Asymptotics and consistency

EC 421, Set 6

Edward Rubin

04 February 2020

Prologue

Schedule

Last Time

Living with heteroskedasticity

Today

Asymptotics and consistency

This week

Our second assignment

Near-ish future

Midterm next week (Thursday?)

`R` showcase

Need speed? `R` allows essentially infinite parallelization.

Three popular packages:

- `future` and `furrr`
- `parallel`
- `foreach`

And here's a nice tutorial.

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: $\boldsymbol{E}\left[\hat{\beta}_j\right] = ?$

2. The **variance** of the estimator's distribution: $\mathrm{Var}\left(\hat{\beta}_j\right) = ?$

which tell us about the **tendency of the estimator** if we took **∞ 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 $n \to \infty$)?

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 ∞).

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"[†]).

† Here, "big" n means $n \to \infty$. 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 α if

$$\lim_{n \to \infty} P\left(\left| B_n - \alpha \right| > \epsilon \right) = 0$$

for any $\varepsilon > 0$.

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

Practically: B's distribution collapses to a spike at α as n approaches ∞.

Probability limits

Equivalent statements:

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

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
- $\text{plim}\left(\dfrac{X}{Y}\right) = \dfrac{\text{plim}(X)}{\text{plim}(Y)}$
- $\text{plim}\big(f(X)\big) = f\big(\text{plim}(X)\big)$

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 α if $\text{plim} B = \alpha$.

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

Consistent estimators

Example: We want to estimate the population mean μ_x (where X~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: $\tilde{X} = \frac{1}{n+1}\sum_{i=1}^{n} x_i$

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

Consistent estimators

To see which are unbiased/biased:

$$\boldsymbol{E}\left[X_1\right] = \mu_x$$

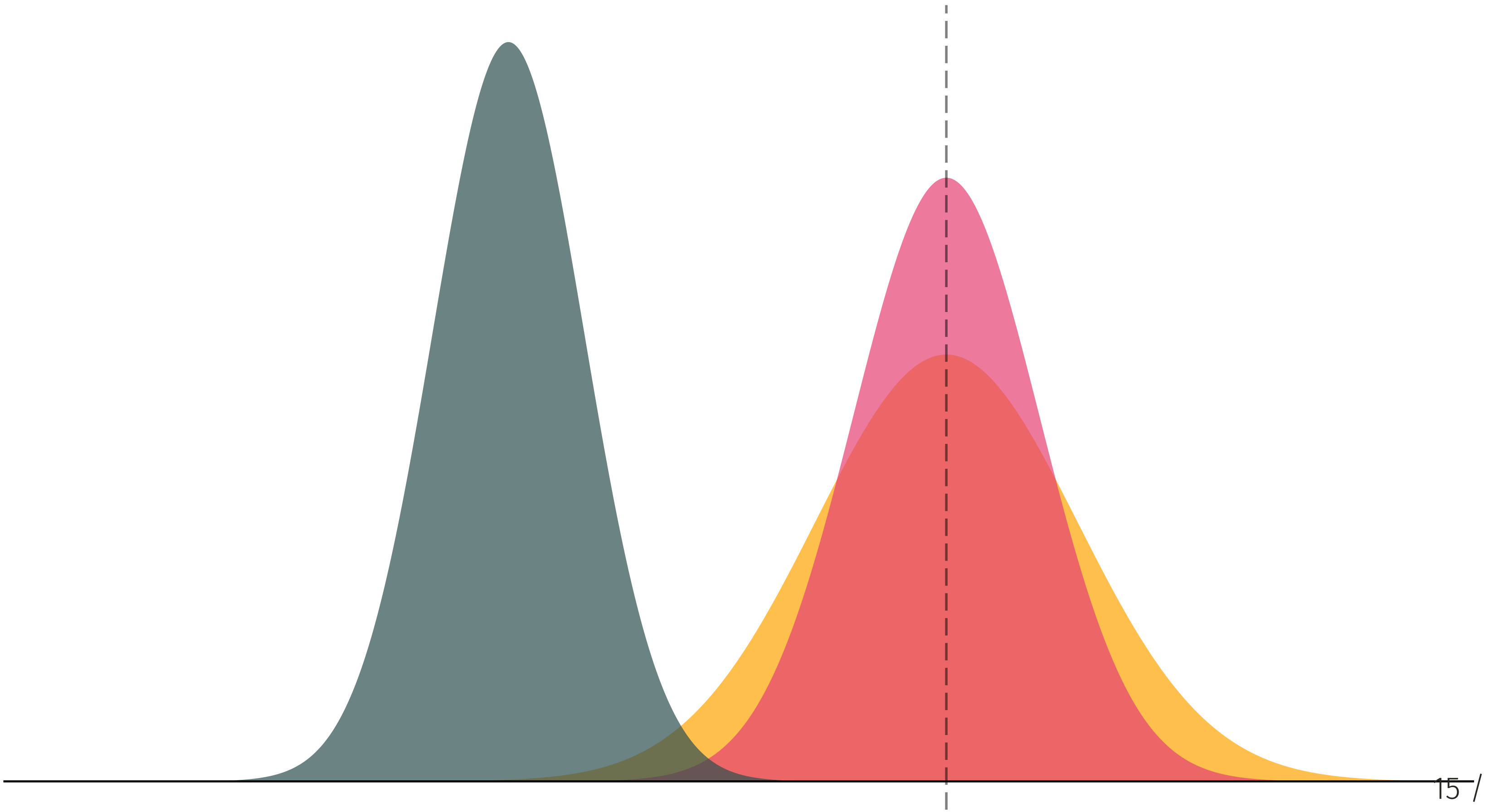
$$\boldsymbol{E}\left[\bar{X}\right] = \boldsymbol{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{E}\left[x_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mu_x = \mu_x$$

