Sampling Distributions and the Law of Large Numbers

Gov 2001: Quantitative Social Science Methods I

Scott Cunningham

Harvard University

Spring 2026

Today's Reading

Required

- **Aronow & Miller**, §3.1–3.2.2: I.I.D., random sampling, WLLN (pp. 91–99)
- **Blackwell**, Ch. 3: Asymptotics (pp. 51–78)

The bridge: Everything so far was about *populations.* Now we ask: how do we learn about populations from *samples*?

The Fundamental Problem of Statistics

What we want: Population parameters

- Population mean: $\mu = \mathbb{E}[Y]$
- Population variance: $\sigma^2 = \text{Var}(Y)$
- Conditional expectation: $\mathbb{E}[Y|X]$

What we have: A sample of n observations

- $Y_1, Y_2, \ldots, Y_n$ drawn from the population

The question: Can we use the sample to learn about the population?

Yes—under the right conditions. That's what this week is about.

The I.I.D. Assumption

Independent and Identically Distributed (I.I.D.)

The sample $Y_1, Y_2, \ldots, Y_n$ is **I.I.D.** if:

1. **Independent**: $Y_i \perp\!\!\!\perp Y_j$ for all $i \neq j$
2. **Identically distributed**: Each Y_i has the same distribution F

Intuition: Each observation is a fresh, independent draw from the same population.

When does this hold?

- Simple random sampling with replacement
- Simple random sampling without replacement (approximately, if population is large)
- *Not* time series, clustered data, or convenience samples

Random Sampling in Practice

Example: We want to know average income in Massachusetts.

Population: All adults in Massachusetts (about 5.5 million)
Parameter: $\mu = \mathbb{E}[\text{Income}]$

How to sample?

- Get a list of all adults (sampling frame)
- Randomly select $n = 1{,}000$ people
- Measure their income: $Y_1, Y_2, \ldots, Y_{1000}$

If the sampling is truly random, the Y_i are approximately I.I.D.

In practice, sampling is never perfect. But I.I.D. is a useful approximation.

The Sample Mean as an Estimator

Natural idea: Estimate μ with the sample mean:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

Key insight: $\bar{Y}$ is itself a *random variable.*

- Different samples give different values of $\bar{Y}$
- $\bar{Y}$ has its own distribution (the **sampling distribution**)
- We want to understand this distribution

Statistics is about understanding how estimators behave across repeated samples.

Properties of the Sample Mean

Assume: $Y_1, \ldots, Y_n$ I.I.D. with $\mathbb{E}[Y_i] = \mu$ and $\text{Var}(Y_i) = \sigma^2$

Expected Value of $\bar{Y}$

$$\mathbb{E}[\bar{Y}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Y_i] = \frac{1}{n} \cdot n\mu = \mu$$

The sample mean is unbiased: On average, $\bar{Y}$ equals the true μ.

This follows from linearity of expectation—no independence needed!

Variance of the Sample Mean

Variance of $\bar{Y}$

$$\begin{aligned}\text{Var}(\bar{Y}) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) \\ &= \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} Y_i\right) \\ &= \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}\end{aligned}$$

Key insight: Variance shrinks as n increases!

- With $n = 100$: $\text{Var}(\bar{Y}) = \sigma^2/100$
- With $n = 10{,}000$: $\text{Var}(\bar{Y}) = \sigma^2/10{,}000$

Larger samples $\Rightarrow$ more precise estimates

The Standard Error

Standard Error of the Mean

$$\text{SE}(\bar{Y}) = \text{SD}(\bar{Y}) = \frac{\sigma}{\sqrt{n}}$$

Interpretation: The typical distance of $\bar{Y}$ from μ across repeated samples.

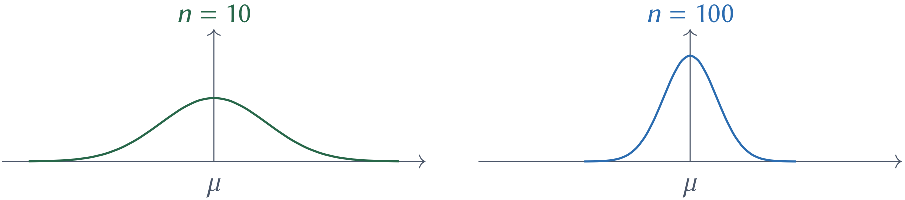
Example: If $\sigma = 20{,}000$ (income SD) and $n = 400$:

$$\text{SE}(\bar{Y}) = \frac{20{,}000}{\sqrt{400}} = \frac{20{,}000}{20} = 1{,}000$$

The sample mean is typically within $1,000 of the true mean.

To cut SE in half, you need 4 times the sample size.

Visualizing the Sampling Distribution



Both centered at μ (unbiased)

Larger $n \Rightarrow$ tighter distribution around μ

As $n \to \infty$, the distribution collapses to a spike at μ.

The Law of Large Numbers: Intuition

The question: What happens to $\bar{Y}$ as $n \to \infty$?

We know:

- $\mathbb{E}[\bar{Y}] = \mu$ (for any n)
- $\text{Var}(\bar{Y}) = \sigma^2/n \to 0$ as $n \to \infty$

Together: $\bar{Y}$ gets closer and closer to μ.

Informal Statement

Sample averages converge to population averages as sample size grows.

This is why statistics works!

Convergence in Probability

Definition

A sequence of random variables X_n **converges in probability** to a constant c, written $X_n \xrightarrow{p} c$, if:

$$\lim_{n \to \infty} \Pr(|X_n - c| > \varepsilon) = 0 \quad \text{for all } \varepsilon > 0$$

In words: The probability that X_n is far from c goes to zero.

