The Delta Method

Gov 2001: Quantitative Social Science Methods I

Scott Cunningham

Harvard University

Spring 2026

Today's Reading

Required

- **Blackwell**, §3.6: The Delta Method
- **Aronow & Miller**, Review §3.2.5–3.2.6

What we're doing:

- The delta method: finding the asymptotic distribution of $h(\hat{\theta})$
- Why this matters: many quantities of interest are *nonlinear* functions
- The multivariate delta method: for functions of multiple parameters

Last inference lecture before the midterm!

The Problem

We know the CLT gives us:

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

But what if we want:

- The distribution of $\log(\bar{X}_n)$?
- The distribution of $\bar{X}_n^2$?
- The distribution of $1/\bar{X}_n$?
- The distribution of $\exp(\hat{\beta})$ (for odds ratios)?

The question: If $\hat{\theta}_n$ is asymptotically normal, what is the asymptotic distribution of $h(\hat{\theta}_n)$?

Answer: The delta method.

Real Applications

Why do we need nonlinear functions of estimators?

- **Risk ratios**: $\hat{p}_1/\hat{p}_0$ (ratio of two proportions)
- **Odds ratios**: $\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_0/(1-\hat{p}_0)}$
- **Log-transformed variables**: CI for μ when $Y = \log(X)$
- **Elasticities**: $\frac{\partial \log Y}{\partial \log X} = \frac{\hat{\beta} \cdot \bar{X}}{\bar{Y}}$
- **Marginal effects**: $\frac{\partial \Pr(Y=1)}{\partial X}$ in probit/logit
- **Treatment effects**: $\hat{\beta}_1/\hat{\beta}_2$ (ratio of coefficients)

All of these require knowing the distribution of a *function* of estimators.

The Delta Method: Statement

Theorem (Delta Method)

Suppose $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, V)$.

If h is continuously differentiable at θ with $h'(\theta) \neq 0$, then:

$$\sqrt{n}(h(\hat{\theta}_n) - h(\theta)) \xrightarrow{d} N\left(0, [h'(\theta)]^2 V\right)$$

In words:

- The transformation $h(\hat{\theta})$ is also asymptotically normal
- The variance gets multiplied by $[h'(\theta)]^2$
- The "spread" depends on how steep h is at θ

Intuition: Taylor Expansion

The delta method is a first-order Taylor approximation.

Near θ, any smooth function is approximately linear:

$$h(\hat{\theta}_n) \approx h(\theta) + h'(\theta)(\hat{\theta}_n - \theta)$$

Rearranging:

$$h(\hat{\theta}_n) - h(\theta) \approx h'(\theta)(\hat{\theta}_n - \theta)$$

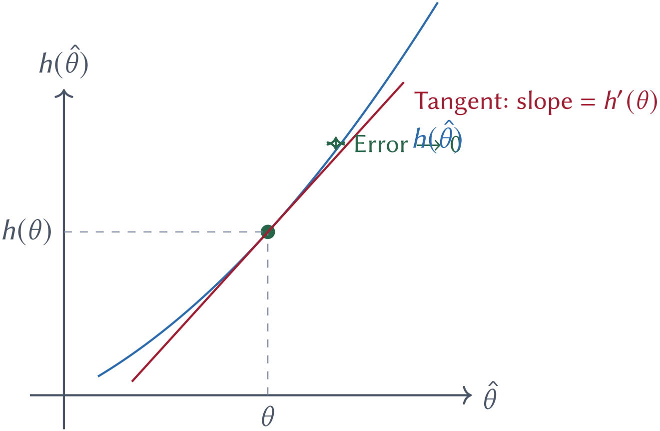
Multiply by $\sqrt{n}$:

$$\sqrt{n}(h(\hat{\theta}_n) - h(\theta)) \approx h'(\theta) \cdot \sqrt{n}(\hat{\theta}_n - \theta)$$

Since $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, V)$:

$$\sqrt{n}(h(\hat{\theta}_n) - h(\theta)) \xrightarrow{d} h'(\theta) \cdot N(0, V) = N(0, [h'(\theta)]^2 V)$$

Visual Intuition: The Tangent Line



Key insight: Near θ, the curve $\approx$ the tangent line.

