The Delta Method

The delta method is an important theorem in asymptotic statistics, which allows us to get the asymptotic distribution of a transformation of a sequence of random variables that itself converge in distribution. That is, suppose that we have a sequence of real numbers $(r_n)_n$ such that $r_n \to \infty$, and a sequence of random variables, $\{X, X_n, n \geq 1\}$ such that

$$r_n(X_n - \theta) \xrightarrow{d} X,$$

for some real number θ. If $\phi : \mathbb{R} \to \mathbb{R}$ is a sufficiently well behaved function, then when can we also conclude that

$$r_n(\phi(X_n) - \phi(\theta))$$

converges in distribution and if so to what?

Before, we can answer this question using the delta method, we need some preliminary results and definitions.

Definition 1. Let $\{X, X_n, n \geq 1\}$ be random variables with distribution functions $\{F, F_n, n \geq 1\}$, respectively. Then

a) X_n is said to **converge in probability** to X, written $X_n \xrightarrow{P} X$ if

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0, \ \forall \varepsilon > 0.$$

b) X_n is said to converge in distribution to a random variable X, written $X_n \xrightarrow{d} X$ if

$$F_n(t) \to F(t)$$

for all points $t \in \mathbb{R}$ such that F is continuous at t.

c) $\{X_n, n \geq 1\}$ is said to be **bounded in probability** or **uniformly tight** if $\forall \varepsilon > 0$, $\exists M > 0$ such that

$$\sup_{n\geq 1} P(|X_n| > M) < \varepsilon.$$

Note 1. The following implications always hold

$$X_n \xrightarrow{P} X \implies X_n \xrightarrow{d} X \implies \{X_n, n \geq 1\} \text{ is bounded in probability.}$$

We will also make use of the following, which is part of Slutsky's lemma.

Lemma 1. *Suppose that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$ for some constant $c \in \mathbb{R}$. Then $Y_n X_n \xrightarrow{d} cX$ and $X_n + Y_n \xrightarrow{d} X + c$.*

Note 2. It turns out that $Y_n \xrightarrow{d} c$ if and only if $Y_n \xrightarrow{P} c$, where $c \in \mathbb{R}$ is a constant.

We will frequently work with sequences of random variables that are bounded in probability or that converge to 0 in probability, so it is useful to have some short hand notation to represent such sequences.

Definition 2 (Stochastic o and O Notation)**.** Let $\{X_n, n \geq 1\}$ be a sequences of random variables.

a) By $X_n = o_P(1)$, we mean $X_n \xrightarrow{P} 0$.

b) By $X_n = O_p(1)$, we mean that X_n is bounded in probability.

c) For a sequence of random variables $\{R_n, n \geq 1\}$, we write $X_n = o_P(R_n)$ to mean that $X_n = Y_n R_n$, where Y_n is some sequence of random variables such that $Y_n = o_P(1)$. Similarly, we write $X_n = O_P(R_n)$ to mean that $X_n = Y_n R_n$, where Y_n is some sequence of random variables such that $Y_n = O_P(1)$.

We will need the following stochastic relation to prove the delta method.

Proposition 1. $o_p(O_p(1)) = o_p(1)$. *That is, if* $X_n \xrightarrow{P} 0$ *and* $\{Y_n, n \geq 1\}$ *are bounded in probability, then* $X_n Y_n \xrightarrow{P} 0$.

Proof. Let $\varepsilon > 0$. Let $\delta > 0$ and choose $M > 0$ such that

$$\sup_{n \geq 1} P(|Y_n| \geq M) < \delta.$$

Then

$$\begin{aligned} P(|Y_n X_n| > \varepsilon) &= P(|Y_n X_n| > \varepsilon, |Y_n| \geq M) + P(|Y_n X_n| > \varepsilon, |Y_n| < M) \\ &\leq P(|Y_n| \geq M) + P(|M||X_n| > \varepsilon) \\ &< \delta + P\left(|X_n| > \frac{\varepsilon}{M}\right) \\ &\to \delta. \end{aligned}$$

Since $\delta > 0$ and $\varepsilon > 0$ were arbitrary, we have

$$P(|Y_n X_n| > \varepsilon) \to 0, \forall \varepsilon > 0.$$

□

We write that a function $R(h) = o(|h|^p)$ as $h \to 0$ if

$$\lim_{h \to 0} \frac{R(h)}{|h|^p} = 0.$$

The final result we will need pertains to handling remainder terms in our first order Taylor expansion.

