The Delta Method & Working Session

Completing the asymptotic toolkit — and practicing for Problem Set 6

Gov 2001 · Spring 2026 · Scott Cunningham

Where we are after Monday

Monday's big ideas:

- Every sample gives a different $\bar{Y}_n$ — the sampling distribution captures that variability
- The CLT tells us the *shape*: approximately normal for large n
- We rearranged the CLT to build confidence intervals
- Slutsky lets us plug in $\hat{\sigma}$ for the unknown σ

The CI formula:

$$\bar{Y}_n \ \pm \ z_{\alpha/2} \times \frac{\hat{\sigma}}{\sqrt{n}}$$

This is the engine. Everything today extends it.

Two ways to build a CI — and when to use each

	Chebyshev CI	**CLT CI**
Critical value (95%)	$\sqrt{1/\alpha} = 4.47$	$z_{0.025} = 1.96$
Assumption	None	Large n
Width	Wider	Tighter
When to use	Small n, any shape	Large n

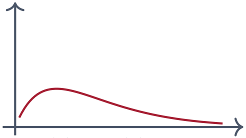
Both are valid. The CLT buys you precision by assuming the normal approximation has kicked in.

What is normal and what is not

The CLT is about the statistic, not the data

Individual treatment effects τ_i

The distribution in the population:

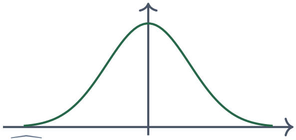


$\tau_i = Y_i(1) - Y_i(0)$ across people

- Skewed, bimodal, discrete
- **Not normal** in general
- Fixed — more data won't change it

Estimated $\widehat{ATE}$ across samples

Normal for large n (CLT):



$\widehat{ATE}$ if we re-ran the experiment

- Symmetric, bell-shaped
- **Always normal** for large n
- Even though τ_i is skewed

The CLT does not say the data become normal

A common confusion: "In large samples, everything is normal"

WRONG: "The CLT says the population distribution approaches a normal as n grows"

The population distribution is *fixed*. Income is skewed at $n = 100$ and still skewed at $n = 1{,}000{,}000$. Collecting more data doesn't change the shape of the world

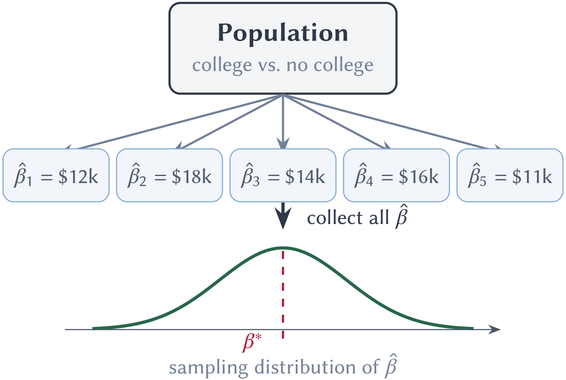
The CLT says: the *sampling distribution of* $\bar{Y}_n$ approaches a normal as n grows

What changes with n: the distribution of the *statistic* we compute, not the distribution of the *data* we observe

The asymptotics always work

College wages example: the sampling distribution of a difference in means

Calculate: $\hat{\beta} = \bar{Y}_{\text{college}} - \bar{Y}_{\text{no college}}$ in each sample



LLN: $\hat{\beta} \xrightarrow{p} \beta^*$.   CLT: normally distributed around β^*.   CI: covers β^*

What is at the center?

The estimand is the population difference in means

β^* is the **estimand**: the population version of the calculation

$\beta^* = \mathbb{E}[\text{wages} \mid \text{college}] - \mathbb{E}[\text{wages} \mid \text{no college}]$

This is just "everybody with college minus everybody without college" in the population

The CLT says $\hat{\beta}$ is:

- Centered on β^* (unbiased)
- Normally distributed (CLT)
- With a variance that shrinks as n grows (law of large numbers)

So far, everything works. But we haven't asked the important question.

What is *not* at the center?

