

# **Advanced Asymptotics**

Gov 2001: Quantitative Social Science Methods I

Scott Cunningham

Harvard University

Spring 2026

# Today's Reading

## Required

- **Aronow & Miller**, §3.2.4–3.2.5: Convergence, CLT, and Standard Errors
- **Blackwell**, §3.5: Slutsky's Theorem

## What we're doing:

- Two types of convergence (and why they matter)
- Slutsky's theorem: the workhorse of practical inference
- Why we can replace  $\sigma$  with  $\hat{\sigma}$  and still do valid inference

# The Big Picture

**Last week:** The CLT tells us that  $\bar{X}_n$  is approximately normal for large  $n$ .

**This week:** How do we actually *use* this in practice?

**The problem:**

- The CLT says:  $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$
- But we don't know  $\sigma$ !
- Can we just plug in  $\hat{\sigma}$ ? Yes, but why?

**Today:** The mathematical machinery that makes this work.

# Two Types of Convergence

As sample size  $n \rightarrow \infty$ , sequences of random variables can converge in different ways.

## Convergence in Probability

$$X_n \xrightarrow{p} c$$

“ $X_n$  gets arbitrarily close to  $c$ ”

The randomness disappears.

Example:  $\bar{X}_n \xrightarrow{p} \mu$  (LLN)

## Convergence in Distribution

$$X_n \xrightarrow{d} X$$

“ $X_n$ ’s distribution approaches  $X$ ’s distribution”

The randomness remains.

Example:  $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$  (CLT)

## Convergence in Probability: Formal Definition

### Definition (Convergence in Probability)

$X_n \xrightarrow{p} c$  if for any  $\epsilon > 0$ :

$$\lim_{n \rightarrow \infty} P(|X_n - c| > \epsilon) = 0$$

**In words:** The probability that  $X_n$  is “far” from  $c$  goes to zero.

### Key examples:

- Law of Large Numbers:  $\bar{X}_n \xrightarrow{p} \mathbb{E}[X]$
- Sample variance:  $S_n^2 \xrightarrow{p} \sigma^2$
- Any consistent estimator:  $\hat{\theta}_n \xrightarrow{p} \theta$

Convergence in probability means the estimator is **consistent**.

## Convergence in Distribution: Formal Definition

### Definition (Convergence in Distribution)

$X_n \xrightarrow{d} X$  if for all  $x$  where  $F_X$  is continuous:

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

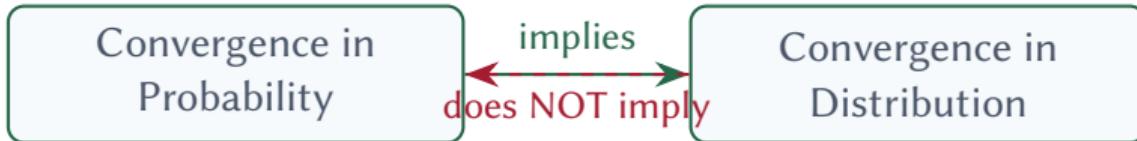
**In words:** The CDF of  $X_n$  approaches the CDF of  $X$  pointwise.

**Key example:** Central Limit Theorem

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

Note:  $X_n$  doesn't converge to a number—it converges to a *distribution*.

# The Relationship Between Them



**Convergence in probability  $\Rightarrow$  Convergence in distribution**, but NOT vice versa.

**Intuition:**

- If  $X_n$  converges to a constant  $c$ , its distribution collapses to a point mass at  $c$
- But a sequence can converge to a distribution without collapsing to a point

**Exception:** If  $X_n \xrightarrow{d} c$  (a constant), then  $X_n \xrightarrow{p} c$ .

# Why Does This Matter?

The CLT gives us:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$$

But we want to use:

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \approx N(0, 1)$$

**Question:** Is this valid? We replaced  $\sigma$  with  $S_n$ .

**Answer:** Yes! Because:

1.  $S_n \xrightarrow{P} \sigma$  (sample std dev is consistent)
2. Slutsky's theorem tells us this substitution is okay

## Slutsky's Theorem: The Workhorse

### Theorem (Slutsky)

If  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{p} c$  (a constant), then:

- $X_n + Y_n \xrightarrow{d} X + c$
- $X_n \cdot Y_n \xrightarrow{d} c \cdot X$
- $X_n / Y_n \xrightarrow{d} X/c$  (if  $c \neq 0$ )

**In words:** If you have something converging in distribution, you can:

- Add/subtract something converging to a constant
- Multiply/divide by something converging to a constant

...and the limiting distribution is what you'd expect.

# Why Slutsky's Theorem Is So Useful

The CLT says:

$$\underbrace{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}_{\xrightarrow{d} N(0,1)} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

We want to show:

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \xrightarrow{d} N(0, 1)$$

Rewrite as:

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} = \underbrace{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}_{\xrightarrow{d} N(0,1)} \cdot \underbrace{\frac{\sigma}{S_n}}_{\xrightarrow{p} 1}$$

# The Power of Slutsky

This is why practical inference works.