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

Consistency

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

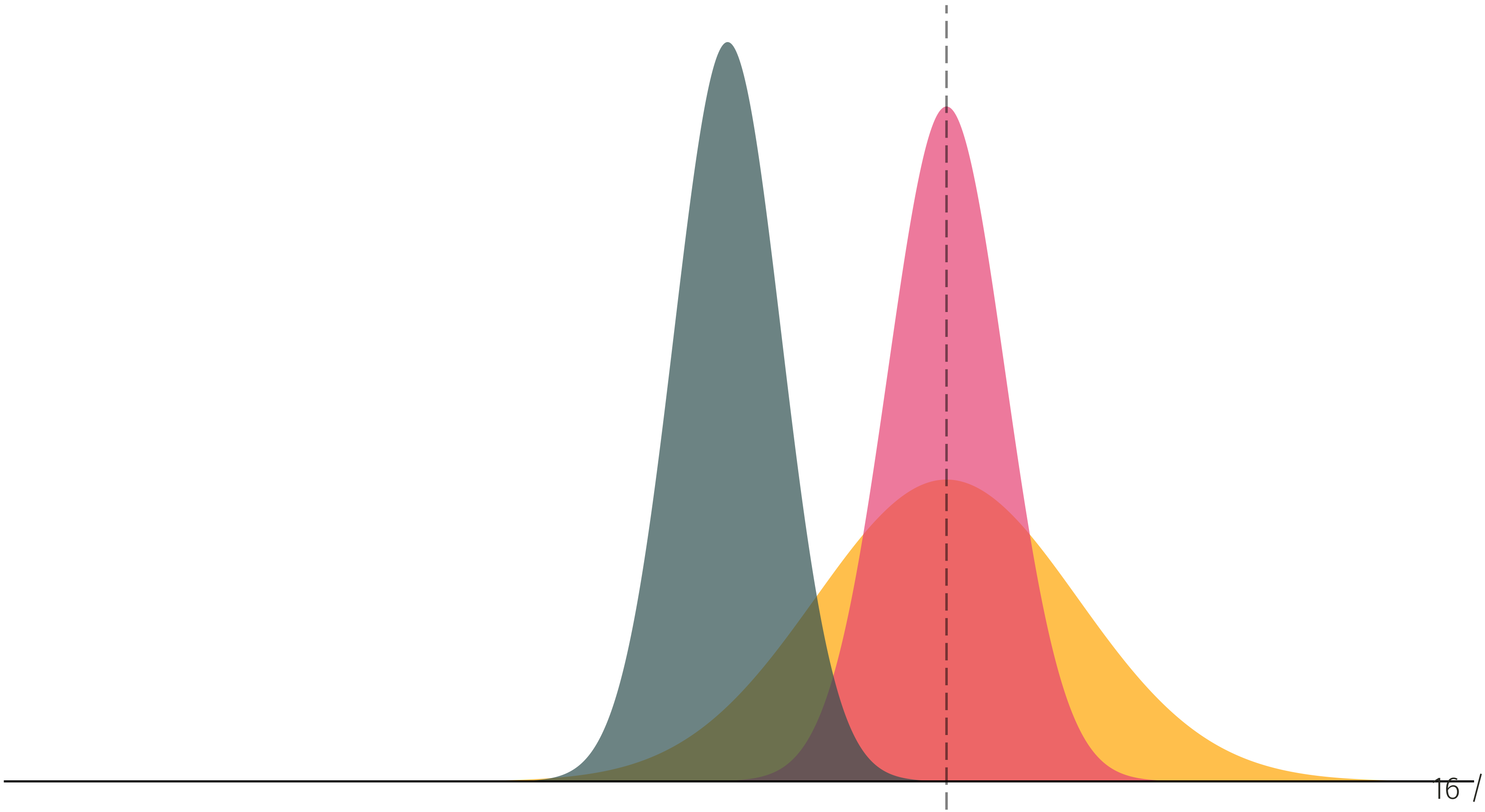
$n = 2$



Consistency

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

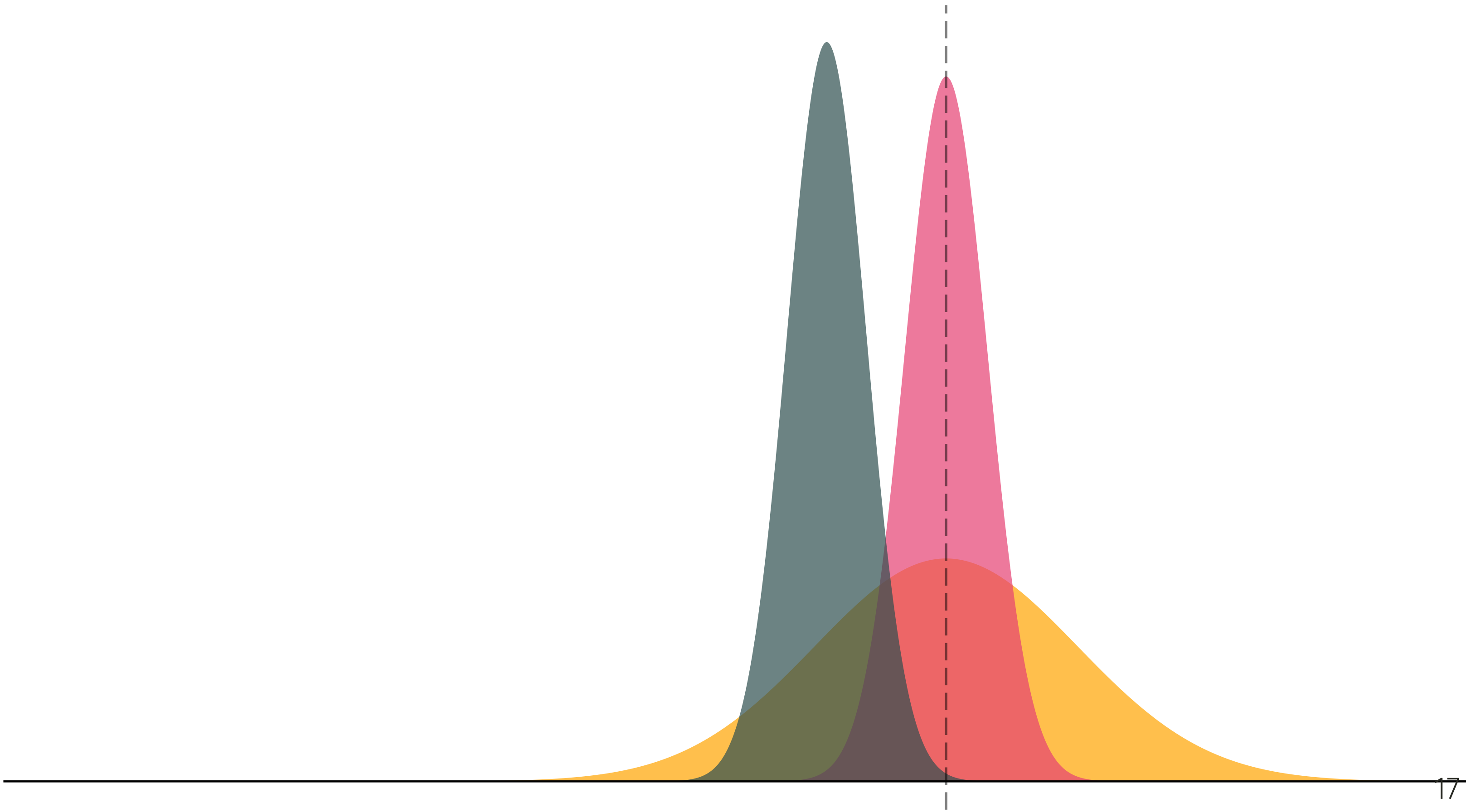
$n = 5$



Consistency

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

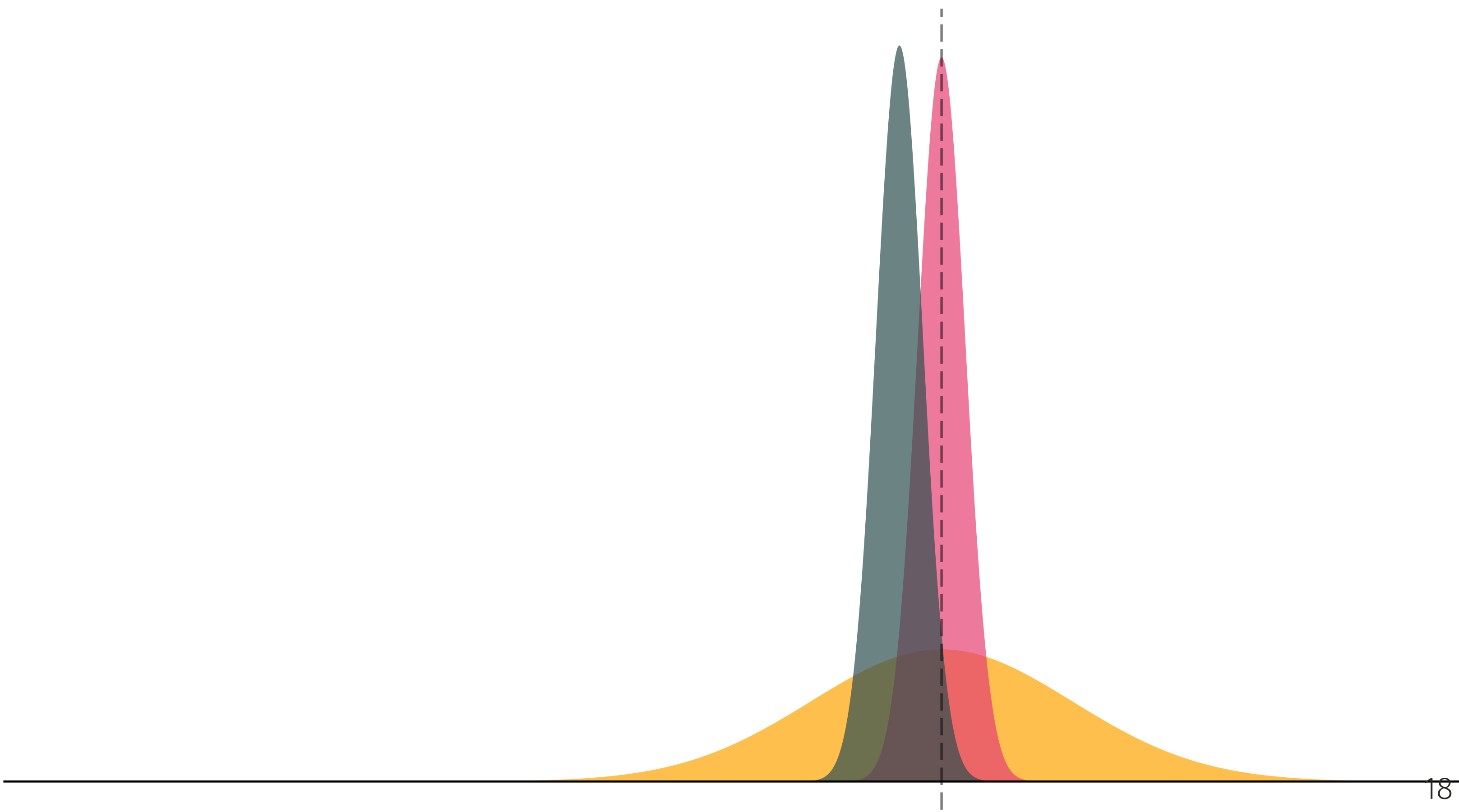
$n = 10$



Consistency

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

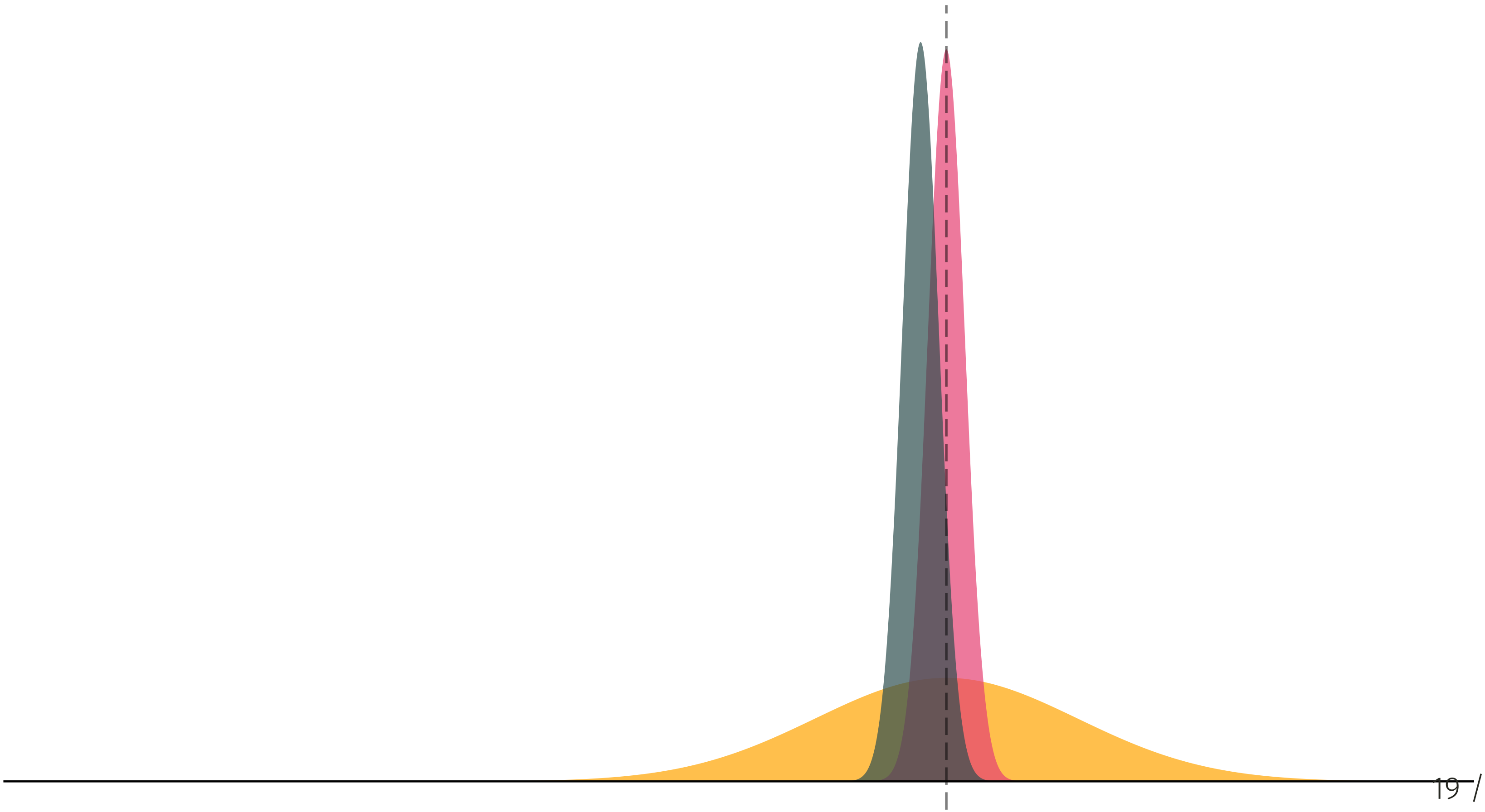
$n = 30$



Consistency

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

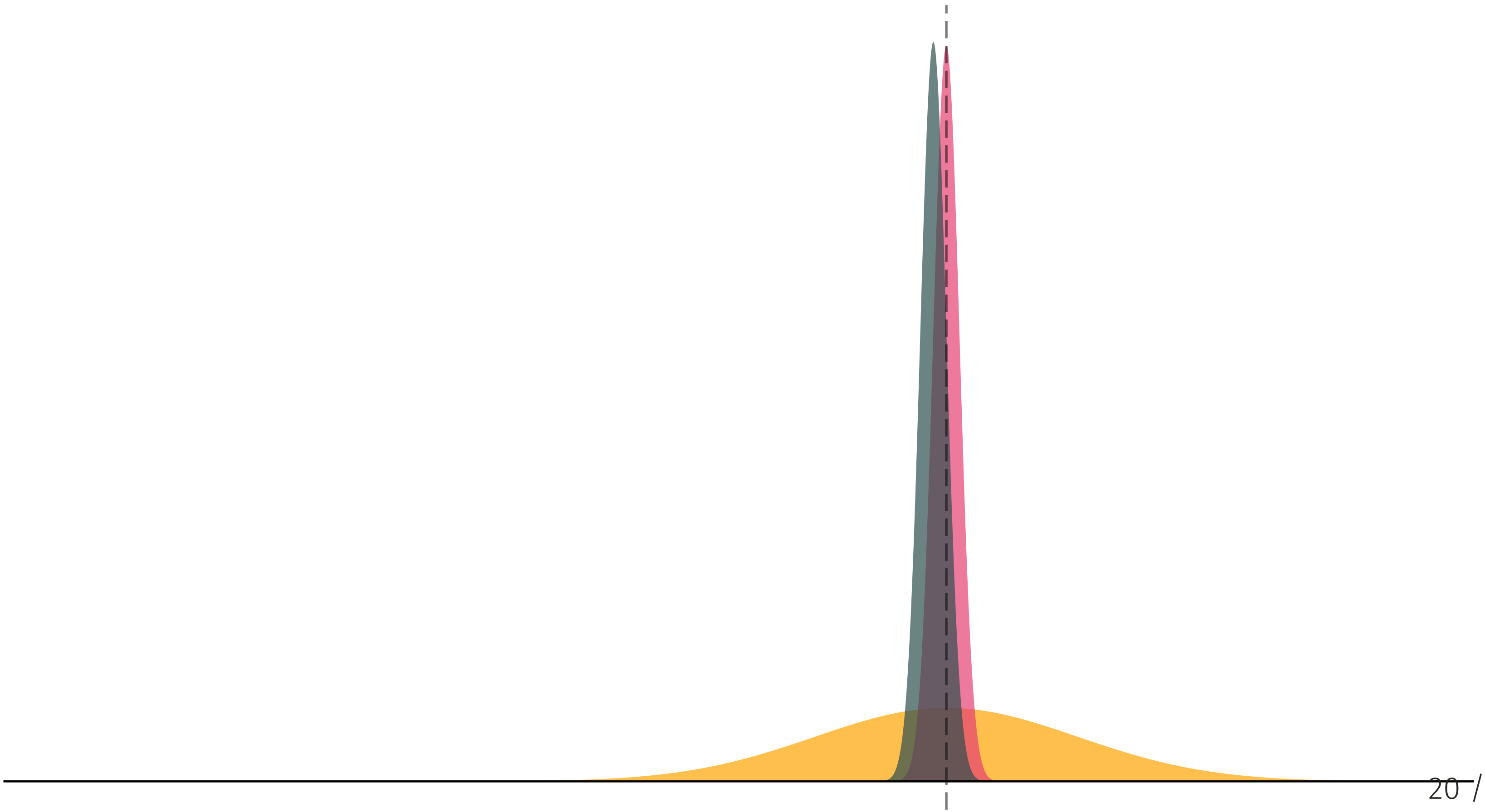
$n = 50$



Consistency

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

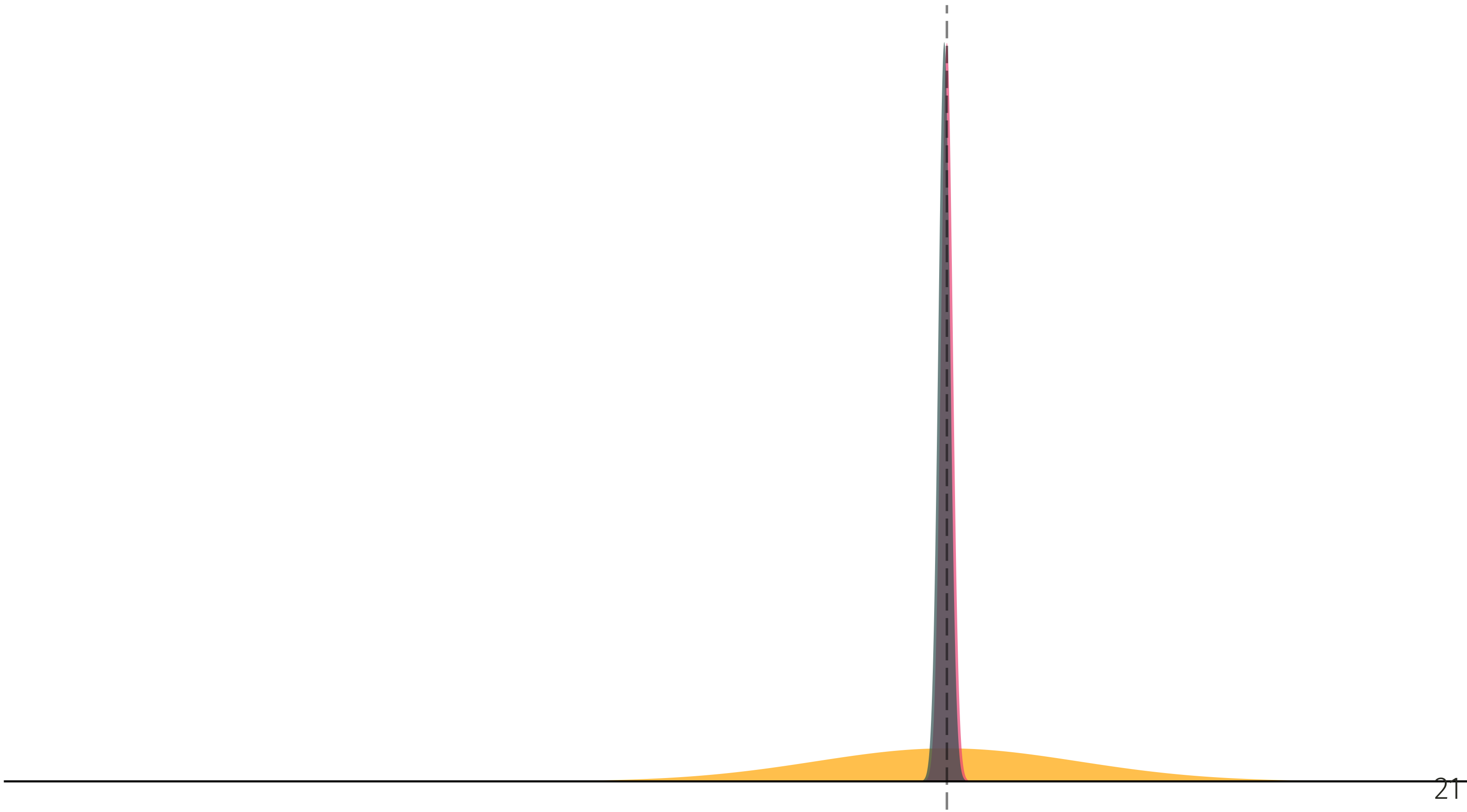
$n = 100$



Consistency

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

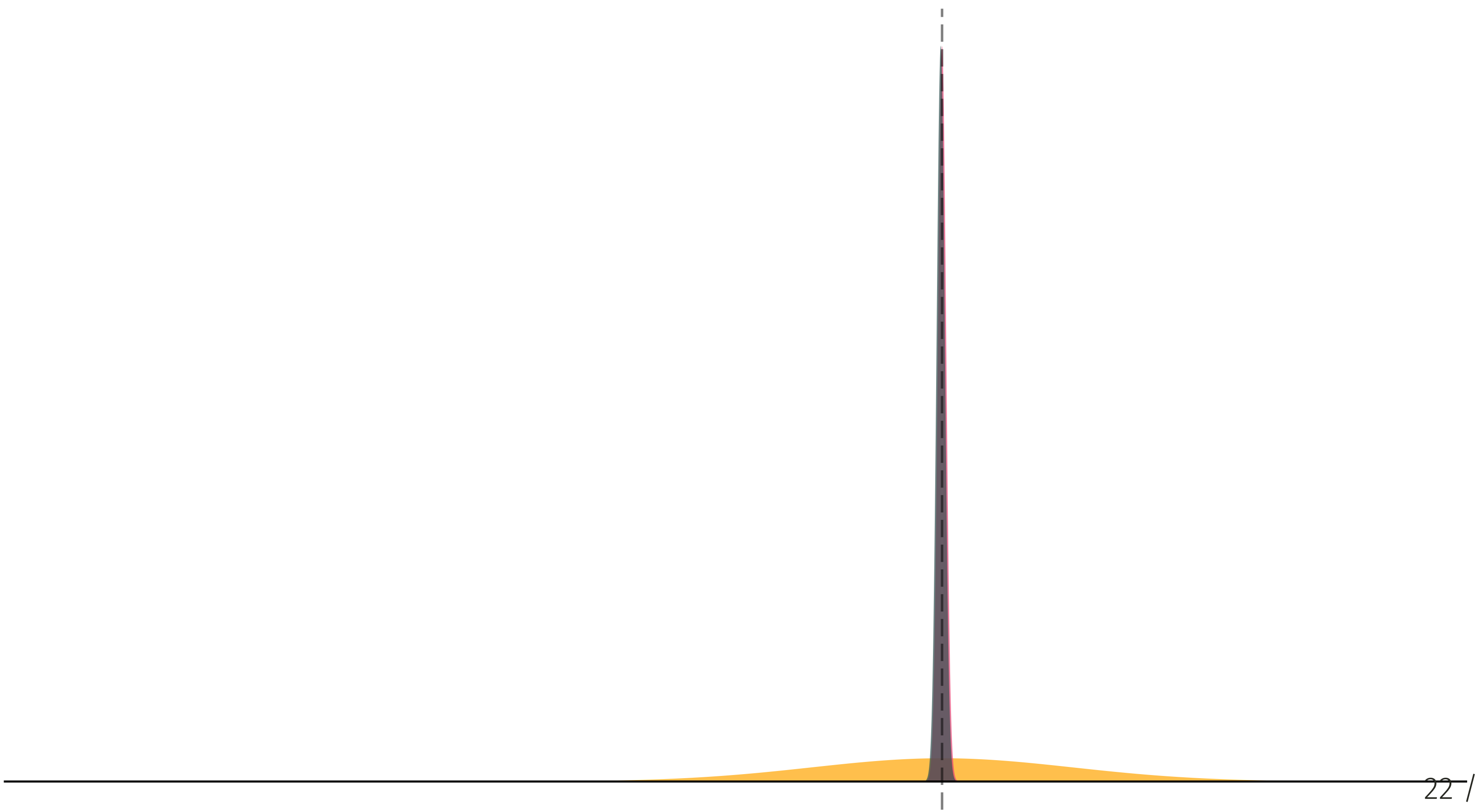
$n = 500$



Consistency

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

