LECTURE 6: LARGE SAMPLE THEORY

MECO 7312.
INSTRUCTOR: DR. KHAI CHIONG
OCTOBER 15, 2025

1. Continuous Mapping Theorem

Suppose that the sequence of random variable Y_n converges in probability to θ as $n \to \infty$. Then continuous functions of Y_n also converge to functions of θ. That is,

$Y_n \xrightarrow{p} \theta$. If g is a continuous function, then $g(Y_n) \xrightarrow{p} g(\theta)$.

$Y_n \xrightarrow{a.s} \theta$. If g is a continuous function, then $g(Y_n) \xrightarrow{a.s} g(\theta)$.

Suppose that the sequence of random variable Y_n converges in distribution to Y as $n \to \infty$. Then continuous functions of Y_n also converge to functions of Y. That is,

$Y_n \xrightarrow{d} Y$. If g is a continuous function, then $g(Y_n) \xrightarrow{d} g(Y)$.

1.1. Example: sample standard deviation

Previously we saw that the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ converges in probability to $\sigma^2 \equiv \text{Var}(X_i)$. Let $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2}$ be the sample standard deviation. It follows from the continuous mapping theorem that s converges in probability to σ because $\sqrt{S^2} \xrightarrow{p} \sqrt{\sigma^2}$.

Although the sample standard deviation S is a consistent estimator of σ, it is a biased estimator of σ.

From Jensen's inequality, if g is a convex function, then

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$$
$$\mathbb{E}[-g(X)] \leq -g(\mathbb{E}[X])$$

If g is a convex function, then $-g$ is a concave function. For a strictly concave function g, we have $\mathbb{E}[g(X)] < g(\mathbb{E}[X])$. Since $f(x) = \sqrt{x}$ is a concave function, and $\mathbb{E}[S^2] = \sigma^2$, it follows that

$$\mathbb{E}[\sqrt{S^2}] < \sqrt{\mathbb{E}[S^2]}$$
$$\mathbb{E}[S] < \sqrt{\sigma^2} = \sigma$$

Therefore, the sample standard deviation is a biased estimator of the true standard deviation (it underestimates).

2. Central Limit Theorem

Let $X_1, X_2, \ldots$ be a sequence of i.i.d random variables with $\mathbb{E}[X_i] = \mu$ and $0 < \text{Var}(X_i) = \sigma^2 < \infty$. Define $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. The Law of Large Numbers tells us that $\bar{X}$ converges in probability to μ.[1] That is, $\bar{X} - \mu \to^p 0$

Now consider $\sqrt{n}(\bar{X} - \mu)$. As $n \to \infty$, we have two conflicting convergence: (i) $\bar{X} - \mu \to 0$ in probability, (ii) but $\sqrt{n} \to \infty$. Somehow, they balance each other out in the sense that $\sqrt{n}(\bar{X} - \mu)$ converges to a random variable as $n \to \infty$. This random variable is $\mathcal{N}(0, \sigma^2)$, regardless of what the underlying distribution of X is.

Central Limit Theorem (Lindeberg-Levy): $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ converges in distribution to $\mathcal{N}(0,1)$ as $n \to \infty$. That is, $\lim_{n\to\infty} P(\sqrt{n}(\bar{X}_n - \mu)/\sigma \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$ for all $x \in \mathbb{R}$. Equivalently, $\sqrt{n}(\bar{X}_n - \mu)$ converges in distribution to $\mathcal{N}(0, \sigma^2)$ as $n \to \infty$.

The sample mean is root-n consistent: $(\bar{X}_n - \mu)/\sigma$ decays to zero at rate $\frac{1}{\sqrt{n}}$ (in probability) as $n \to \infty$. Equivalently, scaling by $\sqrt{n}$ yields a non-degenerate limit. Under the Lindeberg–Lévy CLT, the only assumptions are i.i.d. with finite variance. Some texts (e.g., Casella–Berger) assume the moment generating function exists in a neighborhood of 0 to give an mgf-based proof; this is a proof technique, not a requirement of the theorem itself.

2.1. Asymptotic approximation

When the underlying data-generating process is Normal, we know that the sample mean $\bar{X}_n$ is distributed according to $\mathcal{N}(\mu, \frac{\sigma^2}{n})$.

What if the data-generating process is not Normally distributed. For example, if X_i is Uniformly distributed, what is the distribution of the sample mean $\bar{X}_n$? In practice, we do not know the data-generating process, which is why CLT is important.

