The Delta Method: A Complete Proof

Statistical Inference – Session 2
Based on Casella and Berger, Theorem 5.5.24

1. Statement of the theorem

Theorem 1 (Delta Method). *Let Y_n be a sequence of random variables such that*

$$\sqrt{n}(Y_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2).$$

Suppose that g is differentiable at the point θ, with $g'(\theta) \neq 0$. Then

$$\sqrt{n}\big[g(Y_n) - g(\theta)\big] \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sigma^2[g'(\theta)]^2\big)\,.$$

The proof in Casella and Berger uses Taylor expansion and Slutsky's theorem. The key step that is usually abbreviated is the treatment of the Taylor remainder. We make that step explicit.

2. The main idea

The delta method is simply a first-order local linearization. Around θ,

$$g(y) \approx g(\theta) + g'(\theta)(y - \theta).$$

Putting $y = Y_n$ and multiplying by $\sqrt{n}$ suggests

$$\sqrt{n}\,[g(Y_n) - g(\theta)] \approx g'(\theta)\sqrt{n}\,(Y_n - \theta).$$

Since the expression on the right has a known limiting distribution, the result follows if we can prove that the approximation error is asymptotically negligible.

3. A rigorous treatment of the remainder

Write

$$g(Y_n) - g(\theta) = g'(\theta)(Y_n - \theta) + R_n,$$

where

$$R_n := g(Y_n) - g(\theta) - g'(\theta)(Y_n - \theta).$$

The required result will follow once we establish

$$\sqrt{n}R_n \xrightarrow{p} 0.$$

Because g is differentiable at θ, by the definition of the derivative,

$$\lim_{y \to \theta} \frac{g(y) - g(\theta) - g'(\theta)(y - \theta)}{y - \theta} = 0.$$

Define

$$h(y) = \begin{cases} \dfrac{g(y) - g(\theta) - g'(\theta)(y-\theta)}{y-\theta}, & y \neq \theta, \\ 0, & y = \theta. \end{cases}$$

Then $h(y) \to 0$ as $y \to \theta$. Moreover,

$$R_n = (Y_n - \theta)h(Y_n).$$

Thus the entire problem reduces to showing that $h(Y_n) \xrightarrow{p} 0$.

Why $Y_n \xrightarrow{p} \theta$

We are given

$$\sqrt{n}(Y_n - \theta) \xrightarrow{\mathcal{D}} Z, \qquad Z \sim \mathcal{N}(0, \sigma^2).$$

Convergence in distribution implies tightness. Hence $\{\sqrt{n}(Y_n - \theta)\}$ is bounded in probability. Since $1/\sqrt{n} \to 0$,

$$Y_n - \theta = \frac{1}{\sqrt{n}} \sqrt{n}(Y_n - \theta) \xrightarrow{p} 0.$$

Therefore

$$Y_n \xrightarrow{p} \theta.$$

Why $h(Y_n) \xrightarrow{p} 0$

We now use the following elementary consequence of convergence in probability: if $Y_n \xrightarrow{p} \theta$ and h is continuous at θ, then $h(Y_n) \xrightarrow{p} h(\theta)$. In our construction, $h(y) \to 0$ as $y \to \theta$, so h is continuous at θ and $h(\theta) = 0$. Hence

$$h(Y_n) \xrightarrow{p} 0.$$

For completeness, this can be verified directly. Let $\varepsilon > 0$. Since $h(y) \to 0$ as $y \to \theta$, there exists $\delta > 0$ such that

$$|y - \theta| < \delta \quad \Longrightarrow \quad |h(y)| < \varepsilon.$$

Therefore

$$\Pr\{|h(Y_n)| > \varepsilon\} \leq \Pr\{|Y_n - \theta| \geq \delta\} \longrightarrow 0.$$

Thus $h(Y_n) \xrightarrow{p} 0$.

