Delta Method

Statistics · Module 7 — Asymptotic Theory

Node 7.4

Position in Knowledge Graph

System: Statistics. Structural Domain: Module 7 — Asymptotic Theory. Node 7.4 — Delta Method.

The delta method transfers asymptotic normality from a parameter to any smooth function of that parameter. It depends on asymptotic normality (Node 7.3) and enables the general M-estimation framework (Node 7.5) and likelihood asymptotics (Node 7.6). It is the primary tool for constructing confidence intervals for derived quantities.

Formal Definition

Let $\hat{\theta}_n$ be asymptotically normal:

$$\sqrt{n}\,(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V(\theta)), \quad \theta \in \Theta \subseteq \mathbb{R}^k.$$

Let $g : \Theta \to \mathbb{R}^m$ be continuously differentiable at θ with Jacobian $\dot{g}(\theta) = \partial g / \partial \theta^\top \in \mathbb{R}^{m \times k}$.

First-order delta method. If $\dot{g}(\theta) \neq 0$ (has full row rank), then

$$\sqrt{n}\,\big(g(\hat{\theta}_n) - g(\theta)\big) \xrightarrow{d} \mathcal{N}\big(0,\ \dot{g}(\theta)\, V(\theta)\, \dot{g}(\theta)^\top\big)\,.$$

Second-order delta method. When $g : \mathbb{R} \to \mathbb{R}$ with $g'(\theta) = 0$ and $g''(\theta) \neq 0$:

$$n\big(g(\hat{\theta}_n) - g(\theta)\big) \xrightarrow{d} \frac{g''(\theta)}{2}\, V(\theta)\, \chi^2_1.$$

Intuition

A smooth function of an estimator inherits the estimator's asymptotic Gaussianity because, locally, a smooth function is approximately linear. The Jacobian $\dot{g}(\theta)$ acts as a magnification factor: it stretches or compresses the Gaussian fluctuations of $\hat{\theta}_n$ into Gaussian fluctuations of $g(\hat{\theta}_n)$. When the first derivative vanishes, the linear approximation is trivial and the curvature (second derivative) determines the leading-order behavior, producing a chi-squared rather than Gaussian limit.

Structural Role

The delta method extends asymptotic normality from the parameter θ to any smooth function $g(\theta)$. It is the primary tool for constructing asymptotic confidence intervals and tests for derived quantities — odds ratios, relative risks, variance components, functions of regression coefficients — and is used directly in M-estimation (Node 7.5), variance-stabilizing transformations, and the construction of asymptotic pivots (Node 4.5).

Mathematical Development

Proof of the first-order delta method.

By a first-order Taylor expansion of g around θ:

$$g(\hat{\theta}_n) = g(\theta) + \dot{g}(\theta)(\hat{\theta}_n - \theta) + R_n,$$

where the remainder $R_n = O(\|\hat{\theta}_n - \theta\|^2)$. Since $\hat{\theta}_n - \theta = O_p(n^{-1/2})$, we have $R_n = O_p(n^{-1})$, so $\sqrt{n}R_n = O_p(n^{-1/2}) = o_p(1)$.

Therefore:

$$\sqrt{n}\big(g(\hat{\theta}_n) - g(\theta)\big) = \dot{g}(\theta) \cdot \sqrt{n}(\hat{\theta}_n - \theta) + o_p(1).$$

Since $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} Z \sim \mathcal{N}(0, V(\theta))$ and $\dot{g}(\theta)$ is a fixed matrix, Slutsky's theorem gives

$$\sqrt{n}\big(g(\hat{\theta}_n) - g(\theta)\big) \xrightarrow{d} \dot{g}(\theta)Z \sim \mathcal{N}(0,\, \dot{g}(\theta)V(\theta)\dot{g}(\theta)^\top). \quad \square$$

Scalar case. For $g : \mathbb{R} \to \mathbb{R}$ and scalar θ:

$$\sqrt{n}\big(g(\hat{\theta}_n) - g(\theta)\big) \xrightarrow{d} \mathcal{N}\big(0,\ [g'(\theta)]^2 V(\theta)\big)\,.$$

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

Proof of the second-order delta method.

