LECTURE 6: LARGE SAMPLE THEORY

MECO 7312.
INSTRUCTOR: DR. KHAI CHIONG
SEPTEMBER 29, 2021

1. Continuous Mapping Theorem

Suppose that the sequence of random variable X_n converges in probability to θ as $n \to \infty$. 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 $n \to \infty$. 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 actually 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 $\mathbb{E}[g(X)] < g(\mathbb{E}[X])$. Then recall that $\mathbb{E}[S^2] = \sigma^2$. Since $f(x) = \sqrt{x}$ is a concave function, it follows that

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

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 $\text{Var}(X_i) = \sigma^2 < \infty$. 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, $\bar{X} - \mu \to^p 0$

However 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 $\sqrt{n} \to \infty$. Magically, 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, denote $G_n(x)$ as the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$, then we have $\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$ for all $x \in \mathbb{R}$. Equivalently, $\sqrt{n}(\bar{X}_n - \mu)$ converges in distribution to $\mathcal{N}(0, \sigma^2)$ as $n \to \infty$.

$\sqrt{n}$ 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 $\frac{1}{\sqrt{n}}$. 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 distribution. For example, if X_i is Uniformly distributed, what is the distribution of the sample mean $\bar{X}_n$? More generally and practically, we do not even know what the data-generating process is.

[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 Approximation is to appeal to asymptotically large n (hence using CLT), in order to say something about the distribution of a statistic, *even when n is finite, for any underlying population density.*

Even when n is finite and not large, we can usually take $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ to be the asymptotic approximation for the distribution of $\bar{X}$. This approximation holds remarkably well.

2.2. Simulating the Central Limit Theorem

We can check for ourselves using simulations. Take X_i to be exponentially distributed, i.e. the pdf of X_i is $f(x) = \lambda e^{-\lambda x}$. Check using simulations that the distribution of $\bar{X}$ is approximately $\mathcal{N}(\frac{1}{\lambda}, \frac{1}{n\lambda^2})$. Try it with Uniform, Discrete Uniform, Binomial, Geometric, etc.

The following lines in R generate a random sample of sample size $n = 100$ from the Exponential distribution with parameter $\lambda = 2$. The random sampling is repeated $s = 10000$ times. The simulated dataset is stored in a $n \times s$ matrix called X.

```
n <- 100
s <- 10000
lambda <- 2
X <- matrix(rexp(n*s,lambda), nrow = n, ncol=s)
```

We then compute the sample mean for each random sample, i.e. taking the mean of the matrix X for each column, and store it in a s-dimensional vector m. Now we plot the histogram of m. This histogram is the sampling distribution of $\bar{X}$. This looks somewhat like a Normal distribution.

```
m <- apply(X,2,mean)
hist(m,freq = 0,breaks = 100)
```

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 $\bar{X} \sim \mathcal{N}(\frac{1}{\lambda}, \frac{1}{n\lambda^2})$.

In the following line, we overlay the previous histogram with a plot of the pdf of $\mathcal{N}(\frac{1}{\lambda}, \frac{1}{n\lambda^2})$. Indeed, we see that it aligns closely with the CLT prediction.

```
z <- seq(min(m),max(m),length.out = 1000)
lines(z,dnorm(z,mean=1/lambda,sd=1/(sqrt(n)*lambda)))
```

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.

3. Slutsky's theorem

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

$$Y_n X_n \xrightarrow{d} aX \text{ in distribution} \tag{1}$$

$$X_n + Y_n \xrightarrow{d} X + a \text{ in distribution} \tag{2}$$

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

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

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 $S_n \xrightarrow{p} \sigma$ and $\frac{\sigma}{S_n} \xrightarrow{p} 1$ by the Continuous Mapping Theorem. We have:

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

Hence, *for large n, the distribution of $\bar{X}$ is approximately $\mathcal{N}(\mu, \frac{S^2}{n})$.*[2]

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

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

[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.

