

# Asymptotics and consistency

EC 421, Set 6

Edward Rubin

30 January 2019

# Prologue

# Schedule

## Last Time

Living with heteroskedasticity

## Today

Asymptotics and consistency

## This week

- Our second assignment
- Survey (for credit)

## Near-ish future

Midterm on Feb. 12<sup>th</sup>

# R showcase

Need speed? R allows essentially infinite parallelization.

Three popular packages:

- `future`
- `parallel`
- `foreach`

And here's a nice [tutorial](#).

# Consistency

# Consistency

## Welcome to asymptopia

*Previously:* We examined estimators (e.g.,  $\hat{\beta}_j$ ) and their properties using

# Consistency

## Welcome to asymptopia

Previously: We examined estimators (e.g.,  $\hat{\beta}_j$ ) and their properties using

1. The **mean** of the estimator's distribution:  $E[\hat{\beta}_j] = ?$

# Consistency

## Welcome to asymptopia

Previously: We examined estimators (e.g.,  $\hat{\beta}_j$ ) and their properties using

1. The **mean** of the estimator's distribution:  $E[\hat{\beta}_j] = ?$
2. The **variance** of the estimator's distribution:  $\text{Var}(\hat{\beta}_j) = ?$

# Consistency

## Welcome to asymptopia

Previously: We examined estimators (e.g.,  $\hat{\beta}_j$ ) and their properties using

1. The **mean** of the estimator's distribution:  $E[\hat{\beta}_j] = ?$
2. The **variance** of the estimator's distribution:  $\text{Var}(\hat{\beta}_j) = ?$

which tell us about the **tendency of the estimator** if we took  **$\infty$  samples**, each with **sample size  $n$** .

# Consistency

## Welcome to asymptopia

Previously: We examined estimators (e.g.,  $\hat{\beta}_j$ ) and their properties using

1. The **mean** of the estimator's distribution:  $E[\hat{\beta}_j] = ?$
2. The **variance** of the estimator's distribution:  $\text{Var}(\hat{\beta}_j) = ?$

which tell us about the **tendency of the estimator** if we took  **$\infty$  samples**, each with **sample size  $n$** .

This approach misses something.

# Consistency

Welcome to asymptopia

**New question:**

How does our estimator behave as our sample gets larger (as  $n \rightarrow \infty$ )?

# Consistency

## Welcome to asymptopia

### New question:

How does our estimator behave as our sample gets larger (as  $n \rightarrow \infty$ )?

This *new question* forms a new way to think about the properties of estimators: **asymptotic properties** (or large-sample properties).

# Consistency

## Welcome to asymptopia

### New question:

How does our estimator behave as our sample gets larger (as  $n \rightarrow \infty$ )?

This *new question* forms a new way to think about the properties of estimators: **asymptotic properties** (or large-sample properties).

A "good" estimator will become indistinguishable from the parameter it estimates when  $n$  is very large (close to  $\infty$ ).

# Consistency

## Probability limits

Just as the *expected value* helped us characterize **the finite-sample distribution of an estimator** with sample size  $n$ ,

# Consistency

## Probability limits

Just as the *expected value* helped us characterize **the finite-sample distribution of an estimator** with sample size  $n$ ,

the *probability limit* helps us analyze **the asymptotic distribution of an estimator** (the distribution of the estimator as  $n$  gets "big"<sup>†</sup>).

<sup>†</sup>: Here, "big"  $n$  means  $n \rightarrow \infty$ . That's *really* big data.

# Consistency

## Probability limits

Let  $B_n$  be our estimator with sample size  $n$ .

Then the **probability limit** of  $B$  is  $\alpha$  if

$$\lim_{n \rightarrow \infty} P(|B_n - \alpha| > \epsilon) = 0 \quad (1)$$

for any  $\epsilon > 0$ .

# Consistency

## Probability limits

Let  $B_n$  be our estimator with sample size  $n$ .

Then the **probability limit** of  $B$  is  $\alpha$  if

$$\lim_{n \rightarrow \infty} P(|B_n - \alpha| > \epsilon) = 0 \quad (1)$$

for any  $\epsilon > 0$ .

The definition in (1) *essentially* says that as the **sample size** approaches infinity, the probability that  $B_n$  differs from  $\alpha$  by more than a very small number ( $\epsilon$ ) is zero.

# Consistency

## Probability limits

Let  $B_n$  be our estimator with sample size  $n$ .

Then the **probability limit** of  $B$  is  $\alpha$  if

$$\lim_{n \rightarrow \infty} P(|B_n - \alpha| > \epsilon) = 0 \quad (1)$$

for any  $\epsilon > 0$ .

The definition in (1) *essentially* says that as the **sample size** approaches infinity, the probability that  $B_n$  differs from  $\alpha$  by more than a very small number ( $\epsilon$ ) is zero.

*Practically:*  $B$ 's distribution collapses to a spike at  $\alpha$  as  $n$  approaches  $\infty$ .

# Consistency

## Probability limits

Equivalent statements:

- The probability limit of  $B_n$  is  $\alpha$ .
- $\text{plim } B = \alpha$
- $B$  converges in probability to  $\alpha$ .

# Consistency

## Probability limits

Probability limits have some nice/important properties:

- $\text{plim}(X \times Y) = \text{plim}(X) \times \text{plim}(Y)$
- $\text{plim}(X + Y) = \text{plim}(X) + \text{plim}(Y)$
- $\text{plim}(c) = c$ , where  $c$  is a constant
- $\text{plim}\left(\frac{X}{Y}\right) = \frac{\text{plim}(X)}{\text{plim}(Y)}$
- $\text{plim}(f(X)) = f(\text{plim}(X))$

# Consistency

## Consistent estimators

We say that **an estimator is consistent** if

1. The estimator **has a prob. limit** (its distribution collapses to a spike).
2. This spike is **located at the parameter** the estimator predicts.

# Consistency

## Consistent estimators

We say that **an estimator is consistent** if

1. The estimator **has a prob. limit** (its distribution collapses to a spike).
2. This spike is **located at the parameter** the estimator predicts.

*In other words...*

An estimator is consistent if its asymptotic distribution collapses to a spike located at the estimated parameter.

# Consistency

## Consistent estimators

We say that **an estimator is consistent** if

1. The estimator **has a prob. limit** (its distribution collapses to a spike).
2. This spike is **located at the parameter** the estimator predicts.

*In other words...*

An estimator is consistent if its asymptotic distribution collapses to a spike located at the estimated parameter.

*In math:* The estimator  $B$  is consistent for  $\alpha$  if  $\text{plim } B = \alpha$ .

# Consistency

## Consistent estimators

We say that **an estimator is consistent** if

1. The estimator **has a prob. limit** (its distribution collapses to a spike).
2. This spike is **located at the parameter** the estimator predicts.

*In other words...*

An estimator is consistent if its asymptotic distribution collapses to a spike located at the estimated parameter.

*In math:* The estimator  $B$  is consistent for  $\alpha$  if  $\text{plim } B = \alpha$ .

The estimator is *inconsistent* if  $\text{plim } B \neq \alpha$ .

# Consistency

## Consistent estimators

*Example:* We want to estimate the population mean  $\mu_x$  (where  $X \sim \text{Normal}$ ).

Let's compare the asymptotic distributions of two competing estimators:

1. The first observation:  $X_1$
2. The sample mean:  $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$
3. Some other estimator:  $\widetilde{X} = \frac{1}{n+1} \sum_{i=1}^n x_i$

Note that (1) and (2) are unbiased, but (3) is biased.

# Consistency

## Consistent estimators

To see which are unbiased/biased:

# Consistency

## Consistent estimators

To see which are unbiased/biased:

$$E[X_1] = \mu_x$$

# Consistency

## Consistent estimators

To see which are unbiased/biased:

$$E[X_1] = \mu_x$$

$$E[\bar{X}]$$

# Consistency

## Consistent estimators

To see which are unbiased/biased:

$$E[X_1] = \mu_x$$

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

# Consistency

## Consistent estimators

To see which are unbiased/biased:

$$E[X_1] = \mu_x$$

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

# Consistency

## Consistent estimators

To see which are unbiased/biased:

$$E[X_1] = \mu_x$$

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} \sum_{i=1}^n \mu_x$$

# Consistency

## Consistent estimators

To see which are unbiased/biased:

$$E[X_1] = \mu_x$$

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} \sum_{i=1}^n \mu_x = \mu_x$$

# Consistency

## Consistent estimators

To see which are unbiased/biased:

$$E[X_1] = \mu_x$$

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} \sum_{i=1}^n \mu_x = \mu_x$$

$$E[\tilde{X}]$$

# Consistency

## Consistent estimators

To see which are unbiased/biased:

$$E[X_1] = \mu_x$$

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} \sum_{i=1}^n \mu_x = \mu_x$$

$$E[\tilde{X}] = E\left[\frac{1}{n+1} \sum_{i=1}^n x_i\right]$$

# Consistency

## Consistent estimators

To see which are unbiased/biased:

$$E[X_1] = \mu_x$$

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} \sum_{i=1}^n \mu_x = \mu_x$$

$$E[\tilde{X}] = E\left[\frac{1}{n+1} \sum_{i=1}^n x_i\right] = \frac{1}{n+1} \sum_{i=1}^n E[x_i]$$