[1] Which also implies that $\bar{X}$ converges in distribution to the (degenerate) distribution μ (a constant).

We can use *Asymptotic Approximation* to approximately derive the distribution of $\bar{X}_n$. Starting with the result of the CLT:

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

Rearranging, $\bar{X}$ is approximately distributed as $\mathcal{N}(\mu, \frac{\sigma^2}{n})$, when n is very large. The goal of asymptotic approximations is to appeal to asymptotically large n in order to infer the distribution of a statistic.

Even when n is finite and not large, we can usually take $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ to approximate the distribution of $\bar{X}$. We can use simulations to see that this approximation holds remarkably well in many cases.

2.2. Simulating the Central Limit Theorem

Take X_i to be exponentially distributed, i.e. the pdf of X_i is $f(x) = \lambda e^{-\lambda x}$.

According to the CLT, $\sqrt{n}(\bar{X} - \frac{1}{\lambda}) \to_d \mathcal{N}(0, \frac{1}{\lambda^2})$, where $\mathbb{E}[X] = \frac{1}{\lambda}$ and $\text{Var}(X) = \frac{1}{\lambda^2}$. Therefore the asymptotic approximation for the distribution of $\bar{X}$ is $\bar{X} \sim \mathcal{N}(\frac{1}{\lambda}, \frac{1}{n\lambda^2})$.

We can see from the monte carlo simulation that even when the sample size is not too large ($n = 100$), the asymptotic approximation from the CLT is remarkably accurate. Now if we repeat the above with a smaller sample size, $n = 10$, then we see that the CLT breaks down (especially for skewed/heavy-tailed data). We can repeat the above simulation with other data-generating process.

2.3. Berry–Esseen inequality (how close is CLT at finite n?)

Let $X_1, \ldots, X_n$ be i.i.d. with $\mathbb{E}[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2 \in (0, \infty)$, and finite third absolute central moment

$$\rho \;\equiv\; \mathbb{E}\big[\, |X_i - \mu|^3 \,\big] \;<\; \infty.$$

Then there exists a universal constant C (one may take $C \leq 0.56$; sharper values are known) such that

$$\sup_{x \in \mathbb{R}} \left| P\left(\frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma} \leq x \right) \;-\; \Phi(x) \right| \;\leq\; \frac{C\,\rho}{\sigma^3 \sqrt{n}}.$$

Thus the CLT error decays at the rate $O(n^{-1/2})$, with the constant governed by the standardized third absolute moment ρ/σ^3 (often called the "Lyapunov ratio").

Standardize X to $Z = (X - \mu)/\sigma$. Then

$$\frac{\rho}{\sigma^3} = \frac{\mathbb{E}|X-\mu|^3}{\sigma^3} = \mathbb{E}|Z|^3.$$

Thus the Lyapunov ratio, ρ/σ^3, is the *third absolute standardized moment*—a dimensionless index of *tail heaviness*. Larger values indicate heavier tails and looser Berry–Esseen bounds, leading to a slower normal approximation (a larger sample size is needed to achieve a worst-case tolerable error). Recall that $\mathbb{E}[Z^3]$ is a measure of skewness. The Lyapunov ratio is always larger than the absolute value of skewness, $\mathbb{E}|Z|^3 \ \geq\ \left|\mathbb{E}[Z^3]\right|$.

The Berry–Esseen inequality implies that to guarantee a uniform CLT error $\leq \varepsilon$ it suffices to take

$$n \ \geq\ \left(\frac{C}{\varepsilon}\cdot\frac{\rho}{\sigma^3}\right)^2 \ =\ \left(\frac{C}{\varepsilon}\cdot\mathbb{E}|Z|^3\right)^2.$$

Although this is conservative (worst-case, uniform in x), it cleanly shows how tail weight control the needed sample size.

Distribution	$\mathbb{E}\|Z\|^3$	Required n for $\varepsilon = 0.10,\ C = 0.56$
Bernoulli(1/2)	1	32
Uniform$(0,1)$	$\frac{3\sqrt{3}}{4}$ (≈ 1.299)	53
Exponential (any rate λ)	$\frac{12}{e} - 2$ (≈ 2.415)	183

3. Slutsky's theorem

Suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$ for a constant $a \in \mathbb{R}$. Then

$$X_n + Y_n \xrightarrow{d} X + a, \qquad Y_n X_n \xrightarrow{d} aX, \qquad \frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{a} \ \text{ if } a \neq 0.$$