When $g'(\theta) = 0$, expand to second order:

$$g(\hat{\theta}_n) - g(\theta) = \frac{1}{2}g''(\theta)(\hat{\theta}_n - \theta)^2 + O_p(n^{-3/2}).$$

Multiplying by n:

$$n\big(g(\hat{\theta}_n) - g(\theta)\big) = \frac{g''(\theta)}{2}\left[\sqrt{n}(\hat{\theta}_n - \theta)\right]^2 + o_p(1).$$

Since $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V(\theta))$, the continuous mapping theorem gives

$$\left[\sqrt{n}(\hat{\theta}_n - \theta)\right]^2 \xrightarrow{d} V(\theta)\chi_1^2,$$

so $n(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} \frac{g''(\theta)}{2}V(\theta)\chi_1^2$. $\square$

Multivariate second-order delta method. For $g : \mathbb{R}^k \to \mathbb{R}$ with $\nabla g(\theta) = 0$, let $H = \nabla^2 g(\theta)$ be the Hessian. Then

$$n\big(g(\hat{\theta}_n) - g(\theta)\big) \xrightarrow{d} \frac{1}{2} Z^\top H Z, \quad Z \sim \mathcal{N}(0, V(\theta)).$$

This is a quadratic form in a Gaussian, expressible as a weighted sum of independent χ^2_1 variables with weights equal to the eigenvalues of $HV(\theta)$.

Variance-stabilizing transformations. Choose g so that the asymptotic variance of $g(\hat{\theta}_n)$ does not depend on θ. For scalar θ with asymptotic variance $V(\theta)$, solve

$$[g'(\theta)]^2 V(\theta) = c \quad \Longrightarrow \quad g(\theta) = \int \frac{d\theta}{\sqrt{V(\theta)}}.$$

This produces an asymptotically pivotal quantity: $\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} \mathcal{N}(0, c)$ with c independent of θ.

Application to the MLE. For the MLE with $V(\theta) = I(\theta)^{-1}$:

$$\sqrt{n}\big(g(\hat{\theta}_n) - g(\theta)\big) \xrightarrow{d} \mathcal{N}\big(0,\ \dot{g}(\theta)\, I(\theta)^{-1}\, \dot{g}(\theta)^\top\big)\,.$$

Worked Examples

Example 1 (Log-odds / logit). Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$ with MLE $\hat{p} = \bar{X}_n$ and $V(p) = p(1-p)$. Consider the log-odds $g(p) = \log(p/(1-p))$. Then $g'(p) = 1/(p(1-p))$, so

$$[g'(p)]^2 V(p) = \frac{1}{p^2(1-p)^2} \cdot p(1-p) = \frac{1}{p(1-p)}.$$

Thus $\sqrt{n}(\text{logit}(\hat{p}) - \text{logit}(p)) \xrightarrow{d} \mathcal{N}(0, 1/(p(1-p)))$. A confidence interval for $\text{logit}(p)$ is $\text{logit}(\hat{p}) \pm z_{\alpha/2}/\sqrt{n\hat{p}(1-\hat{p})}$.

Example 2 (Fisher z-transform). Let r be the sample correlation coefficient estimating ρ. For bivariate normal data, $\sqrt{n}(r-\rho)$ is asymptotically $\mathcal{N}(0, (1-\rho^2)^2)$. The Fisher z-transform $g(r) = \frac{1}{2}\log\frac{1+r}{1-r} = \text{arctanh}(r)$ satisfies $g'(\rho) = 1/(1-\rho^2)$, giving

$$[g'(\rho)]^2 (1-\rho^2)^2 = 1.$$

This is a variance-stabilizing transformation: $\sqrt{n}(\text{arctanh}(r) - \text{arctanh}(\rho)) \xrightarrow{d} \mathcal{N}(0, 1)$, with asymptotic variance independent of ρ. The standard approximation uses $n-3$ in place of n for finite-sample correction.