# Consistency

## Consistent estimators

To see which are unbiased/biased:

$$E[X_1] = \mu_x$$

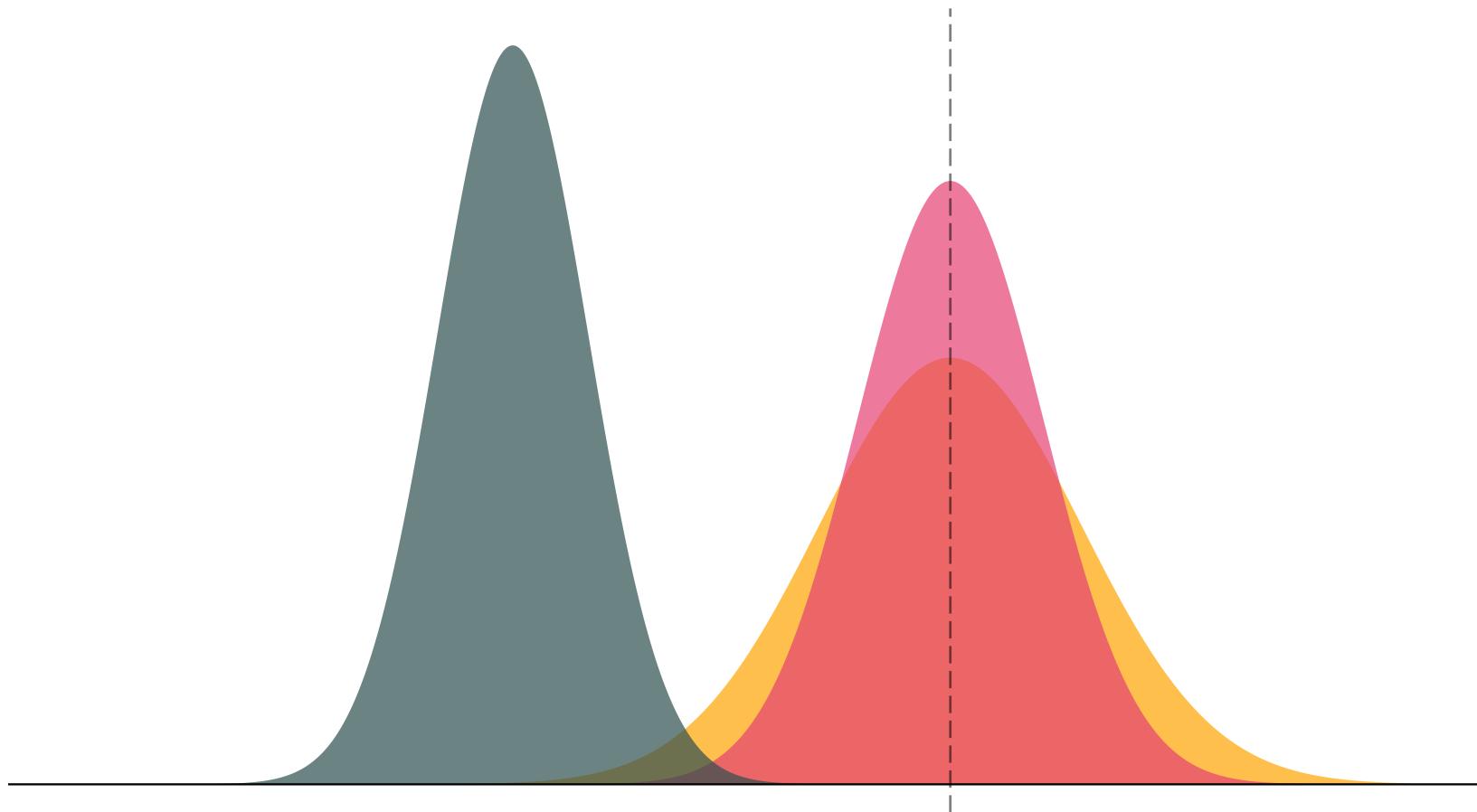
$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} \sum_{i=1}^n \mu_x = \mu_x$$

$$E[\tilde{X}] = E\left[\frac{1}{n+1} \sum_{i=1}^n x_i\right] = \frac{1}{n+1} \sum_{i=1}^n E[x_i] = \frac{n}{n+1} \mu_x$$