The causal effect of college on earnings

β^* is *not* the causal effect of attending college

People who attend college differ from those who don't — in ability, family background, motivation. The population difference in means captures all of that, not just the effect of college itself

So you get a paradox:

- $\hat{\beta}$ is an **unbiased estimator** of β^* ✓
- The CI covers β^* 95% of the time ✓
- But β^* is almost never the quantity you actually care about ×

Precise estimation of the wrong thing.

Start with the estimand, not the estimator

The CLT tells you $\hat{\beta}$ is centered on β^* and normally distributed. It does *not* tell you that β^* is the thing you care about

The estimand is the population quantity you want to learn

The estimator is the calculation you run on your sample

The asymptotics connect the two — but only if they are the *same thing*

This is why Part II of the course begins with: *what are you trying to estimate?*

Regression, causal inference, and research design are all about choosing estimators whose β^* matches the question you're asking.

The correct interpretation of a 95% CI

If we repeated the sampling procedure many times, 95% of the resulting intervals would contain the true parameter μ

It is a statement about the *procedure*, not about any single interval

Values inside the interval are *consistent with* the data. Values far outside are not

WRONG: "There is a 95% probability that μ is in this interval"

μ is fixed. The interval is random. The probability is about the interval, not about μ.

Part I: The Delta Method

The last tool in the asymptotic toolkit

Sometimes we care about a *function* of μ, not μ itself

The problem

- $\bar{Y}_n$ estimates μ (Monday)
- $g(\bar{Y}_n)$ estimates $g(\mu)$ (today)

But: the CLT gives us the distribution of $\bar{Y}_n$, not $g(\bar{Y}_n)$

Examples from social science:

- Average income is μ, but you want a CI for $\log(\mu)$
  (log scale for skewed income data)
- A proportion is p, but you want a CI for $\frac{p}{1-p}$
  (the odds — used in logistic regression)
- An effect is μ, but you want a CI for μ^2
  (squared effect size — common in meta-analysis)

The difficulty: the CLT tells us the distribution of $\bar{Y}$, not of $g(\bar{Y})$

What does "smooth" mean?

The key requirement for the delta method

Smooth = the first derivative $g'(\mu)$ exists and $g'(\mu) \neq 0$

- $g(\mu) = \log(\mu) \quad \longrightarrow \quad g'(\mu) = 1/\mu \quad \checkmark$
- $g(\mu) = e^{\mu} \quad \longrightarrow \quad g'(\mu) = e^{\mu} \quad \checkmark$
- $g(\mu) = \mu^2 \quad \longrightarrow \quad g'(\mu) = 2\mu \quad \checkmark$ (if $\mu \neq 0$)

Why this matters: a smooth function is locally a straight line

A straight line rescales a normal but keeps it normal — that's the whole trick

When it fails: nearest neighbor matching is *not* smooth — small changes in the data reshuffle which units match, and the function jumps discontinuously

We need a way to “transfer” the CLT through a function

Building the intuition

What we know: $\bar{Y}_n$ is approximately normal around μ (the CLT)

What we want: the distribution of $g(\bar{Y}_n)$ around $g(\mu)$

Key insight: if g is smooth, then near μ it's approximately a straight line

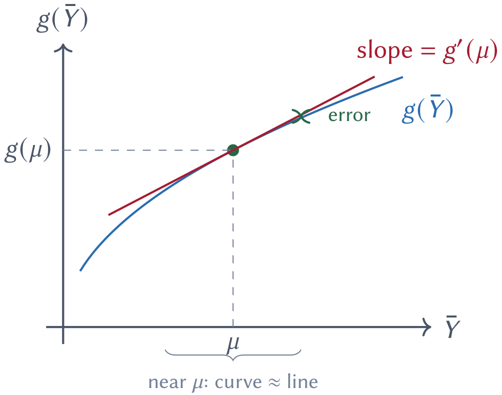
A straight line just *rescales* the normal — stretch it or compress it, but it stays normal

The slope $g'(\mu)$ tells us how much the transformation stretches or compresses:

- Steep slope $\Rightarrow$ small changes in $\bar{Y}$ become large changes in $g(\bar{Y}) \Rightarrow$ *more* variance
- Flat slope $\Rightarrow$ changes in $\bar{Y}$ are dampened $\Rightarrow$ *less* variance

That's the delta method: a linear approximation that preserves normality.