More generally, for any function $g : \mathbb{R}^2 \to \mathbb{R}$ that is continuous at every (x, a) with x in the support of X,

$$g(X_n, Y_n) \xrightarrow{d} g(X, a).$$

The Slutsky's theorem can be used to show that the biased sample variance $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ is nevertheless a consistent estimator of $\sigma^2 \equiv \mathrm{Var}(X_i)$.

$$S^2 \xrightarrow{p} \sigma^2$$

$$\frac{n-1}{n} S^2 \xrightarrow{p} \sigma^2 \ \text{, as } n \to \infty$$

From CLT, we know that $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} \mathcal{N}(0,1)$. What is the limiting distribution if we replace σ by the sample standard deviation S_n. We have seen previously that $S_n^2 \xrightarrow{p} \sigma^2$, therefore $S_n \xrightarrow{p} \sigma$ by the Continuous Mapping Theorem. By applying Slutsky's Theorem to $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ and $S_n \xrightarrow{p} \sigma$,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{d} \mathcal{N}(0,1)$$

Hence, *for large n, the distribution of $\bar{X}$ is approximately $\mathcal{N}(\mu, \frac{S^2}{n})$.*[2] Now recall that if X_i are Normal, then $\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}$ exactly, and as $n \to \infty$, $t_{n-1} \xrightarrow{d} N(0,1)$.

Using Slutsky's theorem, we can also show that:

$$n^{1/3}(\bar{X}_n - \mu)/\sigma = n^{-1/6} n^{1/2}(\bar{X}_n - \mu)/\sigma \to 0$$

4. Delta method

We have derived the asymptotic distribution of the sample mean, that is, $\bar{X} \approx \mathcal{N}(\mu, \frac{\sigma^2}{n})$. What about the sample variance, other statistics and estimator?

Let $X_1, \ldots, X_n$ be iid from a distribution. Suppose we are interested in $g(\bar{X})$. The Taylor's series of g at a is:

$$g(x) = g(a) + g'(a)(x-a) + R(x,a) \tag{1}$$

$R(x,a)$ is the remainder term. The remainder term will be small compared to $g(a) + g'(a)(x-a)$ when x is close to a, and can be ignored. That is, $\lim_{x\to a} R(x,a)/(x-a) = 0$. As a shorthand, we usually write $g(x) = g(a) + g'(a)(x-a) + o(x-a)$, where $o(x-a)$ is a term that is dominated by $x - a$ in the limit.

If we substitute x with $\bar{X}$ and a with $\mu \equiv \mathbb{E}[X_i]$,

$$g(\bar{X}) = g(\mu) + g'(\mu)(\bar{X} - \mu) + o(\bar{X} - \mu) \tag{2}$$

In the limit as $n \to \infty$, we can show that $\sqrt{n} \cdot o(\bar{X} - \mu) \to 0$. Therefore for large n, we have:

[2] However we still do not know what μ is, so how can this result be useful? Well, in the framework of Hypothesis Testing which we will talk about later, if we conjecture that $\mu = \mu_0$, then we would know the entire sampling distribution of $\bar{X}$, and see whether our realized sample mean is consistent with that sampling distribution.

$$\sqrt{n}(g(\bar{X}) - g(\mu)) \approx g'(\mu)\sqrt{n}(\bar{X} - \mu) \tag{3}$$

Since $\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ and $g'(\mu)\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} \mathcal{N}(0, g'(\mu)^2\sigma^2)$, it follows that $\sqrt{n}(g(\bar{X}) - g(\mu)) \xrightarrow{d} \mathcal{N}(0, g'(\mu)^2\sigma^2)$. Therefore, the asymptotic approximation of $g(\bar{X})$ is:

$$g(\bar{X}) \approx \mathcal{N}\left(g(\mu), \frac{g'(\mu)^2\sigma^2}{n}\right) \tag{4}$$

Delta Method. Let Y_n be a sequence of random variances that satisfies $\sqrt{n}(Y_n - \theta) \to \mathcal{N}(0, \sigma^2)$ in distribution. For a given function g such that $g'(\theta)$ exists and is not 0. Then,

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{d} \mathcal{N}(0, \sigma^2 g'(\theta)^2) \tag{5}$$

4.1. Example

For example, suppose $X_1, \dots, X_n$ are iid Bernoulli(p). Then $\mathbb{E}[X_i] = p \equiv \mu$. Therefore the sample mean $\bar{X}$ is a consistent and unbiased estimator of p. The variance is $\text{Var}(X_i) = p(1 - p)$.