# Consistency

Distributions of  $X_1$ ,  $\bar{X}$ , and  $\tilde{X}$

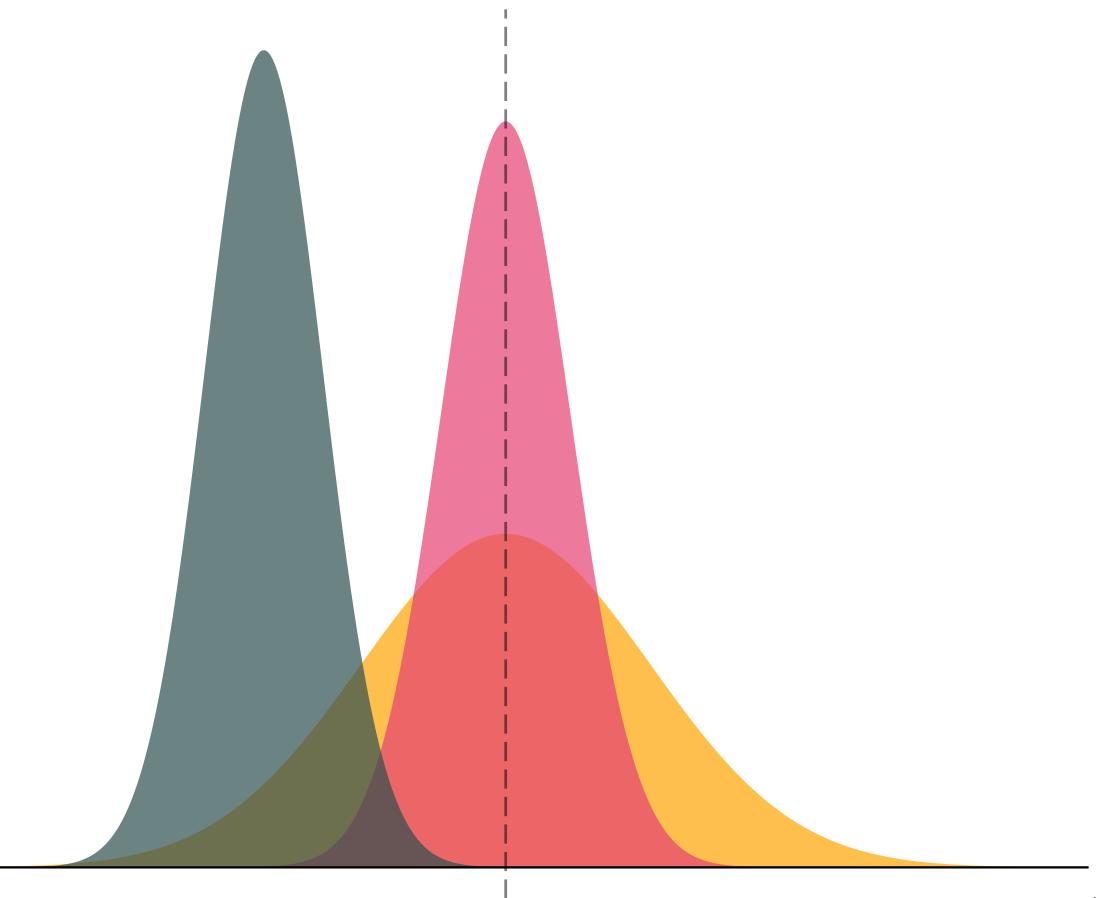
$n = 2$



# Consistency

Distributions of  $X_1$ ,  $\bar{X}$ , and  $\widetilde{X}$

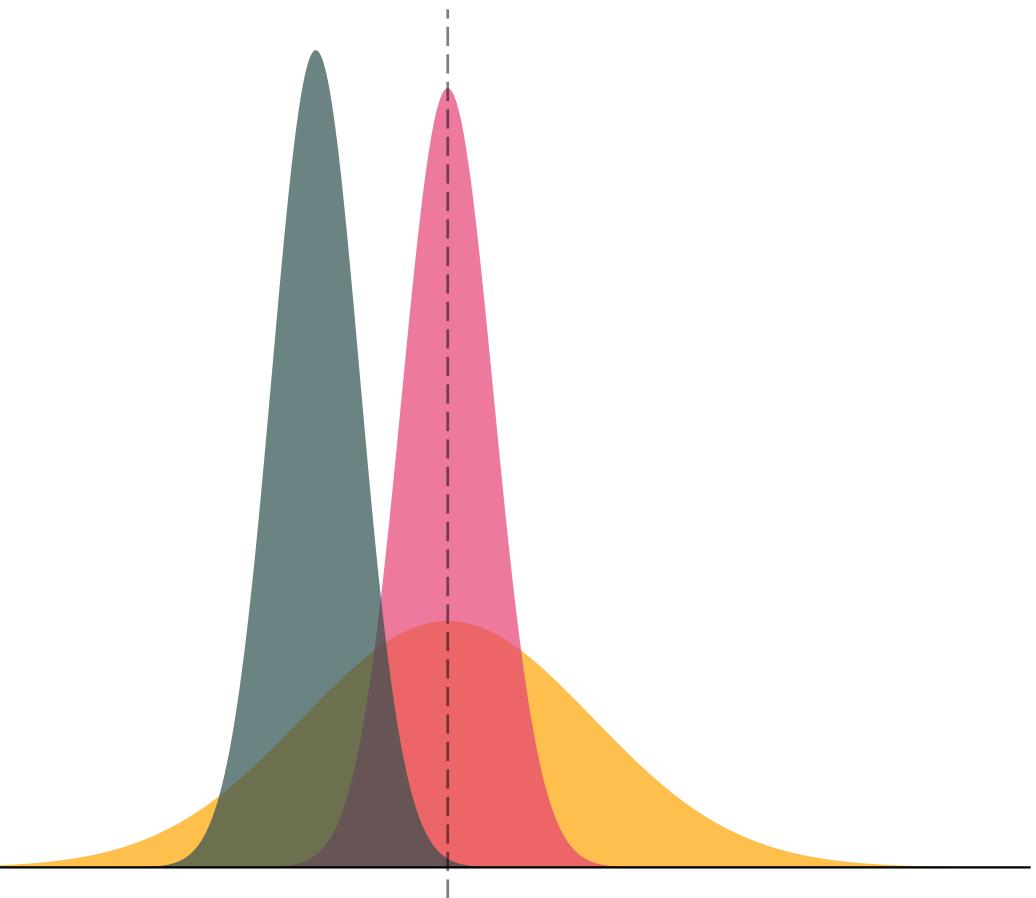
$n = 5$



# Consistency

Distributions of  $X_1$ ,  $\bar{X}$ , and  $\widetilde{X}$

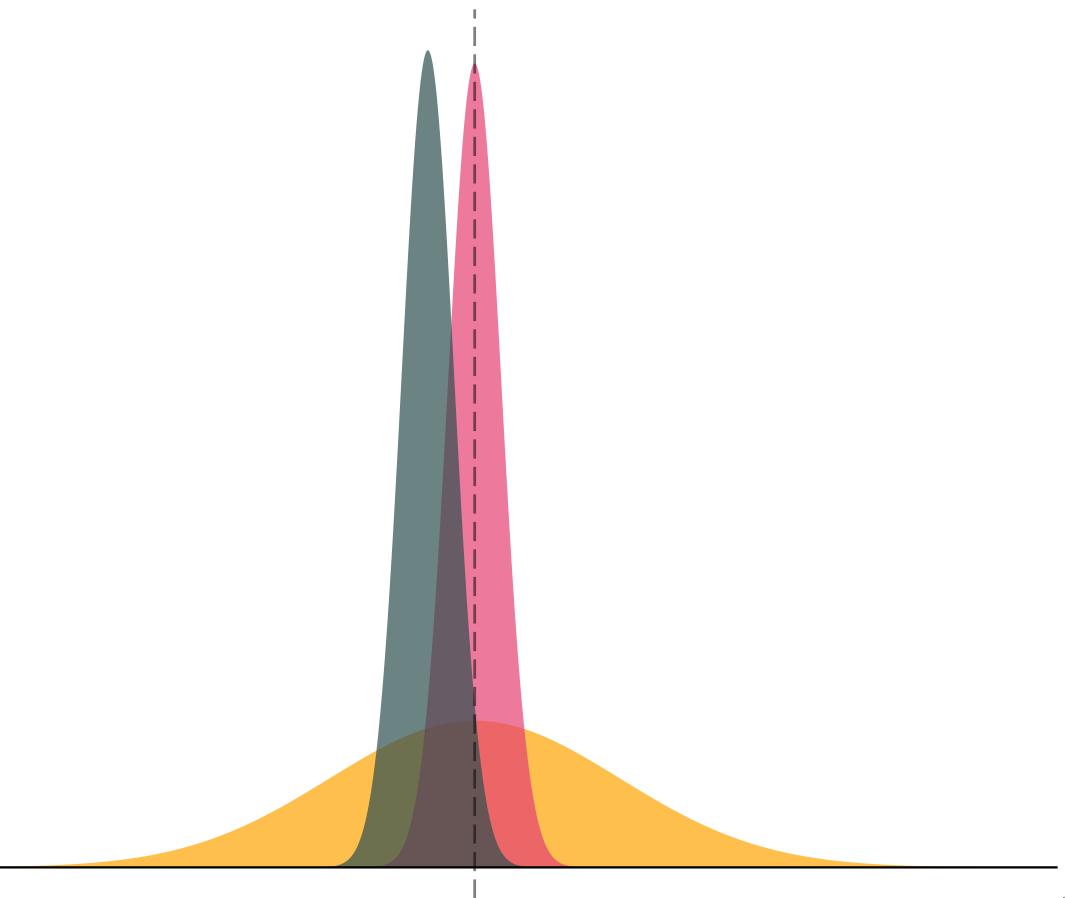
$n = 10$



# Consistency

Distributions of  $X_1$ ,  $\bar{X}$ , and  $\widetilde{X}$

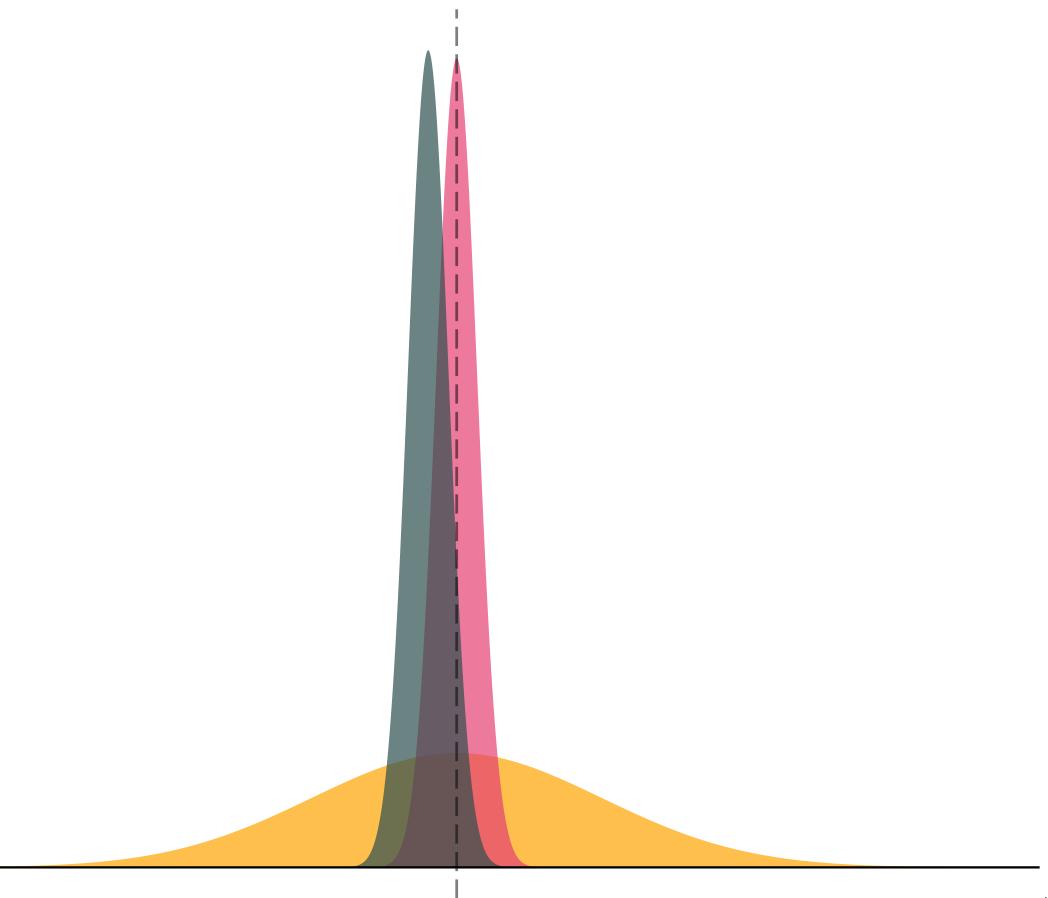
$n = 30$



# Consistency

Distributions of  $X_1$ ,  $\bar{X}$ , and  $\widetilde{X}$

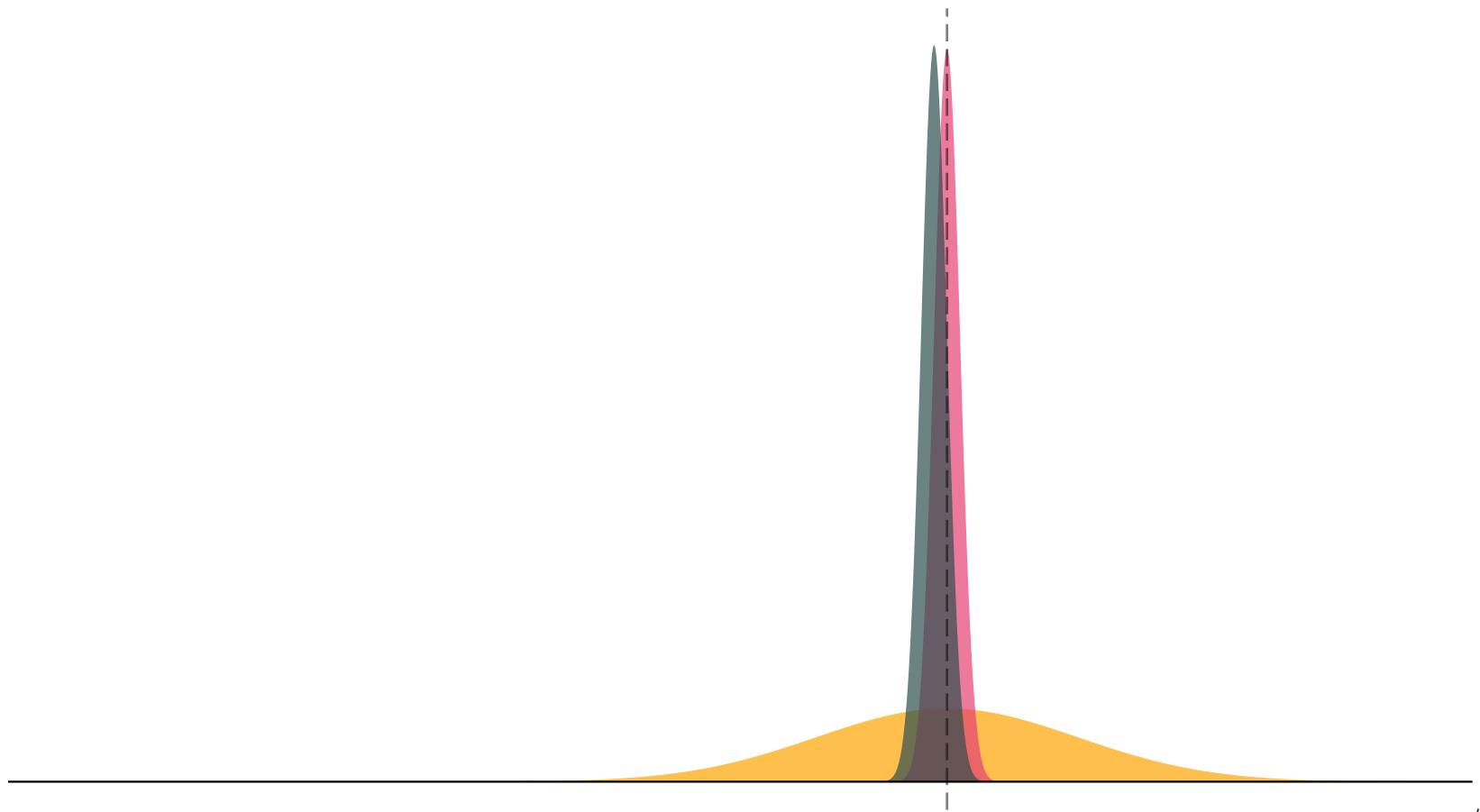
$n = 50$



# Consistency

Distributions of  $X_1$ ,  $\bar{X}$ , and  $\widetilde{X}$

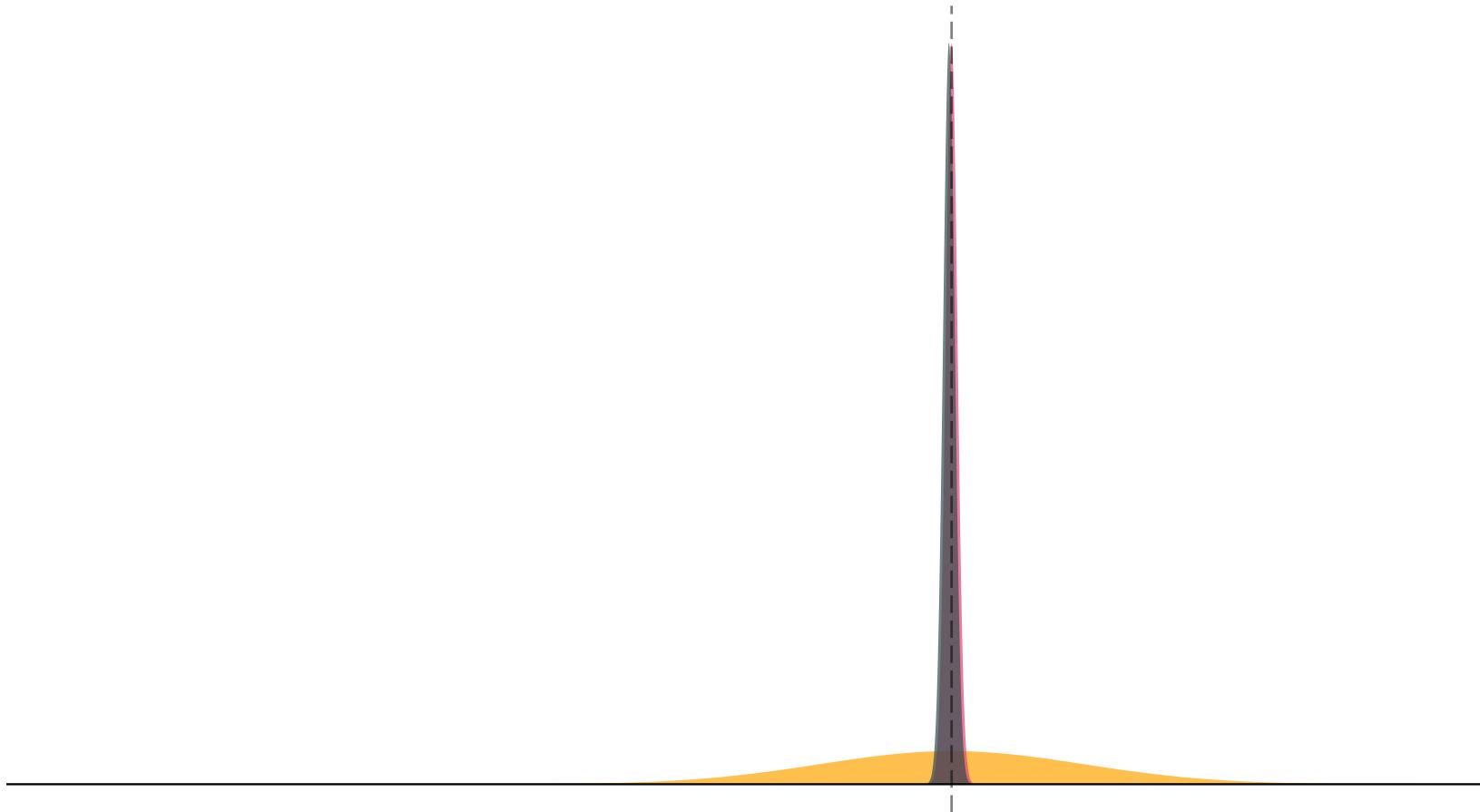
$n = 100$



# Consistency

Distributions of  $X_1$ ,  $\bar{X}$ , and  $\widetilde{X}$

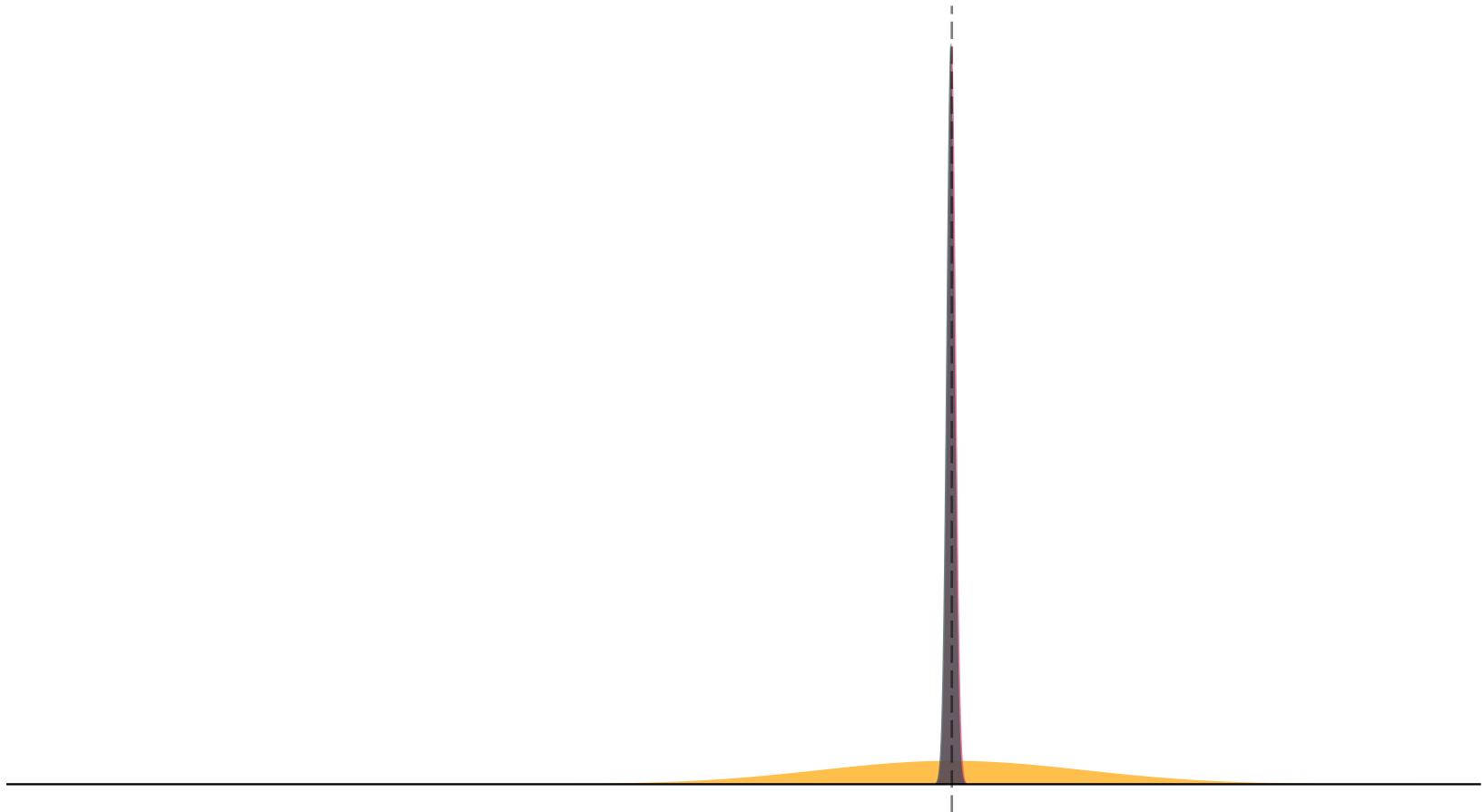
$n = 500$



# Consistency

Distributions of  $X_1$ ,  $\bar{X}$ , and  $\widetilde{X}$

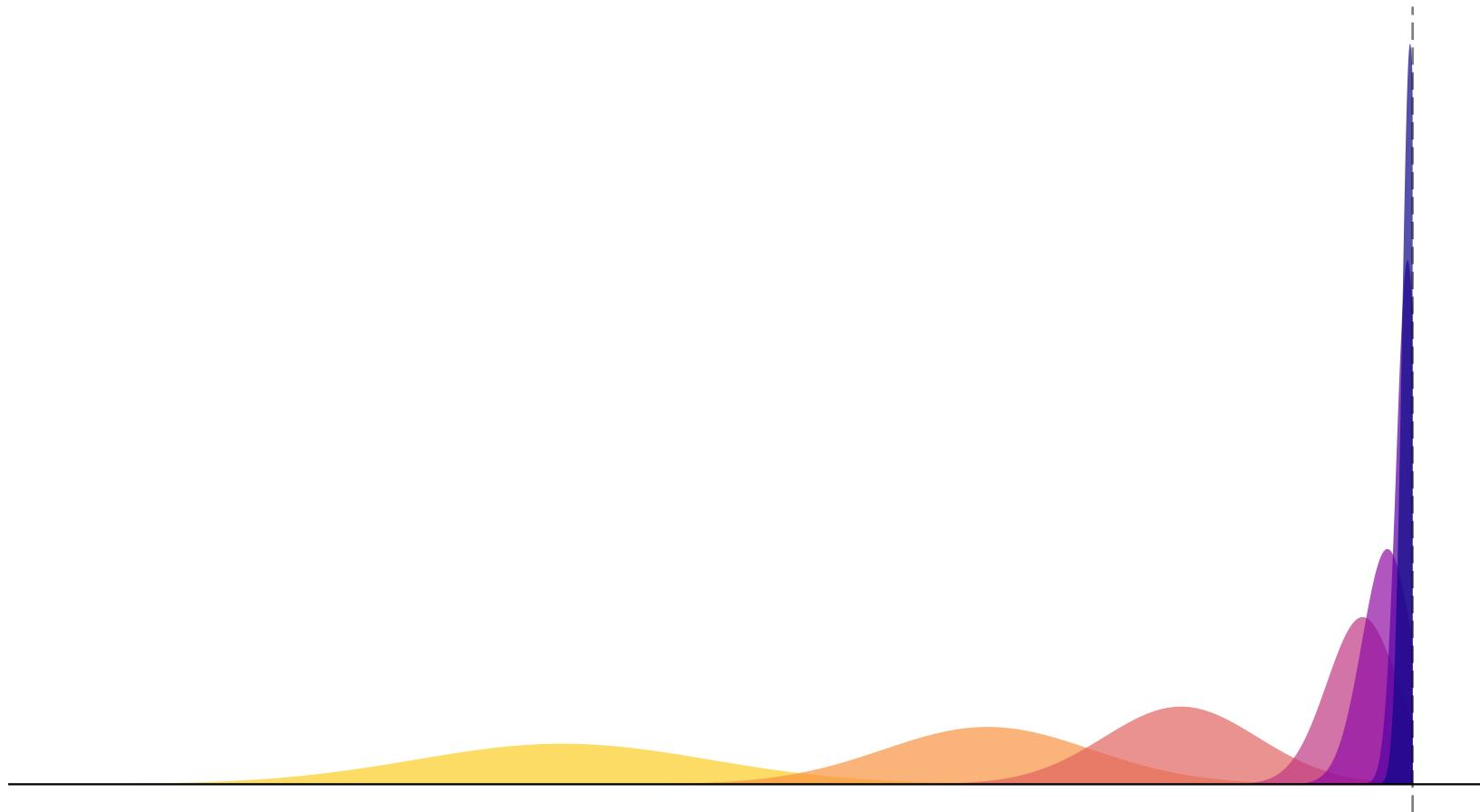
$n = 1000$



# Consistency

The distributions of  $\widetilde{X}$

For  $n$  in {2, 5, 10, 50, 100, 500, 1000}



# Consistency

The takeaway?

# Consistency

## The takeaway?

- An estimator can be unbiased without being consistent

# Consistency

## The takeaway?

- An estimator can be unbiased without being consistent (e.g.,  $\textcolor{blue}{X}_1$ ).

# Consistency

## The takeaway?

- An estimator can be unbiased without being consistent (e.g.,  $\textcolor{blue}{X}_1$ ).
- An estimator can be unbiased and consistent

# Consistency

## The takeaway?

- An estimator can be unbiased without being consistent (e.g.,  $\textcolor{orange}{X}_1$ ).
