

LECTURE 6: LARGE SAMPLE THEORY

MECO 7312.

INSTRUCTOR: DR. KHAI CHIONG

SEPTEMBER 25, 2024

1. Continuous Mapping Theorem

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

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

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

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

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

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

$$\begin{aligned}\mathbb{E}[g(X)] &\geq g(\mathbb{E}[X]) \\ \mathbb{E}[-g(X)] &\leq -g(\mathbb{E}[X])\end{aligned}$$

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

$$\begin{aligned}\mathbb{E}[\sqrt{S^2}] &< \sqrt{\mathbb{E}[S^2]} \\ \mathbb{E}[\sqrt{S^2}] &< \sqrt{\sigma^2} \\ \mathbb{E}[S] &< \sigma\end{aligned}$$

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, \dots 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, $\bar{X} - \mu \xrightarrow{p} 0$

However now consider $\sqrt{n}(\bar{X} - \mu)$. As $n \rightarrow \infty$, we have two conflicting convergence: (i) $\bar{X} - \mu \rightarrow 0$ in probability, (ii) but $\sqrt{n} \rightarrow \infty$. Somehow, they balance each other out in the sense that $\sqrt{n}(\bar{X} - \mu)$ converges to a random variable as $n \rightarrow \infty$. 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 $n \rightarrow \infty$. That is, $\lim_{n \rightarrow \infty} P(\sqrt{n}(\bar{X}_n - \mu)/\sigma \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ for all $x \in \mathbb{R}$. Equivalently, $\sqrt{n}(\bar{X}_n - \mu)$ converges in distribution to $\mathcal{N}(0, \sigma^2)$ as $n \rightarrow \infty$.

\sqrt{n} is also called the “rate of convergence” of the sequence $\bar{X} - \mu$. In another words, $(\bar{X} - \mu)/\sigma$ decays at the same rate to zero as $\frac{1}{\sqrt{n}}$ asymptotically. A weaker form of CLT is proven in Casella-Berger, the proof relies on moment generating function and Taylor’s expansion.

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.

¹Which 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, $\sqrt{n}(\bar{X} - \frac{1}{\lambda}) \rightarrow_d \mathcal{N}(0, \frac{1}{\lambda^2})$, 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. We can repeat the above simulation with other data-generating process.

3. Slutsky's theorem

If $X_n \xrightarrow{d} X$ in distribution, and $Y_n \xrightarrow{p} a$ where a is a constant, then

$$(1) \quad Y_n X_n \xrightarrow{d} aX \text{ in distribution}$$

$$(2) \quad X_n + Y_n \xrightarrow{d} X + a \text{ in distribution}$$

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 $\sigma^2 \equiv \text{Var}(X_i)$.

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

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

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 $\mathcal{N}(\mu, \frac{\sigma^2}{n})$.²

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

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

Similarly,

$$n^{3/4}(\bar{X}_n - \mu)/\sigma = n^{1/4}n^{1/2}(\bar{X}_n - \mu)/\sigma \rightarrow \infty$$

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? Often we are interested in some functions of the sample mean. For example, \bar{X}^2 , $e^{\bar{X}}$, $\log \bar{X}$.

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

$$(3) \quad g(x) = g(a) + g'(a)(x - a) + R(x, a)$$

$R(x, a)$ 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, $\lim_{x \rightarrow a} R(x, a)/(x - a) = 0$. 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.

²However 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.

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

$$(4) \quad g(\bar{X}) = g(\mu) + g'(\mu)(\bar{X} - \mu) + o(\bar{X} - \mu)$$

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

$$(5) \quad \sqrt{n}(g(\bar{X}) - g(\mu)) \approx g'(\mu)\sqrt{n}(\bar{X} - \mu)$$

Since $\sqrt{n}(X - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, by Slutsky's theorem, $g'(\mu)\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} \mathcal{N}(0, g'(\mu)^2\sigma^2)$. 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:

$$(6) \quad g(\bar{X}) \approx \mathcal{N}\left(g(\mu), \frac{g'(\mu)^2\sigma^2}{n}\right)$$