4. The scaled remainder is negligible

We have shown

$$\sqrt{n}R_n = h(Y_n)\sqrt{n}(Y_n - \theta).$$

The first factor satisfies

$$h(Y_n) \xrightarrow{p} 0,$$

while the second factor satisfies

$$\sqrt{n}(Y_n - \theta) \xrightarrow{\mathcal{D}} Z, \qquad Z \sim \mathcal{N}(0, \sigma^2).$$

Convergence in distribution implies boundedness in probability, so the second factor is $O_p(1)$. Hence

$$h(Y_n)\sqrt{n}(Y_n - \theta) = o_p(1)O_p(1) = o_p(1).$$

Consequently,

$$\boxed{\sqrt{n}R_n \xrightarrow{p} 0.}$$

This is the step that is compressed into the phrase "the remainder $\to 0$ as $Y_n \to \theta$" in many textbook proofs. Notice that it is not enough merely to say that $R_n \to 0$: after multiplication by $\sqrt{n}$, we need the stronger statement $\sqrt{n}R_n \xrightarrow{p} 0$.

5. Finish the proof using Slutsky's theorem

From the Taylor decomposition,

$$\sqrt{n}[g(Y_n) - g(\theta)] = g'(\theta)\sqrt{n}(Y_n - \theta) + \sqrt{n}R_n.$$

We have established

$$g'(\theta)\sqrt{n}(Y_n - \theta) \xrightarrow{\mathcal{D}} g'(\theta)Z,$$

where $Z \sim \mathcal{N}(0, \sigma^2)$, and

$$\sqrt{n}R_n \xrightarrow{p} 0.$$

By Slutsky's theorem,

$$\sqrt{n}[g(Y_n) - g(\theta)] \xrightarrow{\mathcal{D}} g'(\theta)Z.$$

Since a constant multiple of a normal random variable is normal,

$$g'(\theta)Z \sim \mathcal{N}\left(0, \sigma^2[g'(\theta)]^2\right).$$

Therefore,

$$\boxed{\sqrt{n}[g(Y_n) - g(\theta)] \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2[g'(\theta)]^2\right).}$$

$\square$

6. Equivalent o_p formulation

A compact way to remember the proof is to write the first-order expansion as

$$g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + o_p(Y_n - \theta).$$

Because

$$Y_n - \theta = O_p(n^{-1/2}),$$

we obtain

$$g(Y_n) - g(\theta) = g'(\theta)(Y_n - \theta) + o_p(n^{-1/2}).$$

Multiplying by $\sqrt{n}$ gives

$$\sqrt{n}[g(Y_n) - g(\theta)] = g'(\theta)\sqrt{n}(Y_n - \theta) + o_p(1),$$

which immediately yields the result by Slutsky.

7. Where each assumption is used

Assumption	Role in the proof
$\sqrt{n}(Y_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$	Gives the limiting distribution and implies $Y_n \xrightarrow{p} \theta$.
g differentiable at θ	Gives the first-order approximation and ensures the remainder is $o(\lvert Y_n - \theta \rvert)$.
$g'(\theta) \neq 0$	Ensures the stated limiting variance is nonzero. The theorem can be extended to $g'(\theta) = 0$, but the first-order limit then degenerates at zero and a second-order delta method may be needed.
Slutsky's theorem	Converts the asymptotically negligible remainder into the desired limiting distribution.

8. Application: $Y_n = \bar{X}$

Suppose

$$X_1, \ldots, X_n \overset{iid}{\sim} (\mu, \sigma^2),$$

with finite variance. By the CLT,

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2).$$

Taking $g(x) = \log x$, for $\mu > 0$,

$$g'(\mu) = \frac{1}{\mu}.$$

The delta method therefore gives

$$\boxed{\sqrt{n}[\log(\bar{X}) - \log \mu] \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\sigma^2}{\mu^2}\right).}$$

Equivalently,

$$\log(\bar{X}) \approx \mathcal{N}\left(\log \mu, \frac{\sigma^2}{n\mu^2}\right)$$

for large n.

9. Teaching takeaway

The delta method is not a mysterious new limit theorem. Its proof consists of three ingredients:

$$\boxed{\begin{array}{l} \text{(1) CLT or another limiting result for } Y_n, \\ \text{(2) first-order Taylor expansion of } g, \\ \text{(3) Slutsky to discard the scaled remainder.} \end{array}}$$

The technically important statement is

$$\boxed{\sqrt{n}R_n = o_p(1),}$$

not merely $R_n \to 0$.