Lecture 9: The law of large numbers and central limit theorem

Theorem 1.14

Let $X_1, X_2, \ldots$ be independent random variables with finite expectations.

(i) (The SLLN). If there is a constant $p \in [1,2]$ such that

$$\sum_{i=1}^{\infty} \frac{E|X_i|^p}{i^p} < \infty, \tag{1}$$

then

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - EX_i) \to_{a.s.} 0.$$

(ii) (The WLLN). If there is a constant $p \in [1,2]$ such that

$$\lim_{n\to\infty} \frac{1}{n^p}\sum_{i=1}^{n} E|X_i|^p = 0, \tag{2}$$

then

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - EX_i) \to_p 0.$$

Remarks

- Note that (1) implies (2) (Lemma 1.6).
- The result in Theorem 1.14(i) is called Kolmogorov's SLLN when $p = 2$ and is due to Marcinkiewicz and Zygmund when $1 \le p < 2$.
- An obvious sufficient condition for (1) with $p \in (1,2]$ is $\sup_n E|X_n|^p < \infty$.
- The WLLN and SLLN have many applications in probability and statistics.

Example 1.32

Let f and g be continuous functions on $[0,1]$ satisfying $0 \le f(x) \le Cg(x)$ for all x, where $C > 0$ is a constant.
We now show that

$$\lim_{n\to\infty} \int_0^1 \int_0^1 \cdots \int_0^1 \frac{\sum_{i=1}^n f(x_i)}{\sum_{i=1}^n g(x_i)} dx_1 dx_2 \cdots dx_n = \frac{\int_0^1 f(x)dx}{\int_0^1 g(x)dx} \qquad (3)$$

(assuming that $\int_0^1 g(x)dx \neq 0$).

Example 1.32 (continued)

$X_1, X_2, ...$ be i.i.d. random variables having the uniform distribution on $[0, 1]$.

By Theorem 1.2,

$$E[f(X_1)] = \int_0^1 f(x)dx < \infty, \quad E[g(X_1)] = \int_0^1 g(x)dx < \infty.$$

By the SLLN (Theorem 1.13(ii)),

$$\frac{1}{n}\sum_{i=1}^n f(X_i) \to_{a.s.} E[f(X_1)], \quad \frac{1}{n}\sum_{i=1}^n g(X_i) \to_{a.s.} E[g(X_1)],$$

By Theorem 1.10(i),

$$\frac{\sum_{i=1}^n f(X_i)}{\sum_{i=1}^n g(X_i)} \to_{a.s.} \frac{E[f(X_1)]}{E[g(X_1)]}. \tag{4}$$

Since the random variable on the left-hand side of (4) is bounded by C, result (3) follows from the dominated convergence theorem and the fact that the left-hand side of (3) is the expectation of the random variable on the left-hand side of (4).

Example

Let $T_n = \sum_{i=1}^{n} X_i$, where X_n's are independent random variables satisfying $P(X_n = \pm n^\theta) = 0.5$ and $\theta > 0$ is a constant.
We want to show that $T_n/n \to_{a.s.} 0$ when $\theta < 0.5$.
For $\theta < 0.5$,

$$\sum_{n=1}^{\infty} \frac{EX_n^2}{n^2} = \sum_{n=1}^{\infty} \frac{n^{2\theta}}{n^2} < \infty.$$

By the Kolmogorov strong law of large numbers, $T_n/n \to_{a.s.} 0$.

Example (Exercise 165)

Let $X_1, X_2, ...$ be independent random variables.
Suppose that

$$\frac{1}{\sigma_n} \sum_{j=1}^{n} (X_j - EX_j) \to_d N(0,1),$$

where $\sigma_n^2 = \mathrm{var}(\sum_{j=1}^{n} X_j)$.