- An estimator can be unbiased and consistent (e.g.,  $\overline{\textcolor{red}{X}}$ ).

# Consistency

## The takeaway?

- An estimator can be unbiased without being consistent (e.g.,  $\textcolor{blue}{X}_1$ ).
- An estimator can be unbiased and consistent (e.g.,  $\overline{\textcolor{red}{X}}$ ).
- An estimator can be biased but consistent

# Consistency

## The takeaway?

- An estimator can be unbiased without being consistent (e.g.,  $\textcolor{blue}{X}_1$ ).
- An estimator can be unbiased and consistent (e.g.,  $\overline{\textcolor{red}{X}}$ ).
- An estimator can be biased but consistent (e.g.,  $\widetilde{X}$ ).

# Consistency

## The takeaway?

- An estimator can be unbiased without being consistent (e.g.,  $\textcolor{blue}{X}_1$ ).
- An estimator can be unbiased and consistent (e.g.,  $\overline{\textcolor{red}{X}}$ ).
- An estimator can be biased but consistent (e.g.,  $\widetilde{X}$ ).
- An estimator can be biased and inconsistent

# Consistency

## The takeaway?

- An estimator can be unbiased without being consistent (e.g.,  $\textcolor{blue}{X}_1$ ).
- An estimator can be unbiased and consistent (e.g.,  $\textcolor{red}{\bar{X}}$ ).
- An estimator can be biased but consistent (e.g.,  $\widetilde{\bar{X}}$ ).
- An estimator can be biased and inconsistent (e.g.,  $\overline{X} - 50$ ).

# Consistency

## The takeaway?

- An estimator can be unbiased without being consistent (e.g.,  $\textcolor{blue}{X}_1$ ).
- An estimator can be unbiased and consistent (e.g.,  $\textcolor{red}{\bar{X}}$ ).