Notation: We also write $\text{plim}\, X_n = c$ (probability limit).

This is a weaker notion than "X_n equals c for large n"—there's still randomness, but it gets negligible.

The Weak Law of Large Numbers

Weak Law of Large Numbers (WLLN)

If $Y_1, Y_2, \dots$ are I.I.D. with $\mathbb{E}[Y_i] = \mu$ and $\text{Var}(Y_i) = \sigma^2 < \infty$, then:

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{p} \mu$$

In words: The sample mean converges in probability to the population mean.

This is foundational: It tells us that with enough data, we can learn the truth.

Why the WLLN is True (Chebyshev's Proof)

Chebyshev's Inequality: For any random variable X with mean μ and variance σ^2:

$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Apply to $\bar{Y}$: $\mathbb{E}[\bar{Y}] = \mu$, $\text{Var}(\bar{Y}) = \sigma^2/n$

For any $\varepsilon > 0$:

$$\Pr(|\bar{Y} - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{Y})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0$$

The bound goes to zero as $n \to \infty$. QED.

What the LLN Means in Practice

The LLN justifies what we do as researchers:

- We collect data (a sample)
- We compute statistics (like $\bar{Y}$)
- We interpret them as estimates of population quantities

The LLN says this works: With enough data, our estimates get arbitrarily close to the truth.

But: The LLN is *asymptotic*. It doesn't tell us:

- How close we are for any *finite n*
- The shape of the sampling distribution
- How to construct confidence intervals

For that, we need the Central Limit Theorem (Wednesday).

Consistency of Estimators

Definition: Consistency

An estimator $\hat{\theta}_n$ is **consistent** for θ if:

$$\hat{\theta}_n \xrightarrow{p} \theta \quad \text{as } n \longrightarrow \infty$$

The LLN tells us: $\bar{Y}$ is a consistent estimator of μ.

Consistency is a minimal requirement:

- If an estimator isn't consistent, we're in trouble
- More data doesn't help us get the right answer
- Example: estimating μ with Y_1 (just the first observation)—unbiased but not consistent!

The Analogy Principle

A general strategy for estimation:

Analogy Principle (Plug-in Principle)

Estimate population quantities by replacing population distributions with sample distributions.

Examples:

- $\mu = \mathbb{E}[Y] \Rightarrow \hat{\mu} = \bar{Y} = \frac{1}{n}\sum Y_i$
- $\sigma^2 = \mathbb{E}[(Y - \mu)^2] \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum(Y_i - \bar{Y})^2$
- $\mathbb{E}[\text{Turnout}|\text{Party} = D] \Rightarrow$ sample mean turnout among observed Democrats

The LLN guarantees these plug-in estimators are consistent.

Example: Election Polling

Setup: Election between candidates A and B.

Parameter: $p = \Pr(\text{vote for A})$ in the population

Sample: Survey $n = 1{,}000$ voters. Let $Y_i = 1$ if voter i supports A.

Estimator: $\hat{p} = \bar{Y} = \frac{1}{n}\sum Y_i$ (sample proportion)

Properties:

- $\mathbb{E}[\hat{p}] = p$ (unbiased)
- $\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$
- $\text{SE}(\hat{p}) = \sqrt{p(1-p)/n}$
- If $p = 0.5$: $\text{SE} = \sqrt{0.25/1000} = 0.0158$ (about 1.6 percentage points)

Polls typically report "margin of error" $\approx 2 \times \text{SE}$.

What About Other Quantities?

The LLN applies to any average:

If $g(Y)$ is any function with $\mathbb{E}[g(Y)] < \infty$:

$$\frac{1}{n}\sum_{i=1}^{n} g(Y_i) \xrightarrow{p} \mathbb{E}[g(Y)]$$

Examples:

- $\frac{1}{n}\sum Y_i^2 \xrightarrow{p} \mathbb{E}[Y^2]$
- $\frac{1}{n}\sum (Y_i - \bar{Y})^2 \xrightarrow{p} \text{Var}(Y)$
- $\frac{1}{n}\sum X_i Y_i \xrightarrow{p} \mathbb{E}[XY]$

This is why sample covariances, sample variances, etc. are consistent.

Continuous Mapping Theorem

Useful result: Consistency is preserved by continuous functions.

Continuous Mapping Theorem

If $X_n \xrightarrow{p} c$ and g is continuous at c, then:

$$g(X_n) \xrightarrow{p} g(c)$$

Example: We know $\bar{Y} \xrightarrow{p} \mu$.
Since $g(x) = x^2$ is continuous:

$$\bar{Y}^2 \xrightarrow{p} \mu^2$$

This lets us prove consistency for complex estimators built from simple pieces.

Key Takeaways

1. **I.I.D.**: Observations are independent draws from the same distribution
2. **Sample mean** $\bar{Y}$ is unbiased: $\mathbb{E}[\bar{Y}] = \mu$
3. **Variance shrinks**: $\text{Var}(\bar{Y}) = \sigma^2/n$
4. **LLN**: $\bar{Y} \xrightarrow{p} \mu$ as $n \to \infty$
5. **Consistency**: Estimator converges to the true parameter
6. **Analogy principle**: Replace population with sample $\Rightarrow$ consistent estimators

Next: The Central Limit Theorem—what the sampling distribution looks like.

Looking Ahead

Wednesday: The Central Limit Theorem

- The LLN says $\bar{Y} \longrightarrow \mu$—but what's the *shape* of the distribution?
- CLT: For large n, $\bar{Y}$ is approximately normal
- This is the foundation of confidence intervals and hypothesis tests

Reading:

- A&M §3.2.3–3.2.4 (CLT, convergence concepts)
- Blackwell Ch. 3 (continue)