The slope $h'(\theta)$ tells us how sensitive h is to changes in $\hat{\theta}$.

Why the Slope Matters

The variance of $h(\hat{\theta})$ is $[h'(\theta)]^2$ times the variance of $\hat{\theta}$.

Steep slope ($|h'(\theta)|$ large):

- Small changes in $\hat{\theta} \Longrightarrow$ large changes in $h(\hat{\theta})$
- Higher variance for $h(\hat{\theta})$
- Example: $h(\theta) = \log(\theta)$ when θ is small

Flat slope ($|h'(\theta)|$ small):

- Changes in $\hat{\theta} \Longrightarrow$ small changes in $h(\hat{\theta})$
- Lower variance for $h(\hat{\theta})$
- Example: $h(\theta) = \log(\theta)$ when θ is large

Bottom line: The transformation can amplify or dampen the variance.

Example 1: Logarithm

Setup: We want a CI for $\log(\mu)$ based on $\bar{X}_n$.

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

Apply delta method with $h(\theta) = \log(\theta)$:

- $h'(\theta) = 1/\theta$
- At $\theta = \mu$: $h'(\mu) = 1/\mu$

Result:

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

Standard error: $\text{SE}[\log(\bar{X}_n)] \approx \frac{\sigma}{\mu\sqrt{n}} = \frac{\sigma/\sqrt{n}}{\mu}$

Note: If μ is small, $1/\mu$ is large $\Rightarrow$ more variance.

Example 2: Exponential (Odds Ratios)

Setup: In logistic regression, $\hat{\beta}$ is the log-odds ratio. We want a CI for $\exp(\hat{\beta})$ (the odds ratio).

What we know: $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V)$

Apply delta method with $h(\theta) = \exp(\theta)$:

- $h'(\theta) = \exp(\theta)$
- At $\theta = \beta$: $h'(\beta) = \exp(\beta)$

Result:

$$\sqrt{n}(\exp(\hat{\beta}) - \exp(\beta)) \xrightarrow{d} N\left(0, \exp(2\beta) \cdot V\right)$$

Standard error: $\text{SE}[\exp(\hat{\beta})] \approx \exp(\beta) \cdot \text{SE}[\hat{\beta}]$

This is why software reports CIs for odds ratios using the delta method.

Example 3: Squared Mean

Setup: We want a CI for μ^2 based on $\bar{X}_n$.

Apply delta method with $h(\theta) = \theta^2$:

- $h'(\theta) = 2\theta$
- At $\theta = \mu$: $h'(\mu) = 2\mu$

Result:

$$\sqrt{n}(\bar{X}_n^2 - \mu^2) \xrightarrow{d} N(0, 4\mu^2\sigma^2)$$

Standard error: $\text{SE}[\bar{X}_n^2] \approx \frac{2|\mu|\sigma}{\sqrt{n}}$

Warning: If $\mu = 0$, then $h'(\mu) = 0$ and the delta method breaks down!
Need higher-order terms when $h'(\theta) = 0$.

The Multivariate Delta Method

What if h depends on multiple parameters?

Theorem (Multivariate Delta Method)

Suppose $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \Sigma)$ where $\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta} \in \mathbb{R}^k$.
If $h : \mathbb{R}^k \to \mathbb{R}$ is continuously differentiable, then:

$$\sqrt{n}(h(\hat{\boldsymbol{\theta}}_n) - h(\boldsymbol{\theta})) \xrightarrow{d} N\left(0, \nabla h(\boldsymbol{\theta})' \Sigma \nabla h(\boldsymbol{\theta})\right)$$

Where $\nabla h(\boldsymbol{\theta}) = \left(\frac{\partial h}{\partial \theta_1}, \ldots, \frac{\partial h}{\partial \theta_k}\right)'$ is the gradient.

The variance is $\nabla h' \Sigma \nabla h$—a quadratic form in the covariance matrix.

Multivariate Example: Ratio of Means

Setup: Two independent samples. Want a CI for μ_1/μ_2.

Let $\theta = (\mu_1, \mu_2)'$ and $h(\theta) = \mu_1/\mu_2$.

