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 $E[\widehat{\theta}_n] = \theta_0$

A Different Estimator of the Population Variance

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

A Different Estimator of the Population Variance

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

$E[\widehat{\sigma}^2] =$

A Different Estimator of the Population Variance

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

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

A Different Estimator of the Population Variance

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

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

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 $\widehat{\sigma}^2$

$E[\widehat{\sigma}^2] - \sigma^2 =$

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 $\widehat{\sigma}^2$

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

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 $\widehat{\sigma}^2$

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

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 $\widehat{\sigma}^2$

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

Procedure versus Result of the Procedure

Procedure versus Result of the Procedure

Procedure = Random Variable

Procedure versus Result of the Procedure

Procedure = Random Variable

- $X_1, \ldots, X_n$ represents procedure of taking a random sample.

Procedure versus Result of the Procedure

Procedure = Random Variable

- $X_1, \ldots, X_n$ represents procedure of taking a random sample.
- $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ represents procedure of taking sample mean

Procedure versus Result of the Procedure

Procedure = Random Variable

- $X_1, \ldots, X_n$ represents procedure of taking a random sample.
- $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ represents procedure of taking sample mean

Sampling Dist. = Probabilistic Behavior of Procedure

Procedure versus Result of the Procedure

Procedure = Random Variable

- $X_1, \ldots, X_n$ represents procedure of taking a random sample.
- $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ represents procedure of taking sample mean

Sampling Dist. = Probabilistic Behavior of Procedure

If I repeat the procedure of taking the mean of a random sample over and over for many samples, what relative frequencies do I get for the sample means?

Procedure versus Result of the Procedure

Procedure = Random Variable

- $X_1, \ldots, X_n$ represents procedure of taking a random sample.
- $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ represents procedure of taking sample mean

Sampling Dist. = Probabilistic Behavior of Procedure

If I repeat the procedure of taking the mean of a random sample over and over for many samples, what relative frequencies do I get for the sample means?

Result of Procedure = Constant

Procedure versus Result of the Procedure

Procedure = Random Variable

- $X_1, \ldots, X_n$ represents procedure of taking a random sample.
- $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ represents procedure of taking sample mean

Sampling Dist. = Probabilistic Behavior of Procedure

If I repeat the procedure of taking the mean of a random sample over and over for many samples, what relative frequencies do I get for the sample means?

Result of Procedure = Constant

- $x_1, \ldots, x_n$ is the result of sampling, the observed data.

Procedure versus Result of the Procedure

Procedure = Random Variable

- $X_1, \ldots, X_n$ represents procedure of taking a random sample.
- $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ represents procedure of taking sample mean

Sampling Dist. = Probabilistic Behavior of Procedure

If I repeat the procedure of taking the mean of a random sample over and over for many samples, what relative frequencies do I get for the sample means?

Result of Procedure = Constant

- $x_1, \ldots, x_n$ is the result of sampling, the observed data.
- $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the result of taking sample mean

Procedure? Long-Run Relative Freqencies?

Why would I advise you not to play the lottery?

Procedure? Long-Run Relative Freqencies?

Why would I advise you not to play the lottery?

- You may sometimes win, but if you play the lottery many times, on average you will lose money.

Procedure? Long-Run Relative Freqencies?

Why would I advise you not to play the lottery?

- You may sometimes win, but if you play the lottery many times, on average you will lose money.
- Let X be a random variable representing lottery winnings. I am arguing that $E[X]$ - Cost of Ticket < 0

Procedure? Long-Run Relative Freqencies?

Why would I advise you not to play the lottery?

- You may sometimes win, but if you play the lottery many times, on average you will lose money.
- Let X be a random variable representing lottery winnings. I am arguing that $E[X]$ - Cost of Ticket < 0

Procedure = Random Variable

Making a habit of playing the lottery. Expectation is negative.

Procedure? Long-Run Relative Freqencies?

Why would I advise you not to play the lottery?

- You may sometimes win, but if you play the lottery many times, on average you will lose money.
- Let X be a random variable representing lottery winnings. I am arguing that $E[X]$ - Cost of Ticket < 0

Procedure = Random Variable

Making a habit of playing the lottery. Expectation is negative.

Result of that Procedure = Constant

How much you win in a *particular* lottery. Could be greater than or less than cost of ticket in any *individual* instance.