Example 3 (Ratio of means). Let (X_i, Y_i) be i.i.d. with $\mathbb{E}[X] = \mu_X$, $\mathbb{E}[Y] = \mu_Y \neq 0$, and covariance matrix Σ. The ratio $g(\mu_X, \mu_Y) = \mu_X/\mu_Y$ has gradient $\nabla g = (1/\mu_Y, -\mu_X/\mu_Y^2)^\top$. By the delta method,

$$\sqrt{n}\left(\frac{\bar{X}}{\bar{Y}} - \frac{\mu_X}{\mu_Y}\right) \xrightarrow{d} \mathcal{N}\left(0,\ \frac{\sigma_X^2}{\mu_Y^2} - \frac{2\mu_X \sigma_{XY}}{\mu_Y^3} + \frac{\mu_X^2 \sigma_Y^2}{\mu_Y^4}\right).$$

Cross-Links

AI. The delta method underlies the Laplace approximation for posterior predictive uncertainty: the variance of a prediction $f(\theta)$ is approximated by $\nabla f(\hat{\theta})^\top \hat{V} \nabla f(\hat{\theta})$, a direct application of the first-order delta method to neural network outputs.

Economics. In econometrics, the delta method is the standard tool for inference on nonlinear functions of regression coefficients, such as elasticities, marginal effects, and impulse response functions.

Linear Algebra. The delta method is the pushforward of a Gaussian measure under a smooth map: the Jacobian acts as a linear map between tangent spaces, transforming the Gaussian by the standard formula for affine transformations of normals.

Quantum Mechanics. Error propagation in quantum parameter estimation uses the same linearization: the variance of a derived quantity $g(\theta)$ under the quantum Cramer-Rao bound is $[g'(\theta)]^2/(n \cdot F_Q(\theta))$, where F_Q is the quantum Fisher information.

Structural Compression

Invariant. Smooth functions of asymptotically normal estimators are asymptotically normal, with variance determined by the Jacobian of the transformation.

Mechanism. First-order Taylor linearization reduces $g(\hat{\theta}_n) - g(\theta)$ to a linear function of $\hat{\theta}_n - \theta$; the linear map transforms the Gaussian limit by the Jacobian. When the Jacobian vanishes, the second-order term dominates, producing a chi-squared limit.

Cross-System Echo. The delta method is the statistical instance of error propagation in measurement theory. In differential geometry, it is the pushforward of a distribution under a smooth map. The variance-stabilizing transformation is the unique reparametrization that makes the Fisher information metric flat.

Exercises

1. Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$. Use the delta method to find the asymptotic distribution of $g(\bar{X}_n) = 1/\bar{X}_n$ (the MLE of λ).
2. For the Poisson MLE $\hat{\lambda} = \bar{X}_n$, apply the delta method to find the asymptotic distribution of $\sqrt{\hat{\lambda}}$. Verify that $g(\lambda) = \sqrt{\lambda}$ is a variance-stabilizing transformation.
3. Prove the multivariate delta method: if $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V)$ in $\mathbb{R}^k$ and $g : \mathbb{R}^k \to \mathbb{R}^m$ is differentiable, derive the asymptotic distribution of $g(\hat{\theta}_n)$.
4. Let $\hat{p}_1$ and $\hat{p}_2$ be independent sample proportions from n_1 and n_2 observations. Use the delta method to derive the asymptotic distribution of the log-relative-risk $\log(\hat{p}_1/\hat{p}_2)$.
5. Apply the second-order delta method to $g(\theta) = (\theta - \mu)^2$ with $\hat{\theta}_n = \bar{X}_n$ and $\theta = \mu$, and identify the limiting distribution.
6. Find the variance-stabilizing transformation for the binomial proportion $\hat{p}$ (with $V(p) = p(1-p)$), and show it is $g(p) = \arcsin(\sqrt{p})$.
7. Argue, structurally, why the delta method breaks down for non-differentiable functions g, giving a specific example where $g(\hat{\theta}_n)$ is asymptotically normal but with a different variance formula, and explaining what replaces the Jacobian.

Statistics · Module 7 — Asymptotic Theory · Node 7.4
Concepts: central-limit-theorem, convergence
Prerequisites: 7.3
Enables: 7.5, 7.6

hbar.university — own.your.cognition