Example (Exercise 165)

We want to show that

$$\frac{1}{n}\sum_{j=1}^{n}(X_j - EX_j) \to_p 0 \quad \text{iff} \quad \sigma_n/n \to 0.$$

If $\sigma_n/n \to 0$, then by Slutsky's theorem,

$$\frac{1}{n}\sum_{j=1}^{n}(X_j - EX_j) = \frac{\sigma_n}{n}\frac{1}{\sigma_n}\sum_{j=1}^{n}(X_j - EX_j) \to_d 0.$$

Assume now σ_n/n does not converge to 0 but $n^{-1}\sum_{j=1}^{n}(X_j - EX_j) \to_p 0$.
Without loss of generality, assume that $\sigma_n/n \to c \in (0,\infty]$.
By Slutsky's theorem,

$$\frac{1}{\sigma_n}\sum_{j=1}^{n}(X_j - EX_j) = \frac{n}{\sigma_n}\frac{1}{n}\sum_{j=1}^{n}(X_j - EX_j) \to_p 0.$$

This contradicts the fact that $\sum_{j=1}^{n}(X_j - EX_j)/\sigma_n \to_d N(0,1)$.
Hence, $n^{-1}\sum_{j=1}^{n}(X_j - EX_j)$ does not converge to 0 in probability.

The central limit theorem

The WLLN and SLLN may not be useful in approximating the distributions of (normalized) sums of independent random variables.

We need to use the *central limit theorem* (CLT), which plays a fundamental role in statistical asymptotic theory.

Theorem 1.15 (Lindeberg's CLT)

Let $\{X_{nj}, j = 1, ..., k_n\}$ be independent random variables with $k_n \to \infty$ as $n \to \infty$ and

$$0 < \sigma_n^2 = \text{var}\left(\sum_{j=1}^{k_n} X_{nj}\right) < \infty, \quad n = 1, 2, ...,$$

If

$$\frac{1}{\sigma_n^2} \sum_{j=1}^{k_n} E\left[(X_{nj} - EX_{nj})^2 I_{\{|X_{nj} - EX_{nj}| > \epsilon\sigma_n\}}\right] \to 0 \quad \text{for any } \epsilon > 0, \qquad (5)$$

then

$$\frac{1}{\sigma_n} \sum_{j=1}^{k_n} (X_{nj} - EX_{nj}) \to_d N(0, 1).$$

Proof

Considering $(X_{nj} - EX_{nj})/\sigma_n$, without loss of generality we may assume $EX_{nj} = 0$ and $\sigma_n^2 = 1$ in this proof.
Let $t \in \mathscr{R}$ be given.
From the inequality

$$|e^{\sqrt{-1}tx} - (1 + \sqrt{-1}tx - t^2x^2/2)| \le \min\{|tx|^2, |tx|^3\},$$

the ch.f. of X_{nj} satisfies

$$\left|\phi_{X_{nj}}(t) - \left(1 - t^2\sigma_{nj}^2/2\right)\right| \le E\left(\min\{|tX_{nj}|^2, |tX_{nj}|^3\}\right),$$

where $\sigma_{nj}^2 = \text{var}(X_{nj})$.
For any $\varepsilon > 0$, the right-hand side of the previous expression is bounded by

$$E(|tX_{nj}|^3 I_{\{|X_{nj}|<\varepsilon\}}) + E(|tX_{nj}|^2 I_{\{|X_{nj}|\ge\varepsilon\}}),$$

which is bounded by

$$\varepsilon|t|^3\sigma_{nj}^2 + t^2 E(X_{nj}^2 I_{\{|X_{nj}|\ge\varepsilon\}}).$$

Proof (continued)

Summing over j and using $\sigma_n^2 = 1$, we obtain that

$$\sum_{j=1}^{k_n} \left| \phi_{X_{nj}}(t) - \left(1 - t^2\sigma_{nj}^2/2\right) \right| \le \sum_{j=1}^{k_n} \{\varepsilon |t|^3 \sigma_{nj}^2 + t^2 E(X_{nj}^2 I_{\{|X_{nj}|\ge\varepsilon\}})\}$$

$$= \varepsilon |t|^3 + t^2 \sum_{j=1}^{k_n} E(X_{nj}^2 I_{\{|X_{nj}|\ge\varepsilon\}}) \to \varepsilon |t|^3$$

by condition (5).