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

Choose n Students from Class List with Replacement

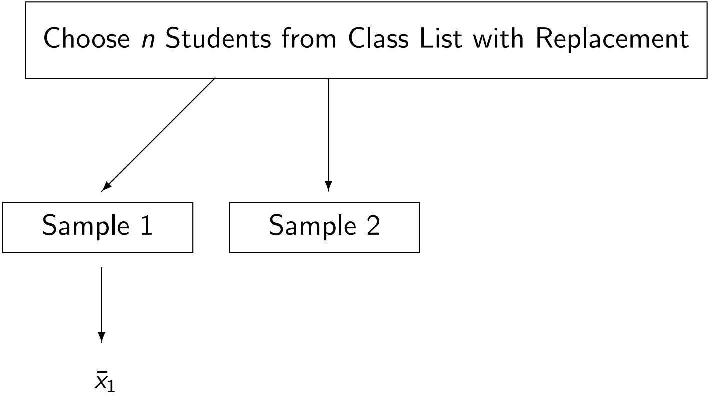
Sampling Distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$\begin{array}{c}\boxed{\text{Choose } n \text{ Students from Class List with Replacement}}\\ \swarrow \\ \boxed{\text{Sample 1}}\end{array}$$

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

$$
\begin{tikzcd}
\boxed{\text{Choose } n \text{ Students from Class List with Replacement}} \arrow[d] \\
\boxed{\text{Sample 1}} \arrow[d] \\
\bar{x}_1
\end{tikzcd}
$$

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



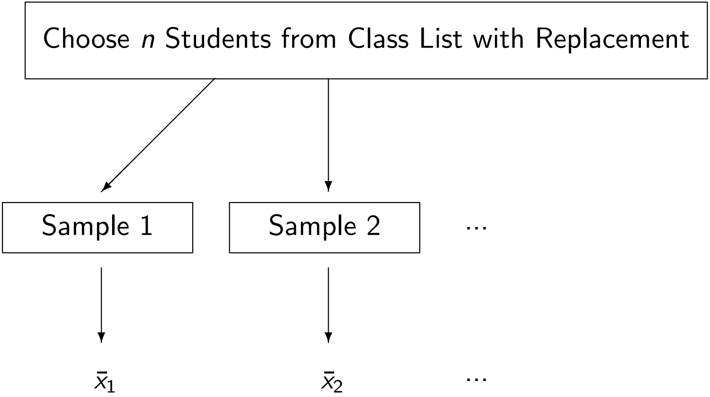
Sampling Distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

Choose n Students from Class List with Replacement

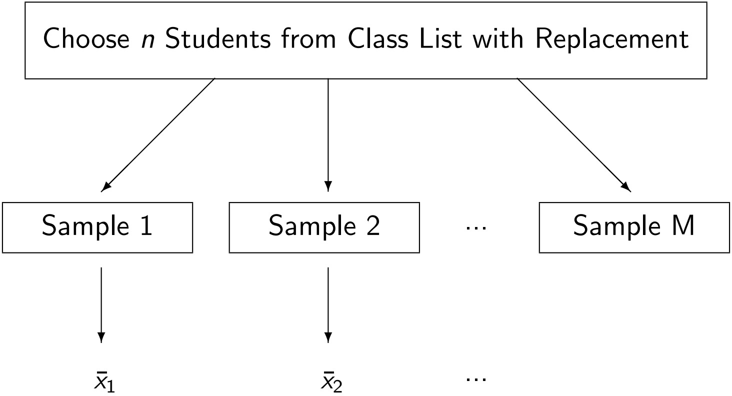
Sample 1 | Sample 2

$\bar{x}_1$ | $\bar{x}_2$

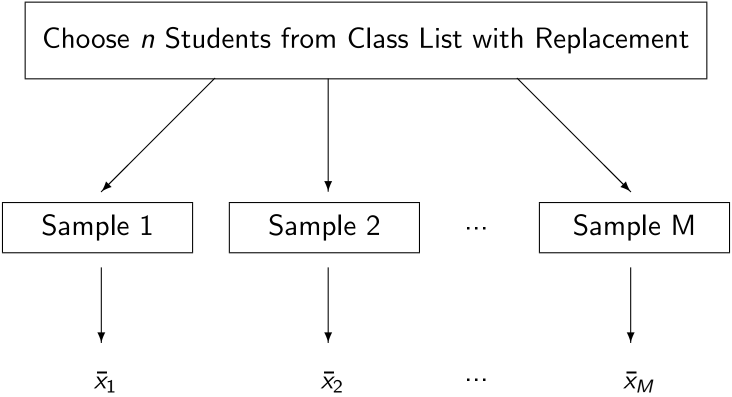
Sampling Distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$