- An estimator can be biased but consistent (e.g.,  $\widetilde{\bar{X}}$ ).
- An estimator can be biased and inconsistent (e.g.,  $\bar{X} - 50$ ).

**Best-case scenario:** The estimator is unbiased and consistent.

# Consistency

## Why consistency (asymptotics)?

1. We cannot always find an unbiased estimator. In these situations, we generally (at least) want consistency.
2. Expected values can be hard/undefined. Probability limits are less constrained, *e.g.*,

$$E[g(X)h(Y)] \text{ vs. } \text{plim}(g(X)h(Y))$$

3. Asymptotics help us move away from assuming the distribution of  $u_i$ .

# Consistency

## Why consistency (asymptotics)?

1. We cannot always find an unbiased estimator. In these situations, we generally (at least) want consistency.
2. Expected values can be hard/undefined. Probability limits are less constrained, *e.g.*,

$$E[g(X)h(Y)] \text{ vs. } \text{plim}(g(X)h(Y))$$

3. Asymptotics help us move away from assuming the distribution of  $u_i$ .

**Caution:** As we saw, consistent estimators can be biased in small samples.

# OLS in asymptopia

# OLS in asymptopia

OLS has two very nice asymptotic properties:

1. Consistency
2. Asymptotic Normality

# OLS in asymptopia

OLS has two very nice asymptotic properties:

1. Consistency
2. Asymptotic Normality

Let's prove #1 for OLS with simple, linear regression, *i.e.*,

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

# OLS in asymptopia

## Proof of consistency

First, recall our previous derivation of  $\hat{\beta}_1$ ,

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_i (x_i - \bar{x}) u_i}{\sum_i (x_i - \bar{x})^2}$$

