Economics 103 – Statistics for Economists

Francis J. DiTraglia

University of Pennsylvania

Lecture 15

Sampling Distributions and Estimation – Part II

Unbiased means "Right on Average"

Bias of an Estimator

Let $\widehat{\theta}_n$ be a sample estimator of a population parameter θ_0. The *bias* of $\widehat{\theta}_n$ is $E[\widehat{\theta}_n] - \theta_0$.

Unbiased Estimator

A sample estimator $\widehat{\theta}_n$ of a population parameter θ_0 is called *unbiased* if $E[\widehat{\theta}_n] = \theta_0$

Why $(n-1)$ for sample variance?

Why $(n-1)$ for sample variance?

We will show that having $n-1$ in the denominator ensures:

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

under random sampling.

Why $(n-1)$ for sample variance?

Step $\#$ 1 – Tedious but straightforward algebra gives:

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

You are not responsible for proving Step $\#1$ on an exam.

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

Why $(n-1)$ for sample variance?

Step # 2 – Take Expectations of Step # 1:

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

Where we have used the linearity of expectation.

Why $(n-1)$ for sample variance?

Step # 3 – Use assumption of random sampling:

$X_1, \ldots, X_n \sim$ iid with mean μ and variance σ^2

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

Since we showed earlier today that $E[\bar{X}] = \mu$ and $Var(\bar{X}) = \sigma^2/n$ under this random sampling assumption.

Why $(n-1)$ for sample variance?

Finally – Divide Step # 3 by $(n-1)$:

$$E[S^2] \quad = \quad E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right] = \frac{(n-1)\sigma^2}{n-1} = \sigma^2$$

Hence, having $(n-1)$ in the denominator ensures that the sample variance is "correct on average," that is *unbiased*.

A Different Estimator of the Population Variance

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$$

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

Bias of $\widehat{\sigma}^2$

$$E[\widehat{\sigma}^2] - \sigma^2 = \frac{(n-1)\sigma^2}{n} - \sigma^2 = \frac{(n-1)\sigma^2}{n} - \frac{n\sigma^2}{n} = -\sigma^2/n$$

How Large is the Average Family?

How many brothers and sisters are in your family, including yourself?

The average number of children per family was about 2.0 twenty years ago.

What's Going On Here?

Biased Sample!

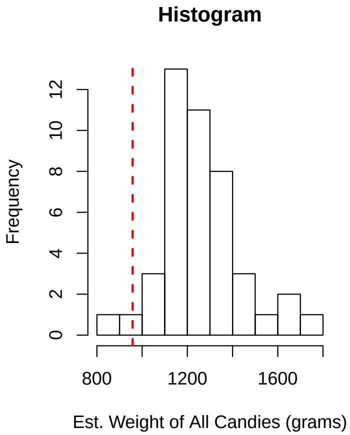
- Zero children $\Rightarrow$ didn't send any to college
- Sampling by *children* so large families oversampled

Candy Weighing: 44 Estimates, Each With $n = 5$

$$\widehat{\theta} = 20 \times (X_1 + \ldots + X_5)$$

Summary of Sampling Dist.	
Overestimates	42
Exactly Correct	0
Underestimates	2
$E[\hat{\theta}]$	1269 grams
$\text{SD}(\widehat{\theta})$	189 grams

Actual Mass: $\theta_0 =$958 grams



What was in the bag?

100 Candies Total:

- 20 Fun Size Snickers Bars (large)
- 30 Reese's Miniatures (medium)
- 50 Tootsie Roll "Midgees" (small)

So What Happened?

Not a random sample! The Snickers bars were *oversampled*.

Could we have avoided this? How?

Let $X_1, X_2, \ldots X_n \sim iid$ mean μ, variance σ^2 and define $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. True or False:

$\bar{X}_n$ *is an unbiased estimator of* μ

(a) True

(b) False

TRUE!

Let $X_1, X_2, \ldots X_n \sim iid$ mean μ, variance σ^2. True or False:

X_1 *is an unbiased estimator of* μ

(a) True

(b) False

TRUE!

How to choose between two unbiased estimators?

Suppose $X_1, X_2, \ldots X_n \sim iid$ with mean μ and variance σ^2

$$\begin{aligned}
E[\bar{X}_n] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu \\
E[X_1] &= \mu \\
Var(\bar{X}_n) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \sigma^2/n \\
Var(X_1) &= \sigma^2
\end{aligned}$$

Efficiency - Compare Unbiased Estimators by Variance

Let $\widehat{\theta}_1$ and $\widehat{\theta}_2$ be unbiased estimators of θ_0. We say that $\widehat{\theta}_1$ is *more efficient* than $\widehat{\theta}_2$ if $Var(\widehat{\theta}_1) < Var(\widehat{\theta}_2)$.

Mean-Squared Error

Except in very simple situations, unbiased estimators are hard to come by. In fact, in many interesting applications there is a *tradeoff* between bias and variance:

- Low bias estimators often have a high variance
- Low variance estimators often have high bias

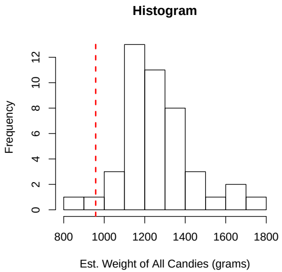
Mean-Squared Error (MSE): Squared Bias plus Variance

$$MSE(\widehat{\theta}) = \text{Bias}(\widehat{\theta})^2 + Var(\widehat{\theta})$$

Root Mean-Squared Error (RMSE): $\sqrt{\text{MSE}}$

Calculate MSE for Candy Experiment

$E[\hat{\theta}]$	1269 grams
θ_0	958 grams
$SD(\widehat{\theta})$	189 grams



$$
\begin{aligned}
\text{Bias} &= 1269 \text{ grams} - 958 \text{ grams} \\
&= 311 \text{ grams} \\
\text{MSE} &= \text{Bias}^2 + \text{Variance} \\
&= (311^2 + 189^2) \text{ grams}^2 \\
&= 1.32442 \times 10^5 \text{ grams}^2 \\
\text{RMSE} &= \sqrt{\text{MSE}} = 364 \text{ grams}
\end{aligned}
$$

Finite Sample versus Asymptotic Properties of Estimators

Finite Sample Properties

For *fixed sample size n* what are the properties of the sampling distribution of $\widehat{\theta}_n$? (E.g. bias and variance.)