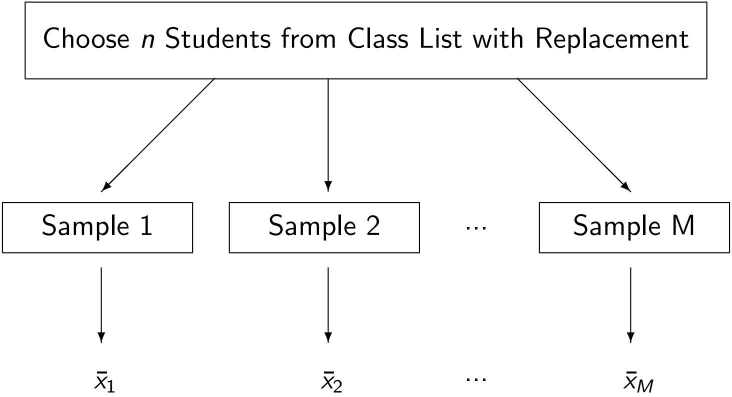
Sampling Distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$



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

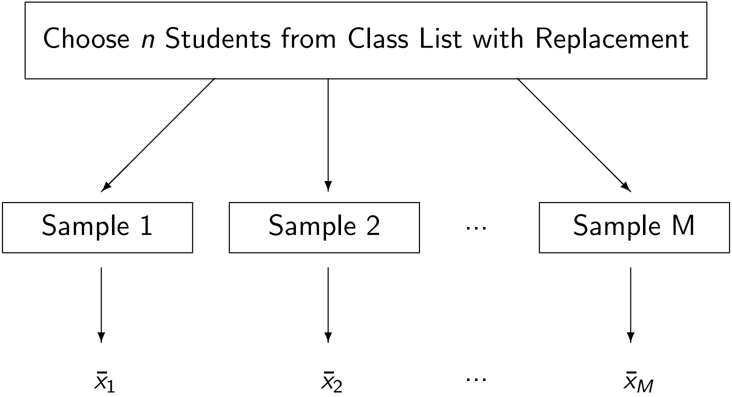


Sampling Distribution of $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$



Repeat M times $\rightarrow$ get M different sample means

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



Repeat M times $\rightarrow$ get M different sample means

Sampling Dist: long run relative frequencies of the $\bar{x}_i$

Height of Econ 103 Students

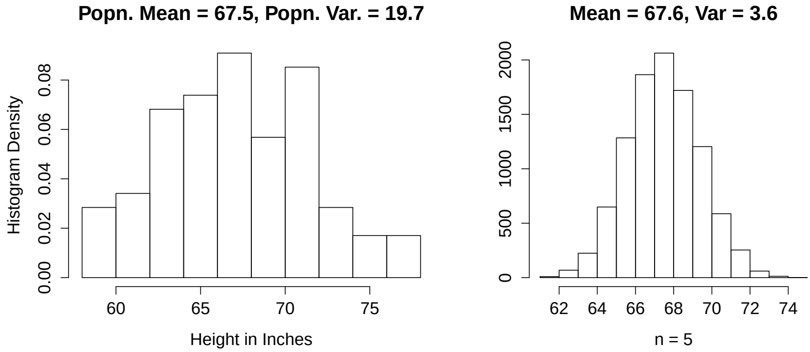


Figure : Left: Population, Right: Sampling distribution of $\bar{X}_5$

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?

What's Going On Here?

Biased Sample!

What's Going On Here?

Biased Sample!

- Zero children $\Rightarrow$ didn't send any to college

What's Going On Here?

Biased Sample!

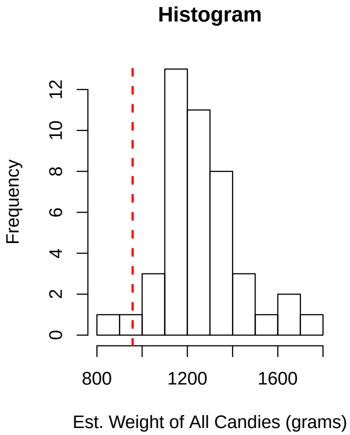
- Zero children $\Rightarrow$ didn't send any to college
- Sampling by *children* so large families oversampled

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: θ_0 =958 grams



What was in the bag?

100 Candies Total:

What was in the bag?

100 Candies Total:

- 20 Fun Size Snickers Bars (large)

What was in the bag?

