 10$. All six point estimators are calculated for each sample, and the point estimator that lies closest to θ is identified and tabulated. Finally, the fraction of times that each estimator is closest to $\theta = 10$ is printed.

```
set.seed(8)
n = 5
theta = 10
nrep = 10000
theta.hat = numeric(6)
count = numeric(6)
for (i in 1:nrep) {
  x = runif(n, 0, theta)
  theta.hat[1] = 2 * mean(x)
  theta.hat[2] = 3 * mean(x)
  theta.hat[3] = max(x)
  theta.hat[4] = (n + 1) * max(x) / n
  theta.hat[5] = (n + 1) * min(x)
  theta.hat[6] = 17
  index = which.min(abs(theta.hat - theta))
  count[index] = count[index] + 1
}
print(count / nrep)
```

The results of the Monte Carlo simulation are given in Table 2.6 for sample sizes $n = 5$, $n = 50$, and $n = 500$. The entries give the fractions of the simulations giving the closest estimator to the true parameter value $\theta = 10$. As expected, the column sums of the entries in the table equal 1. When $n = 5$, even the maligned $\hat{\theta} = 17$ is the closest to $\theta = 10$ for two of the 10,000 random samples. The reader is encouraged to imagine what type

Point estimate	$n = 5$	$n = 50$	$n = 500$
$2\bar{X}$	0.1765	0.0912	0.0328
$3\bar{X}$	0.1323	0.0000	0.0000
$X_{(n)}$	0.3178	0.3749	0.3905
$(n+1)X_{(n)}/n$	0.3262	0.5275	0.5762
$(n+1)X_{(1)}$	0.0470	0.0064	0.0005
17	0.0002	0.0000	0.0000

Table 2.6: Monte Carlo simulation results for a $U(0, 10)$ population.

of data set would lead to this awful estimator outdoing the other estimators. Table 2.6 shows that, by a somewhat narrow margin, the unbiased estimator $(n+1)X_{(n)}/n$ dominates the other estimators for the sample sizes considered here.

In summary, based on $\hat{\theta} = (n+1)X_{(n)}/n$ being (a) an unbiased estimate, (b) the unbiased estimate with the smallest variance, and (c) the estimate that is most likely to be the closest to the population value of θ for several sample sizes in a Monte Carlo experiment, we conclude that $\hat{\theta} = (n+1)X_{(n)}/n$ is the best of the six point estimators. It carries the additional bonus that all of the data values are necessarily less than $\hat{\theta}$, which is a desirable property for this particular population distribution.

This example has brought up three issues concerning point estimators that will be addressed in the paragraphs that follow.

The first issue is motivated by the Monte Carlo simulation experiment. The objective of the experiment was to find the point estimator that was most likely to be closest to the true parameter value. The distance between the estimator $\hat{\theta}$ and the true parameter value θ is an important quantity known as the error of estimation, which is formally defined next.

Definition 2.4 Let $\hat{\theta}$ denote a statistic that is calculated from the sample X_1, X_2, \dots, X_n that is used to estimate the population parameter θ . The *error of estimation* is

$$R(\hat{\theta}, \theta) = |\hat{\theta} - \theta|.$$

The second issue concerns the comparison of the three unbiased estimators and the three biased estimators. Could there ever be circumstances in which one would choose a biased estimator over an unbiased estimator? Consider the generic and idealized presentation of the sampling distributions of two point estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ in Figure 2.17 for a fixed sample size n . The sampling distribution of $\hat{\theta}_1$ is centered over the true parameter value θ , so $\hat{\theta}_1$ is an unbiased estimator of θ , that is, $E[\hat{\theta}_1] = \theta$. The sampling distribution of $\hat{\theta}_2$, however, is not centered over the true parameter value θ , so $\hat{\theta}_2$ is a biased estimator of θ , that is, $E[\hat{\theta}_2] \neq \theta$. But the decision between the two is complicated by the fact that the variance of the second estimator is much smaller than the variance of the first estimator, that is, $V[\hat{\theta}_2] < V[\hat{\theta}_1]$. The choice between the unbiased estimator with the larger variance and the biased estimator with the smaller variance is a difficult one.

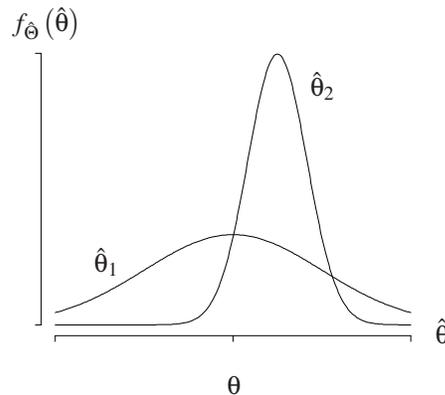


Figure 2.17: Two sampling distributions.