Visual intuition: the tangent line approximation



Near μ, the curve ≈ its tangent line. The CLT concentrates $\bar{Y}_n$ near μ, so this works for large n

The delta method says $g(\bar{Y}_n)$ is also asymptotically normal

The theorem

Start with: $\sqrt{n}(\bar{Y}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$ (the CLT)

If g is differentiable at μ with $g'(\mu) \neq 0$:

$$\boxed{\sqrt{n}\big(g(\bar{Y}_n) - g(\mu)\big) \xrightarrow{d} N\big(0,\ [g'(\mu)]^2\sigma^2\big)}$$

We already expected $g(\bar{Y}_n)$ to be approximately normal — smooth functions preserve normality. The news here is the variance formula: $[g'(\mu)]^2\sigma^2$. Without it, we'd know the shape but not the width — and without the width, no CI.

The derivative controls how much the variance changes

What $[g'(\mu)]^2$ does

The CLT gave us variance σ^2. The delta method gives us variance $[g'(\mu)]^2\sigma^2$.

Steep slope ($|g'(\mu)|$ large): the function *amplifies* small changes in $\bar{Y}_n \longrightarrow$ more variance

Flat slope ($|g'(\mu)|$ small): the function *dampens* small changes in $\bar{Y}_n \longrightarrow$ less variance

This is the whole point of the delta method — $g'(\mu)$ converts the variance of $\bar{Y}_n$ into the variance of $g(\bar{Y}_n)$. That conversion factor is what lets us build the CI.

The delta method CI has the same recipe as before

Estimate ± critical value × SE

Monday's CI for μ:

$$\bar{Y}_n \ \pm \ z_{\alpha/2} \times \frac{\hat{\sigma}}{\sqrt{n}}$$

Delta method CI for $g(\mu)$:

$$g(\bar{Y}_n) \ \pm \ z_{\alpha/2} \times |g'(\bar{Y}_n)| \times \frac{\hat{\sigma}}{\sqrt{n}}$$

The only change: the SE gets multiplied by $|g'(\bar{Y}_n)|$ — the slope evaluated at our sample estimate

Example: CI for log average wages

Setup

Data: $n = 400$ workers, $\bar{Y}_n = \$15{,}000$, $\hat{\sigma} = \$6{,}000$

Goal: a 95% CI for $\ln(\mu)$ — log average wages in the population

Our function: $g(\mu) = \ln(\mu), \quad \text{so} \quad g'(\mu) = 1/\mu$

Step 1: point estimate and derivative

Log wages example

Point estimate:

$g(\bar{Y}_n) = \ln(15{,}000) = 9.62$

Derivative evaluated at our sample estimate:

$$|g'(\bar{Y}_n)| = \frac{1}{\bar{Y}_n} = \frac{1}{15{,}000}$$

This is the conversion factor — it translates the variance of $\bar{Y}_n$ into the variance of $\ln(\bar{Y}_n)$

Step 2: the delta method SE

Log wages example

Formula: $\text{SE} = |g'(\bar{Y}_n)| \times \dfrac{\hat{\sigma}}{\sqrt{n}}$

Plug in:

$$\text{SE} = \frac{1}{15{,}000} \times \frac{6{,}000}{\sqrt{400}} = \frac{1}{15{,}000} \times 300 = 0.02$$

Without the delta method, we'd know $\ln(\bar{Y}_n)$ *is approximately normal but not how wide — now we know the SE is* 0.02

Step 3: the 95% CI

Log wages example

Formula: $g(\bar{Y}_n) \;\pm\; 1.96 \times \mathrm{SE}$

$$9.62 \;\pm\; 1.96 \times 0.02 \;=\; 9.62 \;\pm\; 0.04$$

$$\boxed{95\%\text{ CI for } \ln(\mu): \quad [9.58,\; 9.66]}$$

Under repeated sampling, 95% of intervals constructed this way would contain the true $\ln(\mu)$

What does that CI mean in dollars?

