Lecture 18: Delta Method

Mathematical Statistics I, MATH 60061/70061

Thursday November 18, 2021

Reference: Casella & Berger, 5.5.4

Delta Method

Suppose X_n is a sequence of random variables that satisfy

$$\sqrt{n}(X_n - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

as $n \to \infty$. For a given function g, suppose that $g'(\theta)$ exists and $g'(\theta) \neq 0$. Then

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{d} \mathcal{N}(0, [g'(\theta)]^2\sigma^2),$$

as $n \to \infty$.

In other words, the distribution of $g(X_n)$ can be approximated by

$$\mathcal{N}\left(g(\theta), \frac{[g'(\theta)]^2\sigma^2}{n}\right)$$

for large n.

Multivariate extensions

All convergence concepts can be extended to handle sequences of random variables.

Central Limit Theorem: Suppose $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$, is a sequence of iid random vectors (of dimension k) with $E(\boldsymbol{X}_1) = \boldsymbol{\mu}_{k\times 1}$ and $\mathrm{Cov}(\boldsymbol{X}_1) = \boldsymbol{\Sigma}_{k\times k}$. Let $\bar{\boldsymbol{X}}_n = (\bar{X}_{1+}, \bar{X}_{2+}, \ldots, \bar{X}_{k+})'$ denote the vector of sample means. Then $\sqrt{n}(\bar{\boldsymbol{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathrm{MVN}_k(\mathbf{0}, \boldsymbol{\Sigma})$.

Multivariate Delta Method: Suppose $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$, is a sequence of iid random vectors (of dimension k) that satisfy $\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathrm{MVN}_k(\mathbf{0}, \boldsymbol{\Sigma})$. For a given function $g : \mathbb{R}^k \to \mathbb{R}$, suppose that g is differentiable at $\boldsymbol{\mu}$ and is not zero. Then

$$\sqrt{n}[g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})] \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{x}} \boldsymbol{\Sigma} \frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{x}'}\right)$$

where

$$\frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{x}} = \left(\frac{\partial g(\boldsymbol{x})}{\partial x_1}, \ldots, \frac{\partial g(\boldsymbol{x})}{\partial x_k}\right)\Big|_{\boldsymbol{x}=\boldsymbol{\mu}}.$$

Ratio of sample means

Suppose $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ is a sample of iid random vectors of dimension 2. Define

$$\bar{X}_{1+} = \frac{1}{n} \sum_{j=1}^{n} X_{1j} \quad \text{and} \quad \bar{X}_{2+} = \frac{1}{n} \sum_{j=1}^{n} X_{2j}$$

and denoted by

$$\bar{\boldsymbol{X}}_n = \begin{pmatrix} \bar{X}_{1+} \\ \bar{X}_{2+} \end{pmatrix},$$

the vector of sample means. The multivariate CLT says that

$$\sqrt{n}(\bar{\boldsymbol{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathrm{MVN}_2(\mathbf{0}, \boldsymbol{\Sigma})$$

as $n \to \infty$.

A common summary is the ratio of sample means

$$R = g(\bar{\boldsymbol{X}}_n) = \frac{\bar{X}_{1+}}{\bar{X}_{2+}}.$$

The exact distribution of R is a Cauchy, which is difficult to work with.

We may resort to the large sample distribution of R. With $g(x_1, x_2) = x_1/x_2$, we have

$$\frac{\partial g(x_1, x_2)}{\partial x_1} = \frac{1}{x_2} \quad \text{and} \quad \frac{\partial g(x_1, x_2)}{\partial x_2} = -\frac{x_1}{x_2^2}$$

so that

$$\frac{\partial g(\boldsymbol{\mu})}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{1}{\mu_2} & -\frac{\mu_1}{\mu_2^2} \end{pmatrix}.$$

The multivariate Delta Method says that

$$\sqrt{n}[g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})] = \sqrt{n}\left(R - \frac{\mu_1}{\mu_2}\right) \xrightarrow{d} \mathcal{N}(0, \sigma_R^2),$$

as $n \to \infty$, where

$$\sigma_R^2 = \begin{pmatrix} \frac{1}{\mu_2} & -\frac{\mu_1}{\mu_2^2} \end{pmatrix} \boldsymbol{\Sigma} \begin{pmatrix} \frac{1}{\mu_2} \\ -\frac{\mu_1}{\mu_2^2} \end{pmatrix}$$

Let $X_{(1)}, \ldots, X_{(n)}$ be the order statistics of a random sample $X_1, \ldots, X_n$ from $\mathrm{Unif}(0,1)$. Find the distribution of $X_1/X_{(1)}$.

For $s > 1$,

$$
\begin{aligned}
P\left(\frac{X_1}{X_{(1)}} > s\right) &= \sum_{i=1}^{n} P\left(\frac{X_1}{X_{(1)}} > s, X_{(1)} = X_i\right) \\
&= \sum_{i=2}^{n} P\left(\frac{X_1}{X_{(1)}} > s, X_{(1)} = X_i\right) \\
&= (n-1) P\left(\frac{X_1}{X_{(1)}} > s, X_{(1)} = X_n\right) \\
&= (n-1) P\left(X_1 > sX_n, X_2 > X_n, ..., X_{n-1} > X_n\right) \\
&= (n-1) P\left(sX_n < 1, X_1 > sX_n, X_2 > X_n, ..., X_{n-1} > X\right. \\
&= (n-1) \int_0^{1/s} \left[\int_{sx_n}^{1} \left(\prod_{i=2}^{n-1} \int_{x_n}^{1} dx_i\right) dx_1\right] dx_n \\
&= (n-1) \int_0^{1/s} (1-x_n)^{n-2} (1-sx_n) dx_n
\end{aligned}
$$

Thus, for $s > 1$, with Leibniz integral rule

$$\frac{d}{ds}P\left(\frac{X_1}{X_{(1)}} \leq s\right) = \frac{d}{ds}\left[1 - (n-1)\int_0^{1/s}(1-t)^{n-2}(1-st)dt\right]$$
$$= \int_0^{1/s}(n-1)(1-t)^{n-2}tdt.$$

With integration by parts ($u = t$, $v = -(1-t)^{n-1}$),

$$\int_0^{1/s}(n-1)(1-t)^{n-2}tdt = t\cdot\left[-(1-t)^{n-1}\right]\Big|_0^{1/s} - \int_0^{1/s} -(1-t)^{n-1}dt$$
$$= -\left(\frac{1}{s}\right)\left(1-\frac{1}{s}\right)^{n-1} + \left(-\frac{1}{n}(1-t)^n\right)\Big|_0^{1/s}$$
$$= -\left(\frac{1}{s}\right)\left(1-\frac{1}{s}\right)^{n-1} - \frac{1}{n}\left[\left(1-\frac{1}{s}\right)^n - 1\right]$$
$$= \frac{1}{n}\left[1 - \left(1-\frac{1}{s}\right)^{n-1}\left(1-\frac{1}{s}+\frac{n}{s}\right)\right].$$

For $s \leq 1$, obviously

$$P\left(\frac{X_1}{X_{(1)}} \leq s\right) = 0 \quad \frac{d}{ds} P\left(\frac{X_1}{X_{(1)}} \leq s\right) = 0$$