Without Slutsky:

- We'd need to know  $\sigma$  to use the CLT
- But  $\sigma$  is a population parameter—usually unknown
- Inference would require knowing something we're trying to learn about

With Slutsky:

- We can estimate  $\sigma$  with  $S_n$
- Since  $S_n \xrightarrow{P} \sigma$ , Slutsky says the substitution is valid
- The asymptotic distribution is unchanged

**Key insight:** Consistent estimators can be “plugged in” for population parameters in asymptotic arguments.

## Slutsky in Action: Confidence Intervals

Constructing a 95% CI for  $\mu$ :

**Step 1:** By CLT + Slutsky:

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \xrightarrow{d} N(0, 1)$$

**Step 2:** For large  $n$ , this is approximately standard normal:

$$P\left(-1.96 \leq \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \leq 1.96\right) \approx 0.95$$

**Step 3:** Rearrange to isolate  $\mu$ :

$$P\left(\bar{X}_n - 1.96 \cdot \frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + 1.96 \cdot \frac{S_n}{\sqrt{n}}\right) \approx 0.95$$

**Result:** The familiar CI formula  $\bar{X}_n \pm \widehat{SE}$ .

# A Warning: What Slutsky Does NOT Say

**Slutsky requires one sequence to converge to a constant.**

**NOT TRUE:**

- If  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} Y$ , then  $X_n + Y_n \xrightarrow{d} X + Y$
- This fails because we need to know the *joint* distribution

**Example:** Let  $X_n \xrightarrow{d} N(0, 1)$  and  $Y_n = -X_n$ .

- Both  $X_n \xrightarrow{d} N(0, 1)$  and  $Y_n \xrightarrow{d} N(0, 1)$
- But  $X_n + Y_n = 0$  for all  $n$ !
- So  $X_n + Y_n \xrightarrow{d} 0$ , not  $N(0, 2)$

Slutsky works because converging to a *constant* eliminates this joint distribution problem.

# The Continuous Mapping Theorem

A close relative of Slutsky:

## Theorem (Continuous Mapping)

If  $X_n \xrightarrow{d} X$  and  $g$  is a continuous function, then:

$$g(X_n) \xrightarrow{d} g(X)$$

Similarly, if  $X_n \xrightarrow{p} c$ , then  $g(X_n) \xrightarrow{p} g(c)$ .

**In words:** Continuous functions preserve convergence.

**Examples:**

- If  $\bar{X}_n \xrightarrow{p} \mu$ , then  $\bar{X}_n^2 \xrightarrow{p} \mu^2$
- If  $\bar{X}_n \xrightarrow{p} \mu$  and  $\mu > 0$ , then  $\log(\bar{X}_n) \xrightarrow{p} \log(\mu)$
- If  $\hat{\theta}_n \xrightarrow{p} \theta$ , then  $h(\hat{\theta}_n) \xrightarrow{p} h(\theta)$  for continuous  $h$

## CMT + Slutsky: A Powerful Combination

Many practical results combine these theorems.

**Example:** Proving  $S_n^2$  is consistent for  $\sigma^2$ .

We know:

- $\bar{X}_n \xrightarrow{P} \mu$  (LLN)
- $\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{P} \mathbb{E}[X^2]$  (LLN applied to  $X^2$ )

Since  $S_n^2 = \frac{1}{n} \sum X_i^2 - \bar{X}_n^2$ :

By CMT:  $\bar{X}_n^2 \xrightarrow{P} \mu^2$

By Slutsky (difference):  $S_n^2 \xrightarrow{P} \mathbb{E}[X^2] - \mu^2 = \sigma^2 \quad \checkmark$

This is why we can trust  $S_n^2$  as an estimator of  $\sigma^2$ .

# Multivariate CLT

The CLT extends to vectors.

## Theorem (Multivariate CLT)

Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be iid random vectors in  $\mathbb{R}^k$  with:

- $\mathbb{E}[\mathbf{X}_i] = \boldsymbol{\mu}$
- $\text{Var}(\mathbf{X}_i) = \boldsymbol{\Sigma}$  (the variance-covariance matrix)

Then:

$$\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$$

**In words:** The sample mean vector is asymptotically multivariate normal.

**Why this matters:** Regression has multiple coefficients. We need their joint distribution.

# The Variance-Covariance Matrix

For a random vector  $\mathbf{X} = (X_1, \dots, X_k)'$ :

$$\Sigma = \text{Var}(\mathbf{X}) = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_k) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_k, X_1) & \text{Cov}(X_k, X_2) & \cdots & \text{Var}(X_k) \end{pmatrix}$$

## Properties:

- Diagonal: variances of each component
- Off-diagonal: covariances between components
- Symmetric:  $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$
- Positive semi-definite:  $\mathbf{a}'\Sigma\mathbf{a} \geq 0$  for any  $\mathbf{a}$

We'll use this extensively when we get to regression.

# Why This All Matters for Regression

**Next week:** We start regression. Today's tools are foundational.

**OLS estimator  $\hat{\beta}$ :**

- $\hat{\beta} \xrightarrow{P} \beta$  (consistency, via LLN)
- $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}, \mathbf{V})$  (asymptotic normality, via multivariate CLT)
- We estimate  $\mathbf{V}$  and use Slutsky to justify plugging it in

**All regression inference** — standard errors, t-tests, confidence intervals — relies on:

1. CLT (asymptotic normality)
2. Slutsky (plug-in estimated variances)
3. Continuous Mapping (functions of estimators)

## Summary: The Asymptotic Toolkit

---

Tool	What It Does	When You Use It
LLN	$\bar{X}_n \xrightarrow{p} \mu$	Proving consistency
CLT	$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$	Asymptotic normality
Slutsky	Plug in consistent estimators	Replacing $\sigma$ with $\hat{\sigma}$
CMT	$g(X_n) \xrightarrow{d} g(X)$	Functions of estimators

---

**Together:** These four results are the foundation of frequentist inference.

## For Wednesday

### Reading:

- **Blackwell**, §3.6: The Delta Method
- A&M review: §3.2.5–3.2.6

### Coming up:

- The delta method: asymptotic distribution of  $h(\hat{\theta})$
- Why this matters: risk ratios, odds ratios, any nonlinear function
- Then: midterm review

The inference machinery is now complete. Time to apply it!