# OLS in asymptopia

## Proof of consistency

First, recall our previous derivation of  $\hat{\beta}_1$ ,

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_i (x_i - \bar{x}) u_i}{\sum_i (x_i - \bar{x})^2}$$

Now divide the numerator and denominator by  $1/n$

# OLS in asymptopia

## Proof of consistency

First, recall our previous derivation of  $\hat{\beta}_1$ ,

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_i (x_i - \bar{x}) u_i}{\sum_i (x_i - \bar{x})^2}$$

Now divide the numerator and denominator by  $1/n$

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n} \sum_i (x_i - \bar{x}) u_i}{\frac{1}{n} \sum_i (x_i - \bar{x})^2}$$

# OLS in asymptopia

## Proof of consistency

We actually want to know the probability limit of  $\hat{\beta}_1$ , so

# OLS in asymptopia

## Proof of consistency

We actually want to know the probability limit of  $\hat{\beta}_1$ , so

$$\text{plim } \hat{\beta}_1 = \text{plim} \left( \beta_1 + \frac{\frac{1}{n} \sum_i (x_i - \bar{x}) u_i}{\frac{1}{n} \sum_i (x_i - \bar{x})^2} \right)$$

# OLS in asymptopia

## Proof of consistency

We actually want to know the probability limit of  $\hat{\beta}_1$ , so

$$\text{plim } \hat{\beta}_1 = \text{plim} \left( \beta_1 + \frac{\frac{1}{n} \sum_i (x_i - \bar{x}) u_i}{\frac{1}{n} \sum_i (x_i - \bar{x})^2} \right)$$

which, by the properties of probability limits, gives us

# OLS in asymptopia

## Proof of consistency

We actually want to know the probability limit of  $\hat{\beta}_1$ , so

$$\text{plim } \hat{\beta}_1 = \text{plim} \left( \beta_1 + \frac{\frac{1}{n} \sum_i (x_i - \bar{x}) u_i}{\frac{1}{n} \sum_i (x_i - \bar{x})^2} \right)$$

which, by the properties of probability limits, gives us

$$= \beta_1 + \frac{\text{plim} \left( \frac{1}{n} \sum_i (x_i - \bar{x}) u_i \right)}{\text{plim} \left( \frac{1}{n} \sum_i (x_i - \bar{x})^2 \right)}$$

# OLS in asymptopia

## Proof of consistency

We actually want to know the probability limit of  $\hat{\beta}_1$ , so

$$\text{plim } \hat{\beta}_1 = \text{plim} \left( \beta_1 + \frac{\frac{1}{n} \sum_i (x_i - \bar{x}) u_i}{\frac{1}{n} \sum_i (x_i - \bar{x})^2} \right)$$

which, by the properties of probability limits, gives us

$$= \beta_1 + \frac{\text{plim} \left( \frac{1}{n} \sum_i (x_i - \bar{x}) u_i \right)}{\text{plim} \left( \frac{1}{n} \sum_i (x_i - \bar{x})^2 \right)}$$

The numerator and denominator are, in fact, population quantities

# OLS in asymptopia

## Proof of consistency

We actually want to know the probability limit of  $\hat{\beta}_1$ , so

$$\text{plim } \hat{\beta}_1 = \text{plim} \left( \beta_1 + \frac{\frac{1}{n} \sum_i (x_i - \bar{x}) u_i}{\frac{1}{n} \sum_i (x_i - \bar{x})^2} \right)$$

which, by the properties of probability limits, gives us

$$= \beta_1 + \frac{\text{plim} \left( \frac{1}{n} \sum_i (x_i - \bar{x}) u_i \right)}{\text{plim} \left( \frac{1}{n} \sum_i (x_i - \bar{x})^2 \right)}$$

The numerator and denominator are, in fact, population quantities

$$= \beta_1 + \frac{\text{Cov}(x, u)}{\text{Var}(x)}$$

# OLS in asymptopia

## Proof of consistency

So we have

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x, u)}{\text{Var}(x)}$$

# OLS in asymptopia

## Proof of consistency

So we have

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x, u)}{\text{Var}(x)}$$

By our assumption of exogeneity (plus the law of total expectation)

# OLS in asymptopia

## Proof of consistency

So we have

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x, u)}{\text{Var}(x)}$$

By our assumption of exogeneity (plus the law of total expectation)

$$\text{Cov}(x, u) = 0$$

# OLS in asymptopia

## Proof of consistency

So we have

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x, u)}{\text{Var}(x)}$$

By our assumption of exogeneity (plus the law of total expectation)

$$\text{Cov}(x, u) = 0$$

Combining these two equations yields

# OLS in asymptopia

## Proof of consistency

So we have

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x, u)}{\text{Var}(x)}$$

By our assumption of exogeneity (plus the law of total expectation)

$$\text{Cov}(x, u) = 0$$

Combining these two equations yields

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{0}{\text{Var}(x)} = \beta_1 \quad \text{😎}$$

so long as  $\text{Var}(x) \neq 0$  (which we've assumed).

# OLS in asymptopia

## Asymptotic normality

Up to this point, we made a very specific assumption about the distribution of  $u_i$ —the  $u_i$  came from a normal distribution.

We can relax this assumption—allowing the  $u_i$  to come from any distribution (still assume exogeneity, independence, and homoskedasticity).

We will focus on the **asymptotic distribution** of our estimators (how they are distributed as  $n$  gets large), rather than their finite-sample distribution.

# OLS in asymptopia

## Asymptotic normality

Up to this point, we made a very specific assumption about the distribution of  $u_i$ —the  $u_i$  came from a normal distribution.

We can relax this assumption—allowing the  $u_i$  to come from any distribution (still assume exogeneity, independence, and homoskedasticity).

We will focus on the **asymptotic distribution** of our estimators (how they are distributed as  $n$  gets large), rather than their finite-sample distribution.

As  $n$  approaches  $\infty$ , the distribution of the OLS estimator converges to a normal distribution.

# OLS in asymptopia

## Recap

With a more limited set of assumptions, OLS is **consistent** and is **asymptotically normally distributed**.

## Current assumptions

1. Our data were **randomly sampled** from the population.
2.  $y_i$  is a **linear function** of its parameters and disturbance.
3. There is **no perfect collinearity** in our data.
4. The  $u_i$  have conditional mean of zero (**exogeneity**),  $E[u_i|X_i] = 0$ .
5. The  $u_i$  are **homoskedastic** with **zero correlation** between  $u_i$  and  $u_j$ .

# Omitted-variable bias, redux

# Omitted-variable bias, redux

## Inconsistency?

Imagine we have a population whose true model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad (2)$$

# Omitted-variable bias, redux

## Inconsistency?

Imagine we have a population whose true model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad (2)$$

*Recall<sub>1</sub>:* **Omitted-variable bias** occurs when we omit a variable in our linear regression model (e.g., leaving out  $x_2$ ) such that

# Omitted-variable bias, redux

## Inconsistency?

Imagine we have a population whose true model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad (2)$$

Recall<sub>1</sub>: **Omitted-variable bias** occurs when we omit a variable in our linear regression model (e.g., leaving out  $x_2$ ) such that

1.  $x_2$  affects  $y$ , i.e.,  $\beta_2 \neq 0$ .

# Omitted-variable bias, redux

## Inconsistency?

Imagine we have a population whose true model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad (2)$$

Recall<sub>1</sub>: **Omitted-variable bias** occurs when we omit a variable in our linear regression model (e.g., leaving out  $x_2$ ) such that

1.  $x_2$  affects  $y$ , i.e.,  $\beta_2 \neq 0$ .
2. Correlates with an included explanatory variable, i.e.,  $\text{Cov}(x_1, x_2) \neq 0$ .

# Omitted-variable bias, redux

## Inconsistency?

Imagine we have a population whose true model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad (2)$$

*Recall*<sub>2</sub>: We defined the **bias** of an estimator  $\hat{W}$  for parameter  $\theta$

# Omitted-variable bias, redux

## Inconsistency?

Imagine we have a population whose true model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad (2)$$

*Recall*<sub>2</sub>: We defined the **bias** of an estimator  $W$  for parameter  $\theta$

$$\text{Bias}_{\theta}(W) = E[W] - \theta$$

# Omitted-variable bias, redux

## Inconsistency?

Imagine we have a population whose true model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad (2)$$

We know that omitted-variable bias causes *biased estimates*.

*Question:* Do *omitted variables* also cause *inconsistent estimates*?

# Omitted-variable bias, redux

## Inconsistency?

Imagine we have a population whose true model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad (2)$$

We know that omitted-variable bias causes **biased estimates**.

*Question:* Do *omitted variables* also cause **inconsistent estimates**?

*Answer:* Find  $\text{plim } \hat{\beta}_1$  in a regression that omits  $x_2$ .

# Omitted-variable bias, redux

## Inconsistency?

Imagine we have a population whose true model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad (2)$$

but we instead specify the model as

$$y_i = \beta_0 + \beta_1 x_{1i} + w_i \quad (3)$$

where  $w_i = \beta_2 x_{2i} + u_i$ .