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

$$(7) \quad \sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{d} \mathcal{N}(0, \sigma^2 g'(\theta)^2)$$

4.1. Example

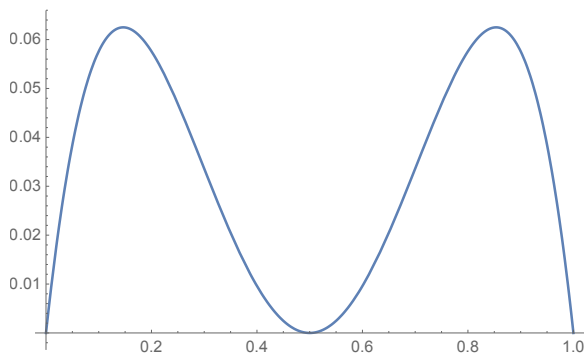
For example, suppose X_1, \dots, X_n are iid Bernoulli(p). Then $\mathbb{E}[X_i] = p \equiv \mu$. Therefore the sample mean \bar{X} is a consistent and unbiased estimator of p . The variance is $\text{Var}(X_i) = p(1 - p)$.

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:

$$(8) \quad \sqrt{n}(\bar{X} - p) \xrightarrow{d} \mathcal{N}(0, p(1 - p)) \text{ as } n \rightarrow \infty$$

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

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

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

$$(10) \quad \sqrt{n}(\bar{X}(1-\bar{X}) - p(1-p)) \xrightarrow{d} \mathcal{N}(0, p(1-p)(1-2p)^2)$$

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

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:

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

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

$$(11) \quad \sqrt{n}(g(\bar{X}) - g(p)) \xrightarrow{d} \mathcal{N}(0, p(1-p)g'(p)^2)$$

$$(12) \quad \sqrt{n}\left(\frac{\bar{X}}{1-\bar{X}} - \frac{p}{1-p}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{p}{(1-p)^3}\right)$$

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?

$$\begin{aligned}\sqrt{n}(\bar{X} - \mu) &\rightarrow_d \mathcal{N}(0, \sigma^2) && \text{from CLT} \\ \sqrt{n}(\bar{X}^2 - \mu^2) &\rightarrow_d \mathcal{N}(0, (2\mu)^2\sigma^2) && \text{from Delta Method}\end{aligned}$$

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 :

$$(13) \quad g(\bar{X}) = g(\mu) + g'(\mu)(\bar{X} - \mu) + \frac{g''(\mu)(\bar{X} - \mu)^2}{2} + R(\bar{X}, \mu)$$

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

$$(14) \quad g(\bar{X}) - g(\mu) \approx \frac{g''(\mu)(\bar{X} - \mu)^2}{2}$$

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,

$$(15) \quad n(g(\bar{X}) - g(\mu)) \xrightarrow{d} \frac{g''(\mu)\sigma^2}{2} \chi_1^2$$

Example:

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

$$\begin{aligned}\sqrt{n}(\bar{X} - 0) &\rightarrow_d \mathcal{N}(0, \sigma^2) && \text{from CLT} \\ n\bar{X}^2 &\rightarrow_d \sigma^2 \chi_1^2 && \text{from second-order Delta Method}\end{aligned}$$

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,

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

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 \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 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 $\boldsymbol{\theta}_n$, if we have:

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

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

$$\sqrt{n}(g(\boldsymbol{\theta}_n) - g(\boldsymbol{\theta})) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{J}_g \mathbf{V} \mathbf{J}_g')$$

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

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

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

4.5. Side remark

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

Later on, we will see that the sampling distribution of coefficients from regressions has a Normal distribution too. In many cases, we are interested in functions of the coefficients. For example, in Probit regression, $P(y = 1) = \Phi(a + bx)$. The estimated coefficients (a, b) has a Normal sampling distribution, but by itself, b has

no meaningful interpretation. Of interest is the marginal effect: $dP(y = 1)/dx = b\phi(a+bx)$. Delta method allows us to compute the standard error of $dP(y = 1)/dx = b\phi(a + bx)$ via asymptotic approximation, which is faster and more accurate than bootstrapping.