Econ 508B: Lecture 10

LLN, CLT and Local Linear Approximation

Hongyi Liu

Washington University in St. Louis

August 5, 2017

Outline

1. Laws of Large Numbers
2. Continuous Mapping Theorem
3. Linderberg-Feller Central Limit Theorem
4. Local Linear Approximation: Delta Method

Outline

1. Laws of Large Numbers
2. Continuous Mapping Theorem
3. Linderberg-Feller Central Limit Theorem
4. Local Linear Approximation: Delta Method

Weak Law of Large Numbers

A sequence $\{X_n\}, n \geq 1$ of random variables is said to follow the WLLN with normalizing sequences of real numbers $\{a_n\}, n \geq 1$ and $\{b_n\}, n \geq 1$ if

$$\frac{S_n - a_n}{b_n} \xrightarrow{p} 0 \text{ as } n \to \infty$$

where $S_n = \sum_{i=1}^{n} X_i$ for $n \geq 1$.

Theorem 1.1

Let $\{X_n\}, n \geq 1$ be a sequence of i.i.d. random variables such that $EX_1^2 < \infty$. Then

$$\bar{X}_n \equiv \frac{X_1 + \cdots + X_n}{n} \xrightarrow{p} EX_1.$$

This has $a_n = nEX_1$ and $b_n = n$.

Proof

Using *Chebychev's inequality*, for any $\epsilon > 0$,

$$P(|\bar{X}_n - EX_1| > \epsilon) \leq \frac{Var(\bar{X}_n)}{\epsilon^2} = \frac{1}{\epsilon^2}\frac{\sigma^2}{n},$$

where $\sigma^2 = Var(X_1) < \infty$. Since

$$\frac{\sigma^2}{n\epsilon^2} \to 0 \text{ as } n \to \infty,$$

Hence $\bar{X}_n \xrightarrow{p} EX_1$.

Example: Sample Standard Error

Let $X_1, ..., X_n$ be i.i.d. real-valued random variables with common mean μ and finite variance σ^2. Consider the unbiased estimator of σ^2

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2,$$

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. According to *WLLN*, we have $\bar{X}_n \xrightarrow{p} \mu$ and $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{p} E(X_1^2) = \mu^2 + \sigma^2$; By **continuous mapping theorem** which will be covered later on, $\bar{X}_n^2 \xrightarrow{p} \mu^2$. Therefore,

$$\frac{n-1}{n}S_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2 \xrightarrow{p} \sigma^2$$

Thus $S_n^2 \xrightarrow{p} \sigma^2$. By **continuous mapping theorem**, then $S_n \xrightarrow{p} \sigma$.

A sequence $\{X_n\}, n \geq 1$ of random variables is said to follow *SLLN* with normalizing sequences $\{a_n\}, \{b_n\}, n \geq 1$ if

$$\frac{S_n - a_n}{b_n} \xrightarrow{a.s.} 0 \text{ as } n \to \infty$$

where $S_n = \sum_{i=1}^{n} X_i$.

Theorem 1.2

Let $\{X_1, ..., X_n\}$ be i.i.d. with $EX_1^4 < \infty$. Then

$$\bar{X}_n \xrightarrow{a.s.} EX_1$$

Complete Convergence

Borel-Cantelli lemma is a useful proposition to understand the connection between **complete convergence** and **convergence almost surely** for *SLLN*. Let us first introduce the notion of *complete convergence.*

Definition 1.1 (complete convergence)

X_n is said to converge completely to X, denoted by $X_n \xrightarrow{c} X$ if $\forall \delta > 0$,

$$\sum_{n=1}^{\infty} P\{|X_n - X| > \delta\} < \infty.$$

What is more, $X_n \xrightarrow{c} X \Rightarrow X_n \xrightarrow{a.s.} X$ as shown by the first lemma of *Borel-Cantelli.*

Proposition 1.1

Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables on some probability space $(\Omega, \mathcal{F}, P)$.

(a) If $\sum_{n=1}^{\infty} P(|X_n| > \epsilon) < \infty$ for each $\epsilon > 0$, then

$$P(\lim_{n\to\infty} X_n = 0) = 1.$$

(b) If $\{X_n\}_{n\geq 1}$ are **pairwise independent** and $P(\lim_{n\to\infty} X_n = 0) = 1$, then $\sum_{n=1}^{\infty} P(|X_n| > \epsilon) < \infty$ for each $\epsilon > 0$.