# Omitted-variable bias, redux

## Inconsistency?

Imagine we have a population whose true model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad (2)$$

but we instead specify the model as

$$y_i = \beta_0 + \beta_1 x_{1i} + w_i \quad (3)$$

where  $w_i = \beta_2 x_{2i} + u_i$ . We estimate (3) via OLS

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{w}_i \quad (4)$$

# Omitted-variable bias, redux

## Inconsistency?

Imagine we have a population whose true model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad (2)$$

but we instead specify the model as

$$y_i = \beta_0 + \beta_1 x_{1i} + w_i \quad (3)$$

where  $w_i = \beta_2 x_{2i} + u_i$ . We estimate (3) via OLS

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{w}_i \quad (4)$$

*Our question:* Is  $\hat{\beta}_1$  consistent for  $\beta_1$  when we omit  $x_2$ ?

# Omitted-variable bias, redux

## Inconsistency?

Imagine we have a population whose true model is

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad (2)$$

but we instead specify the model as

$$y_i = \beta_0 + \beta_1 x_{1i} + w_i \quad (3)$$

where  $w_i = \beta_2 x_{2i} + u_i$ . We estimate (3) via OLS

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{w}_i \quad (4)$$

*Our question:* Is  $\hat{\beta}_1$  consistent for  $\beta_1$  when we omit  $x_2$ ?

$$\text{plim}(\hat{\beta}_1) \stackrel{?}{=} \beta_1$$

# Omitted-variable bias, redux

## Inconsistency?

**Truth:**  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$

**Specified:**  $y_i = \beta_0 + \beta_1 x_{1i} + w_i$

We already showed  $\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, w)}{\text{Var}(x_1)}$

where  $w$  is the disturbance.

# Omitted-variable bias, redux

## Inconsistency?

**Truth:**  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$

**Specified:**  $y_i = \beta_0 + \beta_1 x_{1i} + w_i$

We already showed  $\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, w)}{\text{Var}(x_1)}$

where  $w$  is the disturbance. Here, we know  $w = \beta_2 x_2 + u$ . Thus,

# Omitted-variable bias, redux

## Inconsistency?

**Truth:**  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$       **Specified:**  $y_i = \beta_0 + \beta_1 x_{1i} + w_i$

We already showed  $\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, w)}{\text{Var}(x_1)}$

where  $w$  is the disturbance. Here, we know  $w = \beta_2 x_2 + u$ . Thus,

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, \beta_2 x_2 + u)}{\text{Var}(x_1)}$$

# Omitted-variable bias, redux

## Inconsistency?

**Truth:**  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$       **Specified:**  $y_i = \beta_0 + \beta_1 x_{1i} + w_i$

We already showed  $\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, w)}{\text{Var}(x_1)}$

where  $w$  is the disturbance. Here, we know  $w = \beta_2 x_2 + u$ . Thus,

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, \beta_2 x_2 + u)}{\text{Var}(x_1)}$$

Now, we make use of  $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$

# Omitted-variable bias, redux

## Inconsistency?

**Truth:**  $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$       **Specified:**  $y_i = \beta_0 + \beta_1 x_{1i} + w_i$

We already showed  $\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, w)}{\text{Var}(x_1)}$

where  $w$  is the disturbance. Here, we know  $w = \beta_2 x_2 + u$ . Thus,

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, \beta_2 x_2 + u)}{\text{Var}(x_1)}$$

Now, we make use of  $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, \beta_2 x_2) + \text{Cov}(x_1, u)}{\text{Var}(x_1)}$$

# Omitted-variable bias, redux

## Inconsistency?

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, \beta_2 x_2) + \text{Cov}(x_1, u)}{\text{Var}(x_1)}$$

Now we use the fact that  $\text{Cov}(X, cY) = c \text{Cov}(X, Y)$  for a constant  $c$ .

# Omitted-variable bias, redux

## Inconsistency?

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, \beta_2 x_2) + \text{Cov}(x_1, u)}{\text{Var}(x_1)}$$

Now we use the fact that  $\text{Cov}(X, cY) = c \text{Cov}(X, Y)$  for a constant  $c$ .

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\beta_2 \text{Cov}(x_1, x_2) + \text{Cov}(x_1, u)}{\text{Var}(x_1)}$$

# Omitted-variable bias, redux

## Inconsistency?

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, \beta_2 x_2) + \text{Cov}(x_1, u)}{\text{Var}(x_1)}$$

Now we use the fact that  $\text{Cov}(X, cY) = c \text{Cov}(X, Y)$  for a constant  $c$ .

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\beta_2 \text{Cov}(x_1, x_2) + \text{Cov}(x_1, u)}{\text{Var}(x_1)}$$

As before, our exogeneity (conditional mean zero) assumption implies  $\text{Cov}(x_1, u) = 0$ , which gives us

# Omitted-variable bias, redux

## Inconsistency?

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, \beta_2 x_2) + \text{Cov}(x_1, u)}{\text{Var}(x_1)}$$

Now we use the fact that  $\text{Cov}(X, cY) = c \text{Cov}(X, Y)$  for a constant  $c$ .

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\beta_2 \text{Cov}(x_1, x_2) + \text{Cov}(x_1, u)}{\text{Var}(x_1)}$$

As before, our exogeneity (conditional mean zero) assumption implies  $\text{Cov}(x_1, u) = 0$ , which gives us

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\beta_2 \text{Cov}(x_1, x_2)}{\text{Var}(x_1)}$$

# Omitted-variable bias, redux

## Inconsistency?

Thus, we find that

$$\text{plim } \hat{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)}$$

In other words, an omitted variable will cause OLS to be inconsistent if **both** of the following statements are true:

1. The omitted variable **affects our outcome**, i.e.,  $\beta_2 \neq 0$ .
2. The omitted variable correlates with included explanatory variables, i.e.,  $\text{Cov}(x_1, x_2) \neq 0$ .

If both of these statements are true, then the OLS estimate  $\hat{\beta}_1$  will not converge to  $\beta_1$ , even as  $n$  approaches  $\infty$ .

# Omitted-variable bias, redux

## Signing the bias

Sometimes we're stuck with omitted variable bias.<sup>†</sup>

$$\text{plim } \hat{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)}$$

When this happens, we can often at least know the direction of the inconsistency.

<sup>†</sup> You will often hear the term "omitted-variable bias" when we're actually talking about inconsistency (rather than bias).

# Omitted-variable bias, redux

## Signing the bias

Begin with

$$\text{plim } \hat{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)}$$

We know  $\text{Var}(x_1) > 0$ . Suppose  $\beta_2 > 0$  and  $\text{Cov}(x_1, x_2) > 0$ . Then

# Omitted-variable bias, redux

## Signing the bias

Begin with

$$\text{plim } \hat{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)}$$

We know  $\text{Var}(x_1) > 0$ . Suppose  $\beta_2 > 0$  and  $\text{Cov}(x_1, x_2) > 0$ . Then

$$\text{plim } \hat{\beta}_1 = \beta_1 + (+) \frac{(+)}{(+) \quad (+)} \implies \text{plim } \hat{\beta}_1 > \beta_1$$

∴ In this case, OLS is **biased upward** (estimates are too large).

# Omitted-variable bias, redux

## Signing the bias

Begin with

$$\text{plim } \hat{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)}$$

We know  $\text{Var}(x_1) > 0$ . Suppose  $\beta_2 > 0$  and  $\text{Cov}(x_1, x_2) > 0$ . Then

$$\text{plim } \hat{\beta}_1 = \beta_1 + (+) \frac{(+)}{(+)} \implies \text{plim } \hat{\beta}_1 > \beta_1$$

∴ In this case, OLS is **biased upward** (estimates are too large).

$$\text{Cov}(x_1, x_2) > 0 \quad \text{Cov}(x_1, x_2) < 0$$

$$\begin{array}{ll} \beta_2 > 0 & \text{Upward} \\ \beta_2 < 0 & \end{array}$$

# Omitted-variable bias, redux

## Signing the bias

Begin with

$$\text{plim } \hat{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)}$$

We know  $\text{Var}(x_1) > 0$ . Suppose  $\beta_2 < 0$  and  $\text{Cov}(x_1, x_2) > 0$ . Then

# Omitted-variable bias, redux

## Signing the bias

Begin with

$$\text{plim } \hat{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)}$$

We know  $\text{Var}(x_1) > 0$ . Suppose  $\beta_2 < 0$  and  $\text{Cov}(x_1, x_2) > 0$ . Then

$$\text{plim } \hat{\beta}_1 = \beta_1 + (-) \frac{(+)}{(+) \quad (+)} \implies \text{plim } \hat{\beta}_1 < \beta_1$$

∴ In this case, OLS is **biased downward** (estimates are too small).

$$\text{Cov}(x_1, x_2) > 0 \quad \text{Cov}(x_1, x_2) < 0$$

$\beta_2 > 0$       Upward

$\beta_2 < 0$       Downward

# Omitted-variable bias, redux

## Signing the bias

Begin with

$$\text{plim } \hat{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)}$$

We know  $\text{Var}(x_1) > 0$ . Suppose  $\beta_2 > 0$  and  $\text{Cov}(x_1, x_2) < 0$ . Then

$$\text{plim } \hat{\beta}_1 = \beta_1 + (+) \frac{(-)}{(+) \phantom{+}} \implies \text{plim } \hat{\beta}_1 < \beta_1$$

∴ In this case, OLS is **biased downward** (estimates are too small).