Back-transforming the log CI

Our CI for $\ln(\mu)$: $[9.58,\ 9.66]$

Exponentiate both endpoints:
$[e^{9.58},\ e^{9.66}] = [\$14{,}474,\ \$15{,}643]$

Notice: the CI is not symmetric around $\$15{,}000$ *— that's because* $\ln$ *compresses large values and stretches small ones. This asymmetry is a feature, not a bug, for skewed data like wages.*

Why does the derivative appear?

The Taylor expansion in three lines

Near μ, any smooth function is approximately linear:

$$g(\bar{Y}_n) \approx g(\mu) + g'(\mu)(\bar{Y}_n - \mu)$$

Rearrange:

$$g(\bar{Y}_n) - g(\mu) \approx g'(\mu) \cdot (\bar{Y}_n - \mu)$$

Multiply by $\sqrt{n}$:

$$\sqrt{n}\big(g(\bar{Y}_n) - g(\mu)\big) \approx g'(\mu) \cdot \underbrace{\sqrt{n}(\bar{Y}_n - \mu)}_{\xrightarrow{d} N(0,\sigma^2)}$$

Scaling a normal by $g'(\mu)$ gives another normal: variance gets multiplied by $[g'(\mu)]^2$

The delta method recipe

Four steps you'll use on the problem set

Step 1: Write down your transformation $g(\cdot)$ and compute $g'(\cdot)$

Step 2: Start from the CLT for $\bar{Y}_n$ (or whatever estimator you have)

Step 3: Apply the theorem — multiply the asymptotic variance by $[g'(\mu)]^2$

Step 4: Plug in $\bar{Y}_n$ for μ and $\hat{\sigma}$ for σ (Slutsky) to get a feasible CI

$$\boxed{g(\bar{Y}_n) \;\pm\; 1.96 \times |g'(\bar{Y}_n)| \times \frac{\hat{\sigma}}{\sqrt{n}}}$$

This is the same engine as Monday's CI, with one extra gear: the derivative.

Part II: Practice Problems

Building the muscles

Before we start: what are we actually practicing?

Something I thought would help

I want to try something I haven't done yet this semester.

Before we work through problems, I want to name the *types of thinking* they require — because the pset isn't testing one skill, it's testing several:

- **Recall**: Can you write down a formula from memory?
- **Plug-in calculation**: Can you apply that formula to numbers?
- **Reframing**: Can you see that a question is really asking something you already know how to answer, just in disguise?
- **Applying a theorem**: Can you use the CLT or delta method on a new distribution?
- **Chaining tools**: Can you use LLN, then CLT, then Slutsky, then delta method *in sequence*?

The reason I bring this up

When I was learning this material, I would get stuck on a problem and not know *why* I was stuck.

Was it that I didn't know the formula? Or that I knew the formula but couldn't see where to use it? Or that I needed to use three things in a row and lost track of where I was?

Those are different problems. And the fix for each is different.

As we work through examples today, I'll try to name which kind of move we're making each time — not to give you the answers, but so you can recognize the moves when you see them again

Bounding tail probabilities for commute times

Setup

Context: Commute times in a city follow $X \sim \text{Exp}(1/20)$

So $\mathbb{E}[X] = 20$ minutes, $\quad \text{Var}(X) = 400$

Question: What fraction of commuters could take ≥ 60 minutes?

We'll answer this three ways, each using more information than the last

Markov bound: use only the mean

Commute times example

All we use: $\mathbb{E}[X] = 20$

$$\mathbb{P}(X \geq 60) \leq \frac{\mathbb{E}[X]}{60} = \frac{20}{60} = 0.333$$

At most a third of commuters take ≥ 60 *minutes — a loose bound, but we only needed the mean*

Chebyshev bound: add the variance

Commute times example

Now we also use: $\text{Var}(X) = 400$

The trick: Chebyshev bounds deviations from the mean — so rewrite our question that way

The mean is 20. We're asking about 60. How far apart are they?
$60 - 20 = 40, \quad \text{so} \quad X \geq 60 \implies |X - 20| \geq 40$

Now Chebyshev applies:

$$\mathbb{P}(|X - 20| \geq 40) \leq \frac{\text{Var}(X)}{40^2} = \frac{400}{1600} = 0.25$$

Tighter: at most 25%. Adding the variance bought us 8 percentage points.