100 Candies Total:

- 20 Fun Size Snickers Bars (large)
- 30 Reese's Miniatures (medium)

What was in the bag?

100 Candies Total:

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

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?

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

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

X_1 is an unbiased estimator of μ

(a) True

(b) False

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

$$X_1 \textit{ is an unbiased estimator of } \mu$$

(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

$$E[\bar{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu$$

How to choose between two unbiased estimators?

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

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

How to choose between two unbiased estimators?

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

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

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:

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

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

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):

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

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

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:

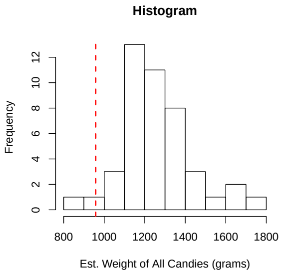
- Low bias estimators often have a high variance
- Low variance estimators often have high bias

Mean Squared Error (MSE): Squared Bias plus Variance

$$\begin{aligned} MSE(\widehat{\theta}) &= \text{Bias}(\widehat{\theta})^2 + Var(\widehat{\theta}) \\ &= \left(E[\widehat{\theta}] - \theta_0\right)^2 + Var(\widehat{\theta}) \end{aligned}$$

Calculate MSE for Candy Experiment

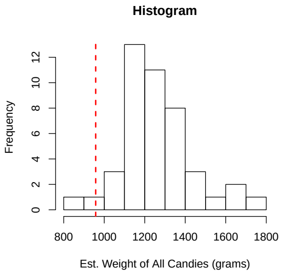
$E[\hat{\theta}]$	1269 grams
θ_0	958 grams
$SD(\widehat{\theta})$	189 grams



$$\text{Bias} \quad =$$

Calculate MSE for Candy Experiment

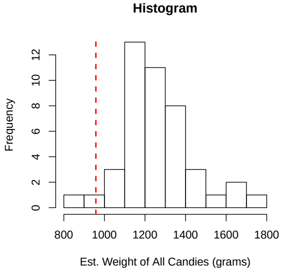
$E[\hat{\theta}]$	1269 grams
θ_0	958 grams
$SD(\widehat{\theta})$	189 grams



$$
\begin{aligned}
\text{Bias} &= 1269 \text{ grams} - 958 \text{ grams} \\
&=
\end{aligned}
$$

Calculate MSE for Candy Experiment

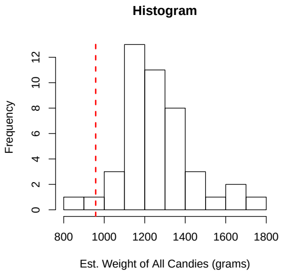
$E[\hat{\theta}]$	1269 grams
θ_0	958 grams
$SD(\widehat{\theta})$	189 grams



$$\begin{aligned} \text{Bias} &= 1269 \text{ grams} - 958 \text{ grams} \\ &= \color{red}{311 \text{ grams}} \\ \text{MSE} &= \end{aligned}$$

Calculate MSE for Candy Experiment

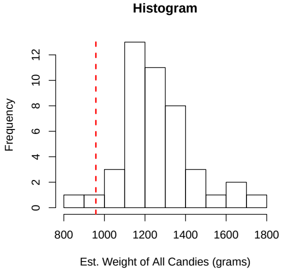
$E[\hat{\theta}]$	1269 grams
θ_0	958 grams
$SD(\widehat{\theta})$	189 grams



$$
\begin{aligned}
\text{Bias} &= 1269 \text{ grams} - 958 \text{ grams} \\
&= \color{red}{311 \text{ grams}} \\
\text{MSE} &= \text{Bias}^2 + \text{Variance} \\
&=
\end{aligned}
$$

Calculate MSE for Candy Experiment

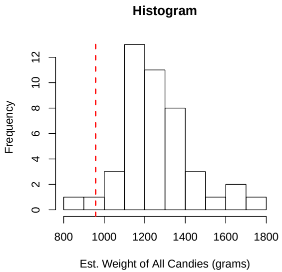
$E[\hat{\theta}]$	1269 grams
θ_0	958 grams
$SD(\widehat{\theta})$	189 grams



$$
\begin{aligned}
\text{Bias} &= 1269 \text{ grams} - 958 \text{ grams} \\
&= {\color{red}311 \text{ grams}} \\
\text{MSE} &= \text{Bias}^2 + \text{Variance} \\
&= (311^2 + 189^2) \text{ grams}^2 \\
&=
\end{aligned}
$$