	$\text{Cov}(x_1, x_2) > 0$	$\text{Cov}(x_1, x_2) < 0$
$\beta_2 > 0$	Upward	Downward
$\beta_2 < 0$	Downward	

# Omitted-variable bias, redux

## Signing the bias

Begin with

$$\text{plim } \hat{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(x_1, x_2)}{\text{Var}(x_1)}$$

We know  $\text{Var}(x_1) > 0$ . Suppose  $\beta_2 < 0$  and  $\text{Cov}(x_1, x_2) < 0$ . Then

$$\text{plim } \hat{\beta}_1 = \beta_1 + (-) \frac{(-)}{(+)} \implies \text{plim } \hat{\beta}_1 > \beta_1$$

∴ In this case, OLS is **biased upward** (estimates are too large).

	$\text{Cov}(x_1, x_2) > 0$	$\text{Cov}(x_1, x_2) < 0$
$\beta_2 > 0$	Upward	Downward
$\beta_2 < 0$	Downward	Upward

# Omitted-variable bias, redux

## Signing the bias

Thus, in cases where we have a sense of

1. the sign of  $\text{Cov}(x_1, x_2)$
2. the sign of  $\beta_2$

we know in which direction inconsistency pushes our estimates.

### Direction of bias

$$\text{Cov}(x_1, x_2) > 0 \quad \text{Cov}(x_1, x_2) < 0$$

$\beta_2 > 0$	Upward	Downward
$\beta_2 < 0$	Downward	Upward

# Measurement error

**Measurement error** in our explanatory variables presents another case in which OLS is inconsistent.

Consider the population model:  $y_i = \beta_0 + \beta_1 z_i + u_i$

- We want to observe  $z_i$  but cannot.
- Instead, we *measure* the variable  $x_i$ , which is  $z_i$  plus some error (noise):

$$x_i = z_i + \omega_i$$

- Assume  $E[\omega_i] = 0$ ,  $\text{Var}(\omega_i) = \sigma_\omega^2$ , and  $\omega$  is independent of  $z$  and  $u$ .

# Measurement error

**Measurement error** in our explanatory variables presents another case in which OLS is inconsistent.

Consider the population model:  $y_i = \beta_0 + \beta_1 z_i + u_i$

- We want to observe  $z_i$  but cannot.
- Instead, we *measure* the variable  $x_i$ , which is  $z_i$  plus some error (noise):

$$x_i = z_i + \omega_i$$

- Assume  $E[\omega_i] = 0$ ,  $\text{Var}(\omega_i) = \sigma_\omega^2$ , and  $\omega$  is independent of  $z$  and  $u$ .

OLS regression of  $y$  and  $x$  will produce inconsistent estimates for  $\beta_1$ .

# Measurement error

## Proof

$$y_i = \beta_0 + \beta_1 z_i + u_i$$

# Measurement error

## Proof

$$\begin{aligned}y_i &= \beta_0 + \beta_1 z_i + u_i \\&= \beta_0 + \beta_1 (x_i - \omega_i) + u_i\end{aligned}$$

# Measurement error

## Proof

$$\begin{aligned}y_i &= \beta_0 + \beta_1 z_i + u_i \\&= \beta_0 + \beta_1 (x_i - \omega_i) + u_i \\&= \beta_0 + \beta_1 x_i + (u_i - \beta_1 \omega_i)\end{aligned}$$

# Measurement error

## Proof

$$\begin{aligned}y_i &= \beta_0 + \beta_1 z_i + u_i \\&= \beta_0 + \beta_1 (x_i - \omega_i) + u_i \\&= \beta_0 + \beta_1 x_i + (u_i - \beta_1 \omega_i) \\&= \beta_0 + \beta_1 x_i + \varepsilon_i\end{aligned}$$

where  $\varepsilon_i = u_i - \beta_1 \omega_i$

# Measurement error

## Proof

$$\begin{aligned}y_i &= \beta_0 + \beta_1 z_i + u_i \\&= \beta_0 + \beta_1 (x_i - \omega_i) + u_i \\&= \beta_0 + \beta_1 x_i + (u_i - \beta_1 \omega_i) \\&= \beta_0 + \beta_1 x_i + \varepsilon_i\end{aligned}$$

where  $\varepsilon_i = u_i - \beta_1 \omega_i$

What happens when we estimate  $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i$ ?

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x, \varepsilon)}{\text{Var}(x)}$$

We will derive the numerator and denominator separately...

# Measurement error

## Proof

The covariance of our noisy variable  $x$  and the disturbance  $\varepsilon$ .

$$\text{Cov}(x, \varepsilon)$$

# Measurement error

## Proof

The covariance of our noisy variable  $x$  and the disturbance  $\varepsilon$ .

$$\text{Cov}(x, \varepsilon) = \text{Cov}([z + \omega], [u - \beta_1 \omega])$$

# Measurement error

## Proof

The covariance of our noisy variable  $x$  and the disturbance  $\varepsilon$ .

$$\begin{aligned}\text{Cov}(x, \varepsilon) &= \text{Cov}([z + \omega], [u - \beta_1 \omega]) \\ &= \text{Cov}(z, u) - \beta_1 \text{Cov}(z, \omega) + \text{Cov}(\omega, u) - \beta_1 \text{Var}(\omega)\end{aligned}$$

# Measurement error

## Proof

The covariance of our noisy variable  $x$  and the disturbance  $\varepsilon$ .

$$\begin{aligned}\text{Cov}(x, \varepsilon) &= \text{Cov}([z + \omega], [u - \beta_1 \omega]) \\ &= \text{Cov}(z, u) - \beta_1 \text{Cov}(z, \omega) + \text{Cov}(\omega, u) - \beta_1 \text{Var}(\omega) \\ &= 0 + 0 + 0 - \beta_1 \sigma_\omega^2\end{aligned}$$

# Measurement error

## Proof

The covariance of our noisy variable  $x$  and the disturbance  $\varepsilon$ .

$$\begin{aligned}\text{Cov}(x, \varepsilon) &= \text{Cov}([z + \omega], [u - \beta_1 \omega]) \\ &= \text{Cov}(z, u) - \beta_1 \text{Cov}(z, \omega) + \text{Cov}(\omega, u) - \beta_1 \text{Var}(\omega) \\ &= 0 + 0 + 0 - \beta_1 \sigma_\omega^2 \\ &= -\beta_1 \sigma_\omega^2\end{aligned}$$

# Measurement error

## Proof

Now for the denominator,  $\text{Var}(x)$ .

$\text{Var}(x)$

# Measurement error

## Proof

Now for the denominator,  $\text{Var}(x)$ .

$$\text{Var}(x) = \text{Var}(z + \omega)$$

# Measurement error

## Proof

Now for the denominator,  $\text{Var}(x)$ .

$$\begin{aligned}\text{Var}(x) &= \text{Var}(z + \omega) \\ &= \text{Var}(z) + \text{Var}(\omega) + 2 \text{Cov}(z, \omega)\end{aligned}$$

# Measurement error

## Proof

Now for the denominator,  $\text{Var}(x)$ .

$$\begin{aligned}\text{Var}(x) &= \text{Var}(z + \omega) \\ &= \text{Var}(z) + \text{Var}(\omega) + 2 \text{Cov}(z, \omega) \\ &= \sigma_z^2 + \sigma_\omega^2\end{aligned}$$

# Measurement error

## Proof

Putting the numerator and denominator back together,

$$\begin{aligned}\text{plim } \hat{\beta}_1 &= \beta_1 + \frac{\text{Cov}(x, \varepsilon)}{\text{Var}(x)} \\&= \beta_1 + \frac{-\beta_1 \sigma_\omega^2}{\sigma_z^2 + \sigma_\omega^2} \\&= \beta_1 - \beta_1 \frac{\sigma_\omega^2}{\sigma_z^2 + \sigma_\omega^2} \\&= \beta_1 \frac{\sigma_z^2 + \sigma_\omega^2}{\sigma_z^2 + \sigma_\omega^2} - \beta_1 \frac{\sigma_\omega^2}{\sigma_z^2 + \sigma_\omega^2} \\&= \beta_1 \frac{\sigma_z^2}{\sigma_z^2 + \sigma_\omega^2}\end{aligned}$$

# Measurement error

## Summary

$$\therefore \text{plim } \hat{\beta}_1 = \beta_1 \frac{\sigma_z^2}{\sigma_z^2 + \sigma_\omega^2}.$$

What does this equation tell us?

# Measurement error

## Summary

$$\therefore \text{plim } \hat{\beta}_1 = \beta_1 \frac{\sigma_z^2}{\sigma_z^2 + \sigma_\omega^2}.$$

What does this equation tell us?

**Measurement error in our explanatory variables** biases the coefficient estimates toward zero.

- This type of bias/inconsistency is often called **attenuation bias**.
- If **the measurement error correlates with the explanatory variables**, we have bigger problems with inconsistency/bias.

# Measurement error

## Summary

What about **measurement in the outcome variable**?

It doesn't really matter—it just increases our standard errors.

# Measurement error

It's everywhere

## General cases

1. We cannot perfectly observe a variable.
2. We use one variable as a *proxy* for another.

## Specific examples

- GDP
- Population
- Crime/police statistics
- Air quality
- Health data
- Proxy *ability* with test scores