The pattern: more information gives tighter bounds

Method	Bound on $\mathbb{P}(X \geq 60)$	Information used
Markov	≤ 0.333	Mean only
Chebyshev	≤ 0.250	Mean + variance
Exact (Exponential)	$= e^{-3} \approx 0.050$	Full distribution

Strategy for problems like this:

- Markov only needs $\mathbb{E}[X]$ and $X \geq 0$
- Chebyshev needs $\mathbb{E}[X]$ and $\text{Var}(X)$ — centers on the mean
- To use Chebyshev for a *one-sided* bound, rewrite as a two-sided event first

But wait — we computed those bounds from sample data

The Markov bound says $\mathbb{P}(X \geq 60) \leq E[X]/60$

But we don't know $E[X]$. We estimated it with $\bar{X}_n = 20$.

Don't overcomplicate this. The bound is a population quantity: $E[X]/60$. In the population, it's just a number. But we don't have the population — we have a sample.

So $\bar{X}_n/60$ is our *estimate* of the bound. And it's a statistic.

What do we know about statistics computed from samples?

The estimated bound has a sampling distribution — like everything else

$\bar{X}_n/60$ is a function of $\bar{X}_n$. A different sample gives a different $\bar{X}_n$, which gives a different estimated bound.

By the CLT: $\bar{X}_n$ is approximately normal for large n
So: $\bar{X}_n/60$ is approximately normal too — it's just a linear rescaling

The same is true for the Chebyshev bound. $\hat{\sigma}^2/t^2$ is a sample quantity. It varies across samples. It converges to the population bound $\text{Var}(X)/t^2$. The CLT gives it a shape.

This is the same move. It's always the same move.

What does it mean to have a CI for a bound?

The bound is a population quantity. You estimated it from a sample. The CI tells you how uncertain that estimate is.

This is a bit strange: you'd have precise knowledge of an imprecise thing — a tight CI for a loose bound.

Every quantity you compute from a sample — a mean, a variance, a bound, a transformed parameter — has this same structure. That's really what the rest of today is about.

CLT in action: average wait times

Setting up the problem

Context: Wait times at a government office. We model each person's wait as:
$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$

What do we know about this distribution?
$\mathbb{E}[X_i] = 1/\lambda, \qquad \text{Var}(X_i) = 1/\lambda^2$

Goal: a 95% CI for the average wait time $1/\lambda$

Ask yourself: what's the estimator? What tool gives us its distribution?

The natural estimator is $\bar{X}$ — and the CLT gives its shape

Wait times: applying the CLT [skill: apply a theorem]

Estimator: $\hat{\mu} = \bar{X}$ (the sample average wait time)

The CLT says: center is $1/\lambda$, variance is $\text{Var}(X_i) = 1/\lambda^2$

$$\sqrt{n}\left(\bar{X} - \frac{1}{\lambda}\right) \xrightarrow{d} N\left(0,\ \frac{1}{\lambda^2}\right)$$

Notice: the variance $1/\lambda^2$ depends on the unknown parameter. We can't use this directly.

Replace the unknown variance with an estimate

Wait times: Slutsky [skill: chaining tools]

The problem: $1/\lambda^2$ is unknown

The fix: Slutsky says we can replace it with $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Why does this work? Because $\hat{\sigma}^2 \xrightarrow{p} 1/\lambda^2$ by the LLN

95% CI for $1/\lambda$:

$$\bar{X} \ \pm \ 1.96 \times \frac{\hat{\sigma}}{\sqrt{n}}$$

We used three tools in sequence: LLN (consistency of $\hat{\sigma}$), CLT (shape), Slutsky (substitution). That's chaining.