Calculate MSE for Candy Experiment

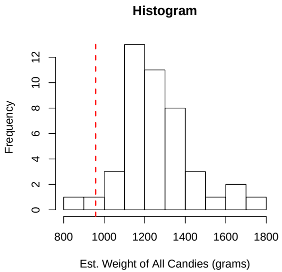
$E[\hat{\theta}]$	1269 grams
θ_0	958 grams
$SD(\widehat{\theta})$	189 grams



$$\begin{aligned}
\text{Bias} &= 1269 \text{ grams} - 958 \text{ grams} \\
&= 311 \text{ grams} \\
\text{MSE} &= \text{Bias}^2 + \text{Variance} \\
&= (311^2 + 189^2) \text{ grams}^2 \\
&= 1.3244 \times 10^5 \text{ grams}^2
\end{aligned}$$

Calculate MSE for Candy Experiment

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



$$
\begin{aligned}
\text{Bias} &= 1269 \text{ grams} - 958 \text{ grams} \\
&= 311 \text{ grams} \\
\text{MSE} &= \text{Bias}^2 + \text{Variance} \\
&= (311^2 + 189^2) \text{ grams}^2 \\
&= 1.3244 \times 10^5 \text{ grams}^2 \\
\text{RMSE} &= \sqrt{\text{MSE}} = 364 \text{ grams}
\end{aligned}
$$

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.)

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."

Why Asymptotics?

Law of Large Numbers

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

Central Limit Theorem

As $n \to \infty$ *nearly all* sampling distributions behave behave like 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*:

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

This makes sense since $\text{MSE}(\widehat{\theta}_n)$ 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.

$$E[\bar{X}_n] \quad =$$

Law of Large Numbers (aka Law of Averages)

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

$$E[\bar{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu$$

Law of Large Numbers (aka Law of Averages)

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

$$E[\bar{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu$$

$$Var(\bar{X}_n) =$$

Law of Large Numbers (aka Law of Averages)

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

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

Law of Large Numbers (aka Law of Averages)

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

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

Law of Large Numbers (aka Law of Averages)

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

$$E[\bar{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu$$

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

$$\begin{aligned}\mathrm{MSE}(\bar{X}_n) &= \mathrm{Bias}(\bar{X}_n)^2 + Var(\bar{X}_n) \\ &= \left(E[\bar{X}_n] - \mu\right)^2 + Var(\bar{X}_n)\end{aligned}$$

Law of Large Numbers (aka Law of Averages)

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

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

Law of Large Numbers (aka Law of Averages)

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

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

Hence $\bar{X}_n$ is consistent for μ

Important!

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

Important!

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

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

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 $\widehat{\sigma}^2$

$E[\widehat{\sigma}^2] - \sigma^2 =$

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 $\widehat{\sigma}^2$

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

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 $\widehat{\sigma}^2$

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

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

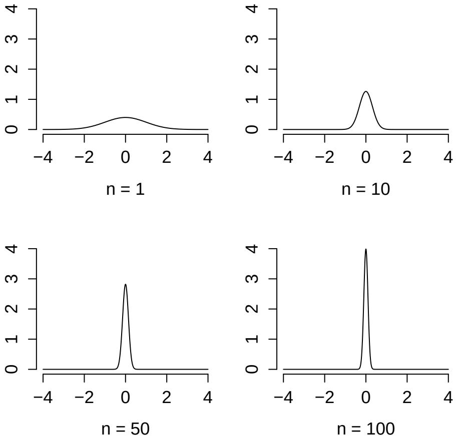


Figure : Sampling Distributions for $\bar{X}_n$ where $X_i \sim$ iid $N(0, 1)$

Another Visualization: Keep Adding Observations

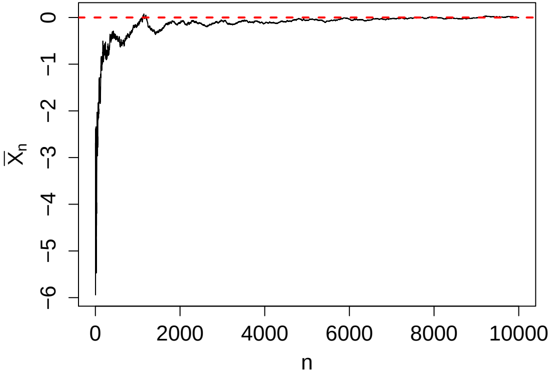


Figure : Running sample means: $X_i \sim$ iid $N(0, 100)$

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!