Consider the random variable $\bar{X}(1-\bar{X})$. This is of interest because it is a (consistent) estimator for the variance of the Bernoulli distribution. We know this by applying the continuous mapping theorem. In fact, the sample variance can be expressed as $S^2 = \frac{n}{n-1}\bar{X}(1 - \bar{X})$ for the Bernoulli distribution. Let $g(x) = x(1 - x)$, then $g'(x) = 1 - 2x$.

First note that $\mathbb{E}[X_i] = p$ and $\text{Var}(X_i) = p(1 - p)$, by CLT:

$$\sqrt{n}(\bar{X} - p) \xrightarrow{d} \mathcal{N}(0, p(1 - p)) \text{ as } n \to \infty \tag{6}$$

By the Delta method, we can derive the sampling distribution of $\bar{X}(1 - \bar{X})$ as $n \to \infty$.

$$\sqrt{n}(g(\bar{X}) - g(p)) \xrightarrow{d} \mathcal{N}(0, p(1 - p)g'(p)^2) \tag{7}$$

$$\sqrt{n}\left(\bar{X}(1 - \bar{X}) - p(1 - p)\right) \xrightarrow{d} \mathcal{N}\left(0, p(1 - p)(1 - 2p)^2\right) \tag{8}$$

Therefore the asymptotic distribution of $\bar{X}(1-\bar{X})$ is $\bar{X}(1-\bar{X}) \approx \mathcal{N}\left(p(1 - p), \frac{p(1-p)(1-2p)^2}{n}\right)$.

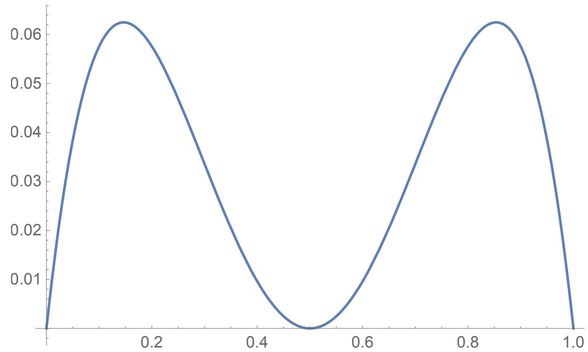

FIGURE 1. $p(1-p)(1-2p)^2$ as a function of p

The asymptotic variance of $\bar{X}(1-\bar{X})$ is $\frac{p(1-p)(1-2p)^2}{n}$. The asymptotic variance of $\bar{X}(1-\bar{X})$ is highest around $p = 0.25$ and $p = 0.75$, see Figure 1. Although $\bar{X}(1-\bar{X})$ is a consistent estimator for the variance of the Bernoulli random variable, the precision of this estimator varies. It is least precise around $p = 0.25$ and $p = 0.75$.

4.2. Another example

Suppose now we are interested in $\frac{p}{1-p}$. This quantity is called the odds ratio. By the Continuous Mapping Theorem, a natural (consistent) estimator for $\frac{p}{1-p}$ would be $\frac{\bar{X}}{1-\bar{X}}$.

Use Delta Method to obtain the asymptotic distribution of $\frac{\bar{X}}{1-\bar{X}}$. From CLT:

$$\sqrt{n}(\bar{X} - p) \xrightarrow{d} \mathcal{N}(0, p(1-p)) \text{ as } n \to \infty$$

Now let $g(x) = \frac{x}{1-x} = \frac{1}{1-x} - 1$. Compute $g'(x) = -\frac{1}{(1-x)^2}$.

$$\sqrt{n}(g(\bar{X}) - g(p)) \xrightarrow{d} \mathcal{N}(0, p(1-p)g'(p)^2) \tag{9}$$

$$\sqrt{n}\left(\frac{\bar{X}}{1-\bar{X}} - \frac{p}{1-p}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{p}{(1-p)^3}\right) \tag{10}$$

Therefore, the asymptotic distribution of $\frac{\bar{X}}{1-\bar{X}}$ is $\frac{\bar{X}}{1-\bar{X}} \approx \mathcal{N}\left(\frac{p}{1-p}, \frac{p}{n(1-p)^3}\right)$.

4.3. Second-order Delta method

What is the asymptotic distribution of $\bar{X}^2$, without assuming Normality?

$$\sqrt{n}(\bar{X} - \mu) \to_d \mathcal{N}(0, \sigma^2) \quad \text{from CLT}$$
$$\sqrt{n}(\bar{X}^2 - \mu^2) \to_d \mathcal{N}(0, (2\mu)^2\sigma^2) \quad \text{from Delta Method}$$

Hence, $\bar{X}^2 \approx \mathcal{N}(\mu^2, \frac{4\mu^2\sigma^2}{n})$. However, what if $\mu = 0$? The asymptotic variance can't be zero! Delta method fails here because $g'(\mu) = 0$. We would need to use second-order Delta Method.