Strategy: consistency before asymptotics

The order matters

When a problem asks for the asymptotic distribution, the steps are always:

Step 1: Show **consistency** (LLN)

- $\bar{X} \xrightarrow{p} \mathbb{E}[X_i]$ — this is almost always the LLN

Step 2: Write down the **CLT**

- $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$ — identify $\sigma^2 = \text{Var}(X_i)$

Step 3: If needed, apply **Slutsky** to replace unknowns

- $\hat{\sigma} \xrightarrow{p} \sigma$ lets you substitute

Step 4: If you need a *function* of μ, apply the **delta method**

Delta method: CI for the rate $\lambda = 1/\mu$

Setting up the problem [skill: reframing]

Context: Same wait time data. We estimated $\hat{\mu} = \bar{X}$ (average wait time).

But a hospital administrator wants to know the *rate*: how many patients per minute? That's $\lambda = 1/\mu$.

The reframing: we already have a CI for μ. Now we need a CI for $g(\mu) = 1/\mu$.

Which tool do we reach for? The function g is smooth, so — delta method.

Step 1: identify the function and its derivative

Rate CI [skill: recall]

$g(\mu) = 1/\mu = \mu^{-1}$

$g'(\mu) = -\mu^{-2} = -1/\mu^2$

The derivative is negative — meaning the rate goes down as the mean goes up. That makes sense: longer average waits = fewer patients per minute.

Step 2: write down the CLT, then apply delta method

Rate CI [skill: apply a theorem]

CLT: $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \mu^2)$
(for Exponential, $\text{Var}(X_i) = 1/\lambda^2 = \mu^2$)

Delta method: multiply variance by $[g'(\mu)]^2 = 1/\mu^4$

$$\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\mu}\right) \xrightarrow{d} N\left(0,\ \frac{1}{\mu^4} \cdot \mu^2\right) = N\left(0,\ \frac{1}{\mu^2}\right)$$

The μ^4 and μ^2 simplify nicely here. That won't always happen — enjoy it when it does.

Step 3: build the CI and plug in numbers

Rate CI   [skill: calculation]

Data: $n = 50$ patients, $\bar{X} = 25$ minutes

Point estimate: $1/\bar{X} = 1/25 = 0.04$ patients/min

Delta method SE: $\dfrac{1}{\bar{X}\sqrt{n}} = \dfrac{1}{25 \times \sqrt{50}} = 0.00566$

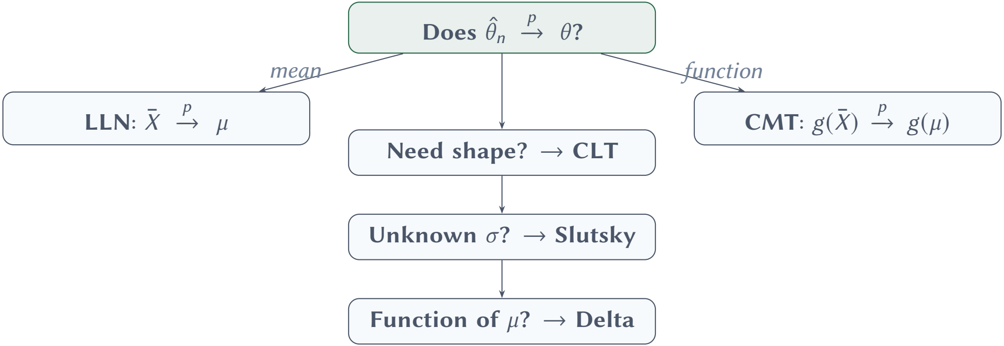
95% CI:

$$\boxed{0.04 \;\pm\; 1.96 \times 0.00566 \;=\; [0.029,\; 0.051]}$$

The administrator can say: the rate is between 0.029 and 0.051 patients per minute, or roughly 2–3 per hour.

Strategy: which tool do I need?