Also by condition (5) and $\sigma_n^2 = 1$,

$$\max_{j\le k_n} \frac{\sigma_{nj}^2}{\sigma_n^2} \le \varepsilon^2 + \max_{j\le k_n} E(X_{nj}^2 I_{\{|X_{nj}|>\varepsilon\}}) \to \varepsilon^2$$

Since $\varepsilon > 0$ is arbitrary and t is fixed,

$$\sum_{j=1}^{k_n} \left| \phi_{X_{nj}}(t) - \left(1 - t^2\sigma_{nj}^2/2\right) \right| \to 0$$

and

$$\lim_{n\to\infty} \max_{j\le k_n} \frac{\sigma_{nj}^2}{\sigma_n^2} = 0. \qquad (6)$$

Proof (continued)

This implies that $1 - t^2\sigma_{nj}^2$ are all between 0 and 1 for large enough n. Using the inequality

$$|a_1 \cdots a_m - b_1 \cdots b_m| \le \sum_{j=1}^{m} |a_j - b_j|$$

for any complex numbers a_j's and b_j's with $|a_j| \le 1$ and $|b_j| \le 1$, $j = 1, ..., m$, we obtain that

$$\left| \prod_{j=1}^{k_n} e^{-t^2\sigma_{nj}^2/2} - \prod_{j=1}^{k_n} \left(1 - t^2\sigma_{nj}^2/2\right) \right| \le \sum_{j=1}^{k_n} \left| e^{-t^2\sigma_{nj}^2/2} - \left(1 - t^2\sigma_{nj}^2/2\right) \right|,$$

which is bounded by

$$t^4 \sum_{j=1}^{k_n} \sigma_{nj}^4 \le t^4 \max_{j \le k_n} \sigma_{nj}^2 \to 0,$$

since $|e^x - 1 - x| \le x^2/2$ if $|x| \le \frac{1}{2}$ and $\sum_{j=1}^{k_n} \sigma_{nj}^2 = \sigma_n^2 = 1$.

Proof (continued)

Then

$$\begin{aligned}\left|\prod_{j=1}^{k_n}\phi_{X_{nj}}(t)-\prod_{j=1}^{k_n}e^{-t^2\sigma_{nj}^2/2}\right| &\leq \sum_{j=1}^{k_n}\left|\phi_{X_{nj}}(t)-e^{-t^2\sigma_{nj}^2/2}\right| \\ &\leq \sum_{j=1}^{k_n}\left|\phi_{X_{nj}}(t)-\left(1-t^2\sigma_{nj}^2/2\right)\right| \\ &\quad +\sum_{j=1}^{k_n}\left|e^{-t^2\sigma_{nj}^2/2}-\left(1-t^2\sigma_{nj}^2/2\right)\right| \\ &\to 0\end{aligned}$$

as previously shown.

Thus,

$$\prod_{j=1}^{k_n}\phi_{X_{nj}}(t)=\prod_{j=1}^{k_n}e^{-t^2\sigma_{nj}^2/2}+o(1)=e^{-t^2/2}+o(1)$$

i.e., the ch.f. of $\sum_{j=1}^{k_n} X_{nj}$ converges to the ch.f. of $N(0,1)$ for every t.
By Theorem 1.9(ii), the result follows.

Remarks

- Condition (5) is called Lindeberg's condition.
- From the proof, Lindeberg's condition implies (6), which is called Feller's condition.
- Feller's condition (6) means that all terms in the sum $\sigma_n^2 = \sum_{j=1}^{k_n} \sigma_{nj}^2$ are uniformly negligible as $n \to \infty$.
- If Feller's condition is assumed, then Lindeberg's condition is not only sufficient but also necessary for the result in Theorem 1.15, which is the well-known Lindeberg-Feller CLT.
- A proof can be found in Billingsley (1995, pp. 359-361).
- Note that neither Lindeberg's condition nor Feller's condition is necessary for the result in Theorem 1.15 (Exercise 158).

Liapounov's condition

A sufficient condition for Lindeberg's condition is the following Liapounov's condition, which is somewhat easier to verify:

$$\frac{1}{\sigma_n^{2+\delta}} \sum_{j=1}^{k_n} E|X_{nj} - EX_{nj}|^{2+\delta} \to 0 \quad \text{for some } \delta > 0. \tag{7}$$