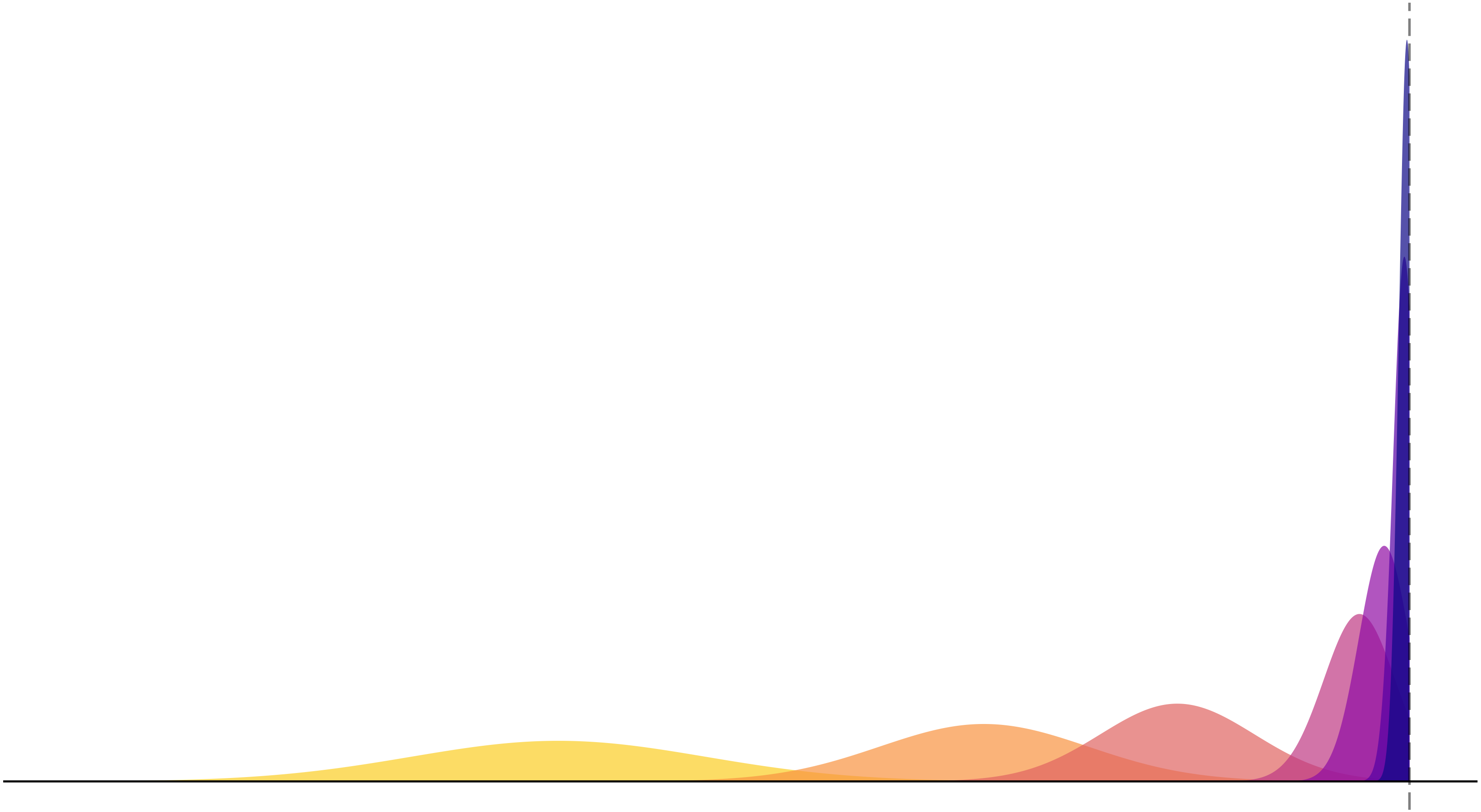
$n = 1000$



Consistency

The distributions of $\tilde{X}$

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



Consistency

The takeaway?

- An estimator can be unbiased without being consistent (e.g., X_1).
- An estimator can be unbiased and consistent (e.g., $\bar{X}$).
- An estimator can be biased but consistent (e.g., $\tilde{X}$).
- 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.*,

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

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

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

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

Proof of consistency

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

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

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

Proof of consistency

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

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