Lemma 2. *Let* $R : D \subset \mathbb{R} \mapsto \mathbb{R}$ *be a function such that* $R(0) = 0$ *and the support of the random variables* $\{X_n, n \geq 1\}$ *lies within* D *such that* $X_n \xrightarrow{P} 0$. *Then for every* $p > 0$, *if* $R(h) = o(|h|^p)$ *as* $h \to 0$, *then* $R(X_n) = o_P(|X_n|^p)$.

Proof. Define $g : D \mapsto \mathbb{R}$ as

$$g(h) = \begin{cases} \dfrac{R(h)}{|h|^p}, & h \neq 0 \\ 0, & h = 0 \end{cases}.$$

Then g is continuous at 0 and $R(h) = g(h)|h|^p$ for all $h \in D$, so $R(X_n) = g(X_n)|X_n|^p$. Because g is continuous at 0, $g(X_n) \xrightarrow{P} g(0) = 0$ by the continuous mapping theorem. Thus, $g(X_n) = o_P(1)$, so

$$R(X_n) = g(X_n)|X_n|^p = |X_n|^p o_P(1) = o_P(|X_n|^p).$$

□

Theorem 1 (Delta Method)**.** *Suppose that $\phi : D \subset \mathbb{R} \mapsto \mathbb{R}$ is differentiable at θ. Let $\{X_n, n \geq 1\}$ be random variables who support lies in D. If $r_n(X_n - \theta) \xrightarrow{d} X$ for some random variable X and some sequence $r_n \to \infty$, then $r_n(\phi(X_n) - \phi(\theta)) \xrightarrow{d} \phi'(\theta)X$.*

Proof. Since $r_n(X_n - \theta)$ converges in distribution to X, it is bounded in probability. Moreover, $r_n \to \infty \implies 1/r_n \to 0$. Hence

$$\frac{1}{r_n} = o_P(1) \text{ and } r_n(X_n - \theta) = O_P(1) \implies X_n - \theta = \frac{1}{r_n} \cdot r_n(X_n - \theta) = o_P(1)O_P(1) = o_P(1).$$

ϕ differentiable at θ implies that

$$R(h) = \phi(\theta + h) - \phi(\theta) - \phi'(\theta)h = o(|h|) \text{ as } h \to 0.$$

Thus,

$$R(X_n - \theta) = \phi(X_n) - \phi(\theta) - \phi'(\theta)(X_n - \theta) = o_P(|X_n - \theta|),$$

by Lemma 2. Mutiplying both sides by r_n, we get

$$r_n(\phi(X_n) - \phi(\theta)) - \phi'(\theta)[r_n(X_n - \theta)] = o_P(|r_n(X_n - \theta)|).$$

By the continuous mapping theorem

$$r_n(X_n - \theta) \xrightarrow{d} X \implies |r_n(X_n - \theta)| \xrightarrow{d} |X| \implies |r_n(X_n - \theta)| = O_P(1).$$

Hence,

$$r_n(\phi(X_n) - \phi(\theta)) - \phi'(\theta)[r_n(X_n - \theta)] = o_P(|r_n(X_n - \theta)|) = o_P(1).$$

This implies that

$$r_n(\phi(X_n) - \phi(\theta)) = \underbrace{r_n(\phi(X_n) - \phi(\theta)) - \phi'(\theta)[r_n(X_n - \theta)]}_{o_P(1)} + \underbrace{\phi'(\theta)[r_n(X_n - \theta)]}_{\xrightarrow{d} \phi'(\theta)X} \xrightarrow{d} \phi'(\theta)X,$$

by Slutsky's lemma. □