Delta method requires that $g'(\mu) \neq 0$, which fails in some cases. Consider the second-order Taylor expansion of the function $g(x)$ about μ:

$$g(\bar{X}) = g(\mu) + g'(\mu)(\bar{X} - \mu) + \frac{g''(\mu)(\bar{X} - \mu)^2}{2} + R(\bar{X}, \mu) \tag{11}$$

Where the remainder term $R(\bar{X}, \mu) \to 0$ as $\bar{X} \to \mu$, and does so at a rate faster than $(\bar{X} - \mu)^2$. When $g'(\mu) = 0$, we have:

$$g(\bar{X}) - g(\mu) \approx \frac{g''(\mu)(\bar{X} - \mu)^2}{2} \tag{12}$$

when n is large. Since $\sqrt{n}(\bar{X} - \mu)/\sigma \xrightarrow{d} \mathcal{N}(0, 1)$, we have $n(\bar{X} - \mu)^2/\sigma^2 \xrightarrow{d} \chi_1^2$ by the Continuous Mapping Theorem. Hence,

$$n(g(\bar{X}) - g(\mu)) \xrightarrow{d} \frac{g''(\mu)\sigma^2}{2}\chi_1^2 \tag{13}$$

Example:

Going back to our example that finding the asymptotic distribution of $\bar{X}^2$ when $\mu = 0$,

$$\sqrt{n}(\bar{X} - 0) \to_d \mathcal{N}(0, \sigma^2) \quad \text{from CLT}$$
$$n\bar{X}^2 \to_d \sigma^2\chi_1^2 \quad \text{from second-order Delta Method}$$

Now χ_1^2 is equivalent to the Gamma distribution with shape parameter $\frac{1}{2}$, and a scale parameter of 2. That is, $\chi_1^2 = \text{Gamma}(\frac{1}{2}, 2)$. Moreover, $c \times \text{Gamma}(\frac{1}{2}, 2) = \text{Gamma}(\frac{1}{2}, 2c)$ for a constant c. Therefore,

$$\bar{X}^2 \approx \frac{\sigma^2}{n}\chi_1^2 \quad \text{asymptotic approximation}$$
$$\bar{X}^2 \approx \text{Gamma}\left(\frac{1}{2}, \frac{2\sigma^2}{n}\right)$$

When $\mu \neq 0$, the asymptotic distribution is $\bar{X}^2 \approx \mathcal{N}(\mu^2, \frac{4\mu^2\sigma^2}{n})$, and $\bar{X}^2$ converges to μ^2 at a rate of $\frac{1}{\sqrt{n}}$. However, if $\mu = 0$, then $\bar{X}^2 \approx \frac{\sigma^2}{n}\chi_1^2$, and $\bar{X}^2$ converges much faster to μ^2, at a rate of $\frac{1}{n}$. For example, if we consider $\sqrt{n}\bar{X}^2$ when $\mu = 0$, then $\sqrt{n}\bar{X}^2$ would converge to zero in probability.

4.4. Multivariate Delta method

Given a sequence of random vectors $\boldsymbol{Y}_n$, if we have:

$$\sqrt{n}(\boldsymbol{Y}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

where $\xrightarrow{d}$ denotes convergence in distribution, $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, and $\boldsymbol{\theta}$ is a p-vector of parameters, the multivariate Delta Method states that for a function $g : \mathbb{R}^p \to \mathbb{R}^q$ that is continuously differentiable at $\boldsymbol{\theta}$, the following asymptotic distribution holds:

$$\sqrt{n}(g(\boldsymbol{Y}_n) - g(\boldsymbol{\theta})) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{J}_g\boldsymbol{\Sigma}\boldsymbol{J}_g^T)$$

where $\boldsymbol{J}_g$ is the Jacobian matrix of g evaluated at $\boldsymbol{\theta}$, which is a $q \times p$ matrix where the element in the ith row and jth column is

$$[\boldsymbol{J}_g]_{ij} = \frac{\partial g_i(\boldsymbol{\theta})}{\partial \theta_j}$$

$$\boldsymbol{J}_g = \begin{bmatrix} \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_p} \\ \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_2(\boldsymbol{\theta})}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_q(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_q(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_q(\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix}$$

Note that when $p = q = 1$, this reduces to the univariate Delta Method.