which, by the properties of probability limits, gives us

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

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

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

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 ∞, 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**), $\boldsymbol{E}\left[u_i \middle| X_i\right] = 0$.
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

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

Recall$_1$: **Omitted-variable bias** occurs when we omit a variable in our linear regression model (*e.g.*, leavining 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.*, $\mathrm{Cov}\left(x_1,\ x_2\right) \neq 0$.

Omitted-variable bias, redux

Inconsistency?

Imagine we have a population whose true model is

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

Recall₂: We defined the **bias** of an estimator W for parameter θ

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

Omitted-variable bias, redux

Inconsistency?

Imagine we have a population whose true model is

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

We know that omitted-variable bias causes biased estimates.

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

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

Omitted-variable bias, redux

Inconsistency?

Imagine we have a population whose true model is

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

but we instead specify the model as

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

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

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

Our question: Is $\hat{\beta}_1$ consistent for β_1 when we omit x_2?

$$\text{plim}\left(\hat{\beta}_1\right) \overset{?}{=} \beta_1$$

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 $\text{plim}\hat{\beta}_1 = \beta_1 + \dfrac{\text{Cov}\left(x_1,\ w\right)}{\text{Var}\left(x_1\right)}$

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

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

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

Omitted-variable bias, redux

Inconsistency?

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

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

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

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

Omitted-variable bias, redux

Inconsistency?

Thus, we find that

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

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.*, $\text{Cov}\left(x_1, x_2\right) \neq 0$.

Omitted-variable bias, redux

Signing the bias

Sometimes we're stuck with omitted variable bias.†

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

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

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

Signing the bias

Begin with

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

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

Signing the bias

Begin with

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

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

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

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

Omitted-variable bias, redux

Signing the bias

Begin with

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

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

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

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

Omitted-variable bias, redux

Signing the bias

Begin with

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

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

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

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

Omitted-variable bias, redux

Signing the bias

Thus, in cases where we have a sense of

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

2. the sign of β_2

we know in which direction inconsistency pushes our estimates.

Direction of bias

	$\text{Cov}(x_1, x_2) > 0$	$\text{Cov}(x_1, x_2) < 0$
$\beta_2 > 0$	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 $\boldsymbol{E}\left[\omega_i\right] = 0$, $\mathrm{Var}\left(\omega_i\right) = \sigma^2_\omega$, and ω is independent of z and u.

OLS regression of y and x will produce inconsistent estimates for β_1.

Measurement error

Proof

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

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 ε.

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

Measurement error

Proof

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

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

Proof

Putting the numerator and denominator back together,

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.

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