Remark 1.1

The interpretation of *Borel-Cantelli lemma* is somewhat ambiguous from the direct definition, however, it will be more intuitive if we use the language of stochastic process or physics.

WLLN&SLLN: finite absolute first moment

Theorem 1.3 (Khintchine's WLLN)

$\{X_1, ..., X_n\}$ are i.i.d., $E|X_1| < \infty . EX_1 = 0$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} 0.$$

Theorem 1.4 (Kolmogorov's SLLN)

$\{X_1, ..., X_n\}$ are i.i.d. random variables, then

$$\frac{S_n - nc}{n} \xrightarrow{a.s.} 0.$$

for some $c \in \mathbb{R}$ if and only if $E|X_1| < \infty$, in which case $EX_1 = c$.

Proof of these LLNs and also CLT will be simpler by working with *characteristic functions*.

Outline

Mann–Wald theorem(Continuous Mapping Theorem)

Definition 2.1

If $X_n \xrightarrow{d} X$ as $n \to \infty$ and $g(\cdot)$ has the set of discontinuity points D_g such that $P(X \in D_g) = 0$, then

$$g(X_n) \xrightarrow{d} g(X) \text{ as } n \to \infty$$

In general, we also have

$$X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$$
$$X_n \xrightarrow{a.s.} X \Rightarrow g(X_n) \xrightarrow{a.s.} g(X)$$

Proof of CMT

For the proof of the first CMT, it can be simply proved by the *Helly-Bray Theorem*.

Theorem 2.1 (Helly-Bray Theorem)

$X_n \xrightarrow{d} X$ if and only if $Ef(X_n) \to Ef(X)$ for all bounded and continuous real-valued functions $f(x)$.

Proof:

Let $f(x)$ be bounded and continuous, it suffices to show that $Ef[g(X_n)] \to Ef[g(X)]$.

- $g(\cdot)$ is continuous, then the composition $f \circ g = f(g(\cdot))$ is also bounded and continuous.
- Hence apply the *Helly-Bray theorem* again proving that $Ef[g(X_n)] \to Ef[g(X)]$ as desired.

Slutsky's Theorem

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ as $n \to \infty$, where c is a constant, then

(1) $X_n + Y_n \xrightarrow{d} X + c$,

(2) $X_n Y_n \xrightarrow{d} Xc$,

(3) $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$ if $c \neq 0$.

Outline

Generic CLT for IID Random Variables

Theorem 3.1

Let $X_1, X_2, ...$ be i.i.d. with $E(X_i^2) < \infty$. Denote $\mu = E(X_i)$ and $\sigma^2 = Var(X_i)$ and define $S_n = \sum_{i=1}^n X_i$. Then

$$\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2).$$

However, this case is too restrictive to be applied in applications. **Two** parts can be simplified:

- It should be enough to have independent random variables for each n rather than an infinite sequence of random variables;
- The assumption of i.i.d. can be simplified into **only independence** and identically distribution is not necessary.

It motivates to use *triangular arrays*, which are collections of random variables $\{X_{n,m}\}_{1 \leq m \leq n}$, where independence is assumed **only** for each fixed n, rather than for an infinite sequence.

Linderberg-Feller CLT

Let $\{X_{n,m}\}_{1 \leq m \leq n}$ be independent for each n, and $E(X_{n,m}) = 0$ and $E(X_{n,m}^2) < \infty$. Let $S_n = \sum_{m=1}^n X_{n,m}$ and suppose the following *Lindeberg-Feller* conditions hold:

- (1): $\exists \sigma^2 \in [0, \infty)$ such that as $n \to \infty$,

$$\sum_{m=1}^{n} E(X_{n,m}^2) \to \sigma^2$$

- (2): For every $\epsilon > 0$, and as $n \to \infty$,

$$\sum_{m=1}^{n} E(X_{n,m}^2 \mathbb{1}\{|X_{n,m}| > \epsilon\}) \to 0.$$

Then $S_n \xrightarrow{d} N(0, \sigma^2)$.