Example 1.33

Let $X_1, X_2, \ldots$ be independent random variables.
Suppose that X_i has the binomial distribution $Bi(p_i, 1)$, $i = 1, 2, \ldots$, and that $\sigma_n^2 = \sum_{i=1}^n \text{var}(X_i) = \sum_{i=1}^n p_i(1-p_i) \to \infty$ as $n \to \infty$.
For each i, $EX_i = p_i$ and
$E|X_i - EX_i|^3 = (1-p_i)^3 p_i + p_i^3(1-p_i) \le 2p_i(1-p_i)$.
Hence $\sum_{i=1}^n E|X_i - EX_i|^3 \le 2\sigma_n^2$, i.e., Liapounov's condition (7) holds with $\delta = 1$.
Thus, by Theorem 1.15,

$$\frac{1}{\sigma_n} \sum_{i=1}^n (X_i - p_i) \to_d N(0,1). \tag{8}$$

It can be shown (exercise) that the condition $\sigma_n \to \infty$ is also necessary for result (8).

The following are useful corollaries of Theorem 1.15 and Theorem 1.9(iii).

Corollary 1.2 (Multivariate CLT)

For i.i.d. random k-vectors $X_1,...,X_n$ with a finite $\Sigma = \text{var}(X_1)$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - EX_1) \to_d N_k(0,\Sigma).$$

Corollary 1.3

Let $X_{ni} \in \mathscr{R}^{m_i}$, $i = 1,...,k_n$, be independent random vectors with $m_i \le m$ (a fixed integer), $n = 1,2,...$, $k_n \to \infty$ as $n \to \infty$, and $\inf_{i,n}\lambda_-[\text{var}(X_{ni})] > 0$, where $\lambda_-[A]$ is the smallest eigenvalue of A.
Let $c_{ni} \in \mathscr{R}^{m_i}$ be vectors such that

$$\lim_{n\to\infty}\left(\max_{1\le i\le k_n}\|c_{ni}\|^2 \Big/ \sum_{i=1}^{k_n}\|c_{ni}\|^2\right) = 0.$$

(i) If $\sup_{i,n} E\|X_{ni}\|^{2+\delta} < \infty$ for some $\delta > 0$, then

$$\sum_{i=1}^{k_n} c_{ni}^{\tau}(X_{ni} - EX_{ni}) \Big/ \left[\sum_{i=1}^{k_n}\text{var}(c_{ni}^{\tau}X_{ni})\right]^{1/2} \to_d N(0,1). \qquad (9)$$

(ii) If whenever $m_i = m_j$, $1 \le i < j \le k_n$, $n = 1,2,...$, X_{ni} and X_{nj} have the same distribution with $E\|X_{ni}\|^2 < \infty$, then (9) holds.

Remarks

- Proving Corollary 1.3 is a good exercise.
- Applications of these corollaries can be found in later chapters.
- More results on the CLT can be found, for example, in Serfling (1980) and Shorack and Wellner (1986).

More on Pólya's theorem

Let Y_n be a sequence of random variables, $\{\mu_n\}$ and $\{\sigma_n\}$ be sequences of real numbers such that $\sigma_n > 0$ for all n, and

$$(Y_n - \mu_n)/\sigma_n \to_d N(0,1).$$

Then, by Proposition 1.16,

$$\lim_{n\to\infty} \sup_x |F_{(Y_n-\mu_n)/\sigma_n}(x) - \Phi(x)| = 0, \quad (10)$$

where Φ is the c.d.f. of $N(0,1)$.

Asymptotic normality

(10) implies that for any sequence of real numbers $\{c_n\}$,

$$\lim_{n\to\infty} \left| P(Y_n \le c_n) - \Phi\left(\frac{c_n - \mu_n}{\sigma_n}\right) \right| = 0,$$

i.e., $P(Y_n \le c_n)$ can be approximated by $\Phi\left(\frac{c_n-\mu_n}{\sigma_n}\right)$, regardless of whether $\{c_n\}$ has a limit.

Since $\Phi\left(\frac{t-\mu_n}{\sigma_n}\right)$ is the c.d.f. of $N(\mu_n, \sigma_n^2)$, Y_n is said to be *asymptotically distributed* as $N(\mu_n, \sigma_n^2)$ or simply *asymptotically normal*.

Examples

- For example, $\sum_{i=1}^{k_n} c_{ni}^{\tau} X_{ni}$ in Corollary 1.3 is asymptotically normal.
- This can be extended to random vectors.
  For example, $\sum_{i=1}^{n} X_i$ in Corollary 1.2 is asymptotically distributed as $N_k(nEX_1, n\Sigma)$.