Asymptotic Properties

What happens to the sampling distribution of $\widehat{\theta}_n$ *as the sample size n gets larger and larger?* (That is, $n \to \infty$).

Why Asymptotics?

Law of Large Numbers

Make precise what we mean by "bigger samples are better."

Central Limit Theorem

As $n \to \infty$ *pretty much any* sampling distribution is well-approximated by a normal random variable!

Consistency

Consistency

If an estimator $\widehat{\theta}_n$ (which is a RV) *converges* to θ_0 (a constant) as $n \to \infty$, we say that *$\widehat{\theta}_n$ is consistent for θ_0*.

What does it mean for a *RV* to converge to a *constant?*

For this course we'll use *MSE Consistency*:

$$\lim_{n \to \infty} \mathrm{MSE}(\widehat{\theta}_n) = 0$$

This makes sense since $\mathrm{MSE}(\widehat{\theta}_n)$ is a *constant*, so this is just an ordinary limit from calculus.

Law of Large Numbers (aka Law of Averages)

Let $X_1, X_2, \ldots X_n \sim$ *iid* mean μ, variance σ^2. Then the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is consistent for the population mean μ.

Law of Large Numbers (aka Law of Averages)

Let $X_1, X_2, \ldots X_n \sim iid$ mean μ, variance σ^2.

$$\begin{aligned}
E[\bar{X}_n] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu \\
Var(\bar{X}_n) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \sigma^2/n \\
\text{MSE}(\bar{X}_n) &= \text{Bias}(\bar{X}_n)^2 + Var(\bar{X}_n) \\
&= \left(E[\bar{X}_n] - \mu\right)^2 + Var(\bar{X}_n) \\
&= 0 + \sigma^2/n \\
&\to 0
\end{aligned}$$

Hence $\bar{X}_n$ is consistent for μ

Important!

An estimator *can* be biased but still consistent, as long as the bias disappears as $n \to \infty$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Bias of $\widehat{\sigma}^2$

$$E[\widehat{\sigma}^2] - \sigma^2 = \frac{(n-1)\sigma^2}{n} - \sigma^2 = -\sigma^2/n \to 0$$

Suppose $X_1, X_2, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$. What is the sampling distribution of $\bar{X}_n$?

(a) $\chi^2(n)$

(b) $t(n)$

(c) $F(n, n)$

(d) $N(\mu, \sigma^2/n)$

(e) Not enough information to determine.

But still, how can something random converge to something constant?

Sampling Distribution of $\bar{X}_n$ Collapses to μ

Look at an example where we can directly calculate not only the mean and variance of the sampling distribution of $\bar{X}_n$, but the *sampling distribution itself*:

$$X_1, X_2, \ldots, X_n \sim \text{iid } N(\mu, \sigma^2) \Rightarrow \bar{X}_n \sim N(\mu, \sigma^2/n)$$

Sampling Distribution of $\bar{X}_n$ Collapses to μ

$X_1, X_2, \ldots, X_n \sim$ iid $N(\mu, \sigma^2 \Rightarrow \bar{X}_n \sim N(\mu, \sigma^2/n)$.

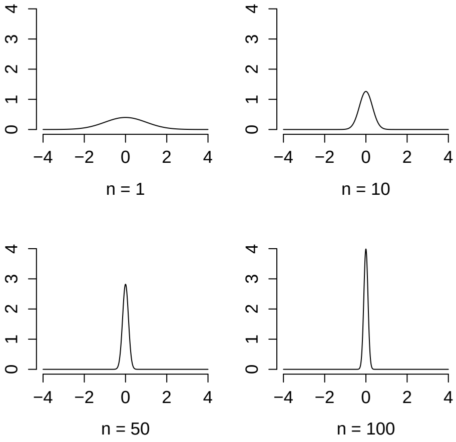


Figure : Sampling Distributions for $\bar{X}_n$ where $X_i \sim$ iid $N(0, 1)$

Another Visualization: Keep Adding Observations

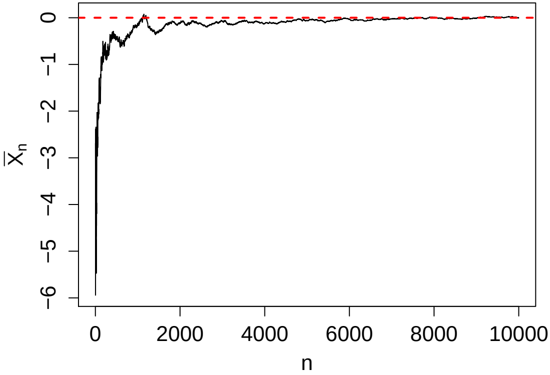


Figure : Running sample means: $X_i \sim$ iid $N(0, 100)$

n	$\bar{X}_n$
1	-2.69
2	-3.18
3	-5.94
4	-4.27
5	-2.62
10	-2.89
20	-5.33
50	-2.94
100	-1.58
500	-0.45
1000	-0.13
5000	-0.05
10000	0.00

Important!

Although I showed two examples involving normal RVs, the LLN holds IN GENERAL!