4.5. Application of Multivariate Delta Method

Delta method underlies computation of standard errors in many statistical packages. See: https://cran.r-project.org/web/packages/modmarg/vignettes/delta-method.html

To see an example where we apply the multivariate Delta Method, let the data-generating process for $Y_1, \ldots, Y_n$ be $P(Y_i = 1|X_i = x_i) = \Phi(\beta_0 + \beta_1 x_i)$. This is the Probit model for a binary outcome $Y_i \in \{0, 1\}$, where the probability of $Y_i = 1$ given a covariate $X_i = x_i$ is modeled as $P(Y_i = 1|X_i = x_i) = \Phi(\beta_0 + \beta_1 x_i)$, where $\Phi(\cdot)$ is the cdf of the standard normal distribution, and $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ are the model parameters. If we let Φ be the logistic cdf, i.e. $\Phi(t) = \frac{e^t}{1+e^t}$, then we have a Logit model.

Later on, we will see that the maximum-likelihood estimator $(\hat{\beta}_0, \hat{\beta}_1)$ has an asymptotic multivariate Normal distribution. Specifically, let $\hat{\boldsymbol{\beta}}_n = (\hat{\beta}_0, \hat{\beta}_1)^T$ be the maximum likelihood estimators (MLEs) of the parameters. Under standard regularity conditions, the MLEs are asymptotically normally distributed:

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma}$ is the asymptotic variance-covariance matrix of the estimators.

After estimating the parameters, we wish to calculate the predicted probability that the outcome is 1 at a given value x, $\hat{P}(Y = 1|X = x) = \Phi(\hat{\beta}_0 + \hat{\beta}_1 x)$. In another words, we wish to derive the asymptotic distribution of $\Phi(\hat{\beta}_0 + \hat{\beta}_1 x)$.

Let $g(\boldsymbol{\beta}) = \Phi(\beta_0 + \beta_1 x)$, we will use multivariate Delta method to derive

$$\sqrt{n}(g(\hat{\boldsymbol{\beta}}) - g(\boldsymbol{\beta})) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{J}_g \boldsymbol{\Sigma} \boldsymbol{J}_g^T)$$

The Jacobian of $g(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta} = (\beta_0, \beta_1)$ is:

$$\boldsymbol{J}_g = \begin{bmatrix} \frac{\partial g}{\partial \beta_0} & \frac{\partial g}{\partial \beta_1} \end{bmatrix}.$$

$$\frac{\partial g}{\partial \beta_0} = \phi(\beta_0 + \beta_1 x),$$

and

$$\frac{\partial g}{\partial \beta_1} = x\phi(\beta_0 + \beta_1 x),$$

where $\phi(\cdot)$ is the pdf of Φ.

Thus, the Jacobian matrix is:

$$\boldsymbol{J}_g = \begin{bmatrix} \phi(\beta_0 + \beta_1 x) & x\phi(\beta_0 + \beta_1 x) \end{bmatrix}.$$

By the multivariate Delta Method, the asymptotic distribution of the predicted probability is:

$$\sqrt{n}(\Phi(\hat{\beta}_0 + \hat{\beta}_1 x) - \Phi(\beta_0 + \beta_1 x)) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{J}_g \boldsymbol{\Sigma} \boldsymbol{J}_g^T)$$

Even though β_0, β_1 is not known in the formula for the asymptotic variance, we can plug in any consistent estimator of β_0, β_1, which is justified from Slutsky's and the Continuous Mapping Theorem. Note that both Slutsky's and the Continuous Mapping Theorem are similarly defined for random vectors or matrices. For instance, $\boldsymbol{J}_g \boldsymbol{\Sigma} \boldsymbol{J}_g^T$ is a (scalar) continuous function of $\boldsymbol{\beta} = (\beta_0, \beta_1)$. Thus if $\hat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}$, then $\hat{\boldsymbol{J}}_g \boldsymbol{\Sigma} \hat{\boldsymbol{J}}_g^T$ also converges in probability to $\boldsymbol{J}_g \boldsymbol{\Sigma} \boldsymbol{J}_g^T$.

Another quantity of interest is the marginal effect,

$$\frac{\partial \Phi(\hat{\beta}_0 + \hat{\beta}_1 x)}{\partial x} = \hat{\beta}_1 \phi(\hat{\beta}_0 + \hat{\beta}_1 x)$$

Whose asymptotic distribution can be computed following the steps above.