A decision tree



Strategy: the standard proof template

Most asymptotic problems follow this structure

To show $\hat{\theta}_n$ is asymptotically normal:

1. **Identify** the estimator as a sample average (or function of averages)
2. **Compute** $\mathbb{E}[\hat{\theta}_n]$ and $\text{Var}(\hat{\theta}_n)$
3. **Apply LLN** for consistency: $\hat{\theta}_n \xrightarrow{p} \theta$
4. **Apply CLT** for the shape: $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, V)$
5. **If** estimating a function $g(\theta)$: delta method scales the variance by $[g'(\theta)]^2$
6. **If** V is unknown: plug in $\hat{V}$ via Slutsky

The CI always has the same form:

$$\boxed{\hat{\theta}_n \ \pm \ z_{\alpha/2} \times \widehat{\text{SE}}(\hat{\theta}_n)}$$

Chebyshev vs. CLT: estimating average household size

Setup

Context: Census survey. Household size $X_i \overset{\text{iid}}{\sim} \text{Uniform}(1, 7)$

$$\mathbb{E}[X_i] = 4, \quad \text{Var}(X_i) = \frac{(7-1)^2}{12} = 3$$

$$n = 200, \quad \bar{X} = 4.15, \quad \hat{\sigma}^2 = 2.9$$

We'll build the same CI two ways. Same data, same goal — different assumptions.

Chebyshev CI: no distributional assumption

Household size [skill: apply a formula]

Recall: Chebyshev uses $k = 1/\sqrt{\alpha} = 1/\sqrt{0.05} = 4.47$ for 95%

$$\bar{X} \pm 4.47 \times \frac{\hat{\sigma}}{\sqrt{n}} = 4.15 \pm 4.47 \times \frac{\sqrt{2.9}}{\sqrt{200}}$$

$$= 4.15 \pm 4.47 \times 0.120 = 4.15 \pm 0.538$$

Chebyshev 95% CI: $[3.61,\ 4.69]$

CLT CI: assume normality of $\bar{X}$ for large n

Household size [skill: apply a formula]

Same data, same SE — only the critical value changes: 1.96 instead of 4.47

$$\bar{X} \pm 1.96 \times \frac{\hat{\sigma}}{\sqrt{n}} = 4.15 \pm 1.96 \times 0.120$$

$$= 4.15 \pm 0.236$$

$$\boxed{\text{CLT 95\% CI: } [3.91,\ 4.39]}$$

Less than half as wide. The normal assumption buys a lot of precision.

Delta method: CI for log average income

A common transformation in economics

Context: Household income survey, $n = 500$, $\bar{X} = \$52{,}000$, $\hat{\sigma} = \$28{,}000$

Want a CI for $\log(\mu)$ (log scale reduces influence of outliers)

Step 1: $g(\mu) = \log(\mu), \quad g'(\mu) = 1/\mu$

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

Step 3: Delta method $\Rightarrow \sqrt{n}(\log \bar{X} - \log \mu) \xrightarrow{d} N\left(0, \frac{\sigma^2}{\mu^2}\right)$

Step 4: 95% CI for $\log(\mu)$:

$$\log(52{,}000) \pm 1.96 \times \frac{28{,}000}{52{,}000 \times \sqrt{500}} = 10.859 \pm 1.96 \times 0.0241 = [10.81,\ 10.91]$$

The full asymptotic toolkit

Tool	What it does
LLN	$\bar{Y}_n \xrightarrow{p} \mu$ (consistency)
CMT	$g(\bar{Y}_n) \xrightarrow{p} g(\mu)$ (preserve convergence)
CLT	$\sqrt{n}(\bar{Y}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$ (shape)
Slutsky	Plug in $\hat{\sigma} \xrightarrow{p} \sigma$ (make it feasible)
Delta method	$g(\bar{Y}_n)$ is asymptotically normal with variance $[g']^2\sigma^2$

This is the complete toolkit for frequentist inference

Everything in regression (Part II of the course) will build on these five results

Problem Set 6 is due March 31

Skills you'll need:

- Bounding tail probabilities with Markov and Chebyshev
- Showing consistency via LLN
- Deriving asymptotic distributions via CLT
- Applying the delta method for functions of estimators
- Constructing CIs both ways: Chebyshev (conservative) and CLT (tighter)
- Chaining tools: LLN $\rightarrow$ CLT $\rightarrow$ Slutsky $\rightarrow$ delta method

Everything we covered Monday and today builds these muscles

Work with each other. The struggle is where the learning happens