Similarly,

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

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, $\bar{X}^2$, $e^{\bar{X}}$, $\log \bar{X}$. Delta method is the key.

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:

$$g(x) = g(a) + g'(a)(x - a) + R(x, a) \tag{3}$$

$R(x, a)$ 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 $\mu \equiv \mathbb{E}[X_i]$,

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

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

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

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:

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

Note that the Taylor's approximation in Equation 4 is increasingly more accurate as $\bar{X}$ is close to μ. Since $\bar{X} \xrightarrow{p} \mu$, we can then justify the use of this approximation in the limit. We summarized everything in the following theorem.

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,

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

4.1. Example

For example, suppose $X_1, \ldots, X_n$ 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:

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

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

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

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

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 $\frac{\bar{X}}{1-\bar{X}}$.

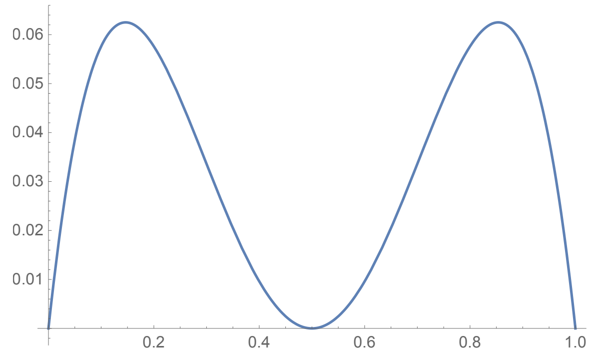

FIGURE 1. $p(1-p)(1-2p)^2$ as a function of p

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

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

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

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. Another example

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

4.4. 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.

4.5. Second-order Delta method (optional)

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:

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

Where the remainder term $R(\bar{X}, \mu) \to 0$ as $\bar{X} \to \mu$, and does so at a rate faster than $(\bar{X} - \mu)^2$. That is, $R(\bar{X}, \mu) = o((\bar{X} - \mu)^2)$. As $n \to \infty$, we have:

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

Since $\sqrt{n}(\bar{X} - \mu)/\sigma \xrightarrow{d} \mathcal{N}(0, 1)$, we have $n(\bar{X} - \mu)^2/\sigma^2 \xrightarrow{d} \chi^2_1$ by the Continuous Mapping Theorem. Hence,

$$n(g(\bar{X}) - g(\mu)) \xrightarrow{d} \frac{g''(\mu)\sigma^2}{2}\chi^2_1 \tag{15}$$

Example:

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

$$\sqrt{n}(\bar{X} - 0) \to_d \mathcal{N}(0, \sigma^2) \quad \text{from CLT}$$
$$n\bar{X}^2 \to_d \sigma^2\chi^2_1 \quad \text{from second-order Delta Method}$$

Now χ^2_1 is equivalent to the Gamma distribution with shape parameter $\frac{1}{2}$, and a scale parameter of 2. That is, $\chi^2_1 = \text{Gamma}(\frac{1}{2}, 2)$. Moreover, $c \times \text{Gamma}(\frac{1}{2}, 2) = \text{Gamma}(\frac{1}{2}, 2c)$ for a constant c. Therefore,

$$\bar{X}^2 \approx \frac{\sigma^2}{n}\chi^2_1 \quad \text{asymptotic approximation}$$
$$\bar{X}^2 \approx \text{Gamma}\left(\frac{1}{2}, \frac{2\sigma^2}{n}\right)$$

When $\mu \neq 0$, $\bar{X}^2$ converges to μ^2 at a rate of $\sqrt{n}$, but suddenly if $\mu = 0$, $\bar{X}^2$ converges much faster to μ^2, at a rate of n. That is, if we consider $\sqrt{n}\bar{X}^2$ when $\mu = 0$, then $\sqrt{n}\bar{X}^2$ would converge to zero in probability.