Gradient:

$$\nabla h = \begin{pmatrix} \partial h/\partial \mu_1 \\ \partial h/\partial \mu_2 \end{pmatrix} = \begin{pmatrix} 1/\mu_2 \\ -\mu_1/\mu_2^2 \end{pmatrix}$$

Covariance matrix (independent samples):

$$\Sigma = \begin{pmatrix} \sigma_1^2/n_1 & 0 \\ 0 & \sigma_2^2/n_2 \end{pmatrix}$$

Asymptotic variance:

$$\text{Var}\left(\frac{\bar{X}_1}{\bar{X}_2}\right) \approx \frac{1}{\mu_2^2} \cdot \frac{\sigma_1^2}{n_1} + \frac{\mu_1^2}{\mu_2^4} \cdot \frac{\sigma_2^2}{n_2}$$

When the Delta Method Fails

The delta method requires $h'(\theta) \neq 0$.

If $h'(\theta) = 0$:

- The first-order Taylor approximation gives zero variance
- Need to use higher-order terms (second-order delta method)
- The limiting distribution may not be normal

Example: $h(\theta) = \theta^2$ when $\theta = 0$.

- $h'(0) = 0$, so delta method gives variance 0
- Actually, $n \cdot (\bar{X}_n^2 - 0) \xrightarrow{d} \sigma^2 \chi_1^2$ (chi-squared!)

Fortunately, this edge case rarely matters in practice.

The Delta Method in Practice

You rarely compute this by hand.

Statistical software does it automatically:

- `deltamethod()` in R's `msm` package
- `nlcom` in Stata
- `car::deltaMethod()` in R
- Many packages compute it internally (e.g., for odds ratios, marginal effects)

What you need to know:

1. Recognize when delta method applies
2. Understand that variance gets scaled by $[h'(\theta)]^2$
3. Know that the result is asymptotically normal
4. Be aware of edge cases ($h'(\theta) = 0$)

Alternative: The Bootstrap

The bootstrap is a computational alternative to the delta method.

Delta method:

- Analytical: derive the formula
- Requires h to be differentiable
- Requires asymptotic normality of $\hat{\theta}$

Bootstrap:

- Computational: resample and compute $h(\hat{\theta}^*)$ many times
- No differentiability requirement
- More flexible, but slower

In practice: Use delta method when h is simple; use bootstrap for complex h or when you're unsure.

Summary: The Asymptotic Toolkit (Complete)

Tool	What It Does
LLN	$\hat{\theta}_n \xrightarrow{p} \theta$ (consistency)
CLT	$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, V)$ (asymptotic normality)
Slutsky	Plug in consistent estimators
CMT	Continuous functions preserve convergence
Delta Method	$h(\hat{\theta}_n)$ is asymptotically normal with variance $[h'(\theta)]^2 V$

This is the complete toolkit for frequentist inference.

Everything in regression will build on these foundations.

What's Next: Regression

After the midterm: Part II of the course.

Regression uses everything we've learned:

- **LLN**: $\hat{\beta} \xrightarrow{p} \beta$ (OLS is consistent)
- **CLT**: $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V)$ (asymptotic normality)
- **Slutsky**: We can estimate V and plug it in
- **Delta method**: For nonlinear functions of $\hat{\beta}$

The inference machinery is complete. Now we apply it to the workhorse of empirical social science: linear regression.

Midterm Preparation

The midterm will cover Weeks 1–8.

Key topics:

- Probability fundamentals (Bayes, conditional probability)
- Random variables, expectation, variance
- Joint distributions, covariance, the CEF
- Sampling distributions, LLN, CLT
- Estimation, confidence intervals, hypothesis testing
- Power analysis
- Convergence, Slutsky, delta method

Focus on concepts: What do these tools *mean*? When do you use them? What are the assumptions?

For the Midterm

Review:

- All lecture slides
- Problem sets 1–3
- Blackwell Chapters 2–4
- A&M Chapters 1–3

Office hours:

- Extended office hours before the midterm (see Canvas)
- Come with specific questions

Good luck! The hard part of the course is behind you. Regression is where it all pays off.