A different but equivalent Lindeberg-Feller CLT

Assume that for each n, $X_{n1}, ..., X_{nn}$ are independent, $EX_{nm} = \mu_{nm}$ and $VarX_{nm} = \sigma_{nm}^2 < \infty$ for $1 \leq m \leq n$. Let

$$Y_{nm} = X_{nm} - \mu_{nm}, T_n = \sum_{m=1}^{n} Y_{nm}, s_n^2 = VarT_n = \sum_{m=1}^{n} \sigma_{nm}^2.$$

As before, T_n/s_n has mean 0 and variance 1. Our goal is to give conditions for

$$\frac{T_n}{s_n} \xrightarrow{d} N(0,1).$$

Lindeberg condition & Feller's condition

Definition 3.1 (Lindeberg condition)

For every $\epsilon > 0$,

$$\frac{1}{s_n^2}\sum_{m=1}^{n} E(Y_{nm}^2 \mathbb{1}\{|Y_{nm}| \geq \epsilon s_n\}) \to 0.$$

Definition 3.2

Intuitively, *Feller's condition* states that the contribution of the tail of each Y_{nm} to s_n^2 should be negligible:

$$\frac{1}{s_n^2}\max_{m \leq n} \sigma_{nm}^2 \to 0.$$

In fact, *Lindeberg's condition* implies *Feller's condition*, which really means that all the terms in the sum s_n^2 are **uniformly asymptotically negligible (UAN)**.

Lyapunov's condition

Usually, the *Lindeberg's condition* is not easy to check with, thus we need to have a more tractable condition to guarantee the *asymptotic normality*.

Definition 3.3

$$\frac{1}{s_n^{2+\delta}} \sum_{m=1}^{n} E(|Y_{nm}|^{2+\delta}) \to 0.$$

To show that, since $|Y_{nm}| \geq \epsilon s_n$ implies $|Y_{nm}/\epsilon s_n|^{\delta} \geq 1$,

$$0 \leq \frac{1}{s_n^2} \sum_{m=1}^{n} E(Y_{nm}^2 \mathbb{1}\{|Y_{nm}| \geq \epsilon s_n\}) \leq \frac{1}{\epsilon^{\delta} s_n^{2+\delta}} \sum_{m=1}^{n} E(Y_{nm}^{2+\delta} \mathbb{1}\{|Y_{nm}| \geq \epsilon s_n\})$$

$$\leq \frac{1}{\epsilon^{\delta} s_n^{2+\delta}} \sum_{m=1}^{n} E(|Y_{nm}^{2+\delta}|); \qquad \text{(Ljapunov's condition)}$$

The right hand side tends to 0, the Lindeberg Condition is satisfied.

Failure of Lyapunov Condition

Lyapunov's condition will fail when the variance is finite, but higher-order moments are infinite.

Example 3.1

Consider a Pareto distribution with the corresponding density as follows

$$f(x) = \frac{c}{|x|^3(\log|x|)^2}, \ |x| > 2$$

with c a normalizing constant.
If $X_0, X_1, ...$ have this density, then

$$\sigma^2 = Var(X) = 2c\int_2^\infty \frac{dx}{x(\log x)^2} < \infty,$$

whereas $E|X|^r = \infty$ for all $r > 2$, so that Lyapunov's condition fails.
Despite this, the CLT still holds.

Outline

Motivation

Question: According to Central Limit Theorem, we have an existing sequence converging to a normal distribution under some condition. However, what if we consider the asymptotic distribution of a function of such a sequence?

Recall the idea of using a first-order (linear) Taylor expansion of a known function, in the neighborhood of that constant limit, is a very useful technique known as the **delta method**, named for the Δ in

$$g(x + \Delta x) \approx g(x) + \Delta x g'(x).$$

Delta Method

Theorem 4.1 (Delta Method)

If $g'(\mu)$ exists and $n^b(X_n - \mu) \xrightarrow{d} X$ for $b > 0$, then

$$n^b\{g(X_n) - g(\mu)\} \xrightarrow{d} g'(\mu)X.$$

Proof:

To show delta method, firstly we have $X_n \xrightarrow{p} \mu$, since $X_n - \mu = n^{-b}n^b(X_n - \mu)$, and $n^b(X_n - \mu) \xrightarrow{d} X$ and $n^{-b} \to 0$, thus $X_n - \mu \xrightarrow{p} 0$ because of **Slutsky's theorem**. Apply the first-order Taylor expansion, then

$$n^b\{g(X_n) - g(\mu)\} = n^b\{g'(\xi_n)(X_n - \mu)\} \text{ for } \xi_n \in [X_n, \mu]/\xi_n \in [\mu, X_n]$$

Since $g'(\cdot)$ is continuous, using **Continuous Mapping Theorem**, $g'(\xi_n) \xrightarrow{p} g'(\mu)$. Then using **Slutsky's theorem** again,

$$n^b\{g(X_n) - g(\mu)\} \xrightarrow{d} g'(\mu)X.$$