Stat 709: Mathematical Statistics
Lecture 16

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Lecture 16: 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

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

If

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

then

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

Lecture 16: 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

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

If

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

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 $\sigma_{nj}^2 = \mathrm{var}(X_{nj})$.
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 (1).

Also by condition (1) 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

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

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 (1) is called Lindeberg's condition.
- From the proof, Lindeberg's condition implies (2), which is called Feller's condition.
- Feller's condition (2) 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:

$$\frac{1}{\sigma_n^{2+\delta}} \sum_{j=1}^{k_n} E|X_{nj} - EX_{nj}|^{2+\delta} \to 0 \quad \text{for some } \delta > 0. \tag{3}$$

Remarks

- Condition (1) is called Lindeberg's condition.
- From the proof, Lindeberg's condition implies (2), which is called Feller's condition.
- Feller's condition (2) 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:

$$\frac{1}{\sigma_n^{2+\delta}} \sum_{j=1}^{k_n} E|X_{nj} - EX_{nj}|^{2+\delta} \to 0 \quad \text{for some } \delta > 0. \tag{3}$$

Example 1.33

Let $X_1, X_2, ...$ be independent random variables.
Suppose that X_i has the binomial distribution $Bi(p_i, 1)$, $i = 1, 2, ...$, and that $\sigma_n^2 = \sum_{i=1}^n \mathrm{var}(X_i) = \sum_{i=1}^n p_i(1-p_i) \to \infty$ 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 (3) holds with $\delta = 1$.
Thus, by Theorem 1.15,

$$\frac{1}{\sigma_n} \sum_{i=1}^n (X_i - p_i) \to_d N(0,1). \tag{4}$$

It can be shown (exercise) that the condition $\sigma_n \to \infty$ is also necessary for result (4).

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

Example 1.33

Let $X_1, X_2, ...$ be independent random variables.
Suppose that X_i has the binomial distribution $Bi(p_i, 1)$, $i = 1, 2, ...$, 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 (3) holds with $\delta = 1$.
Thus, by Theorem 1.15,

$$\frac{1}{\sigma_n} \sum_{i=1}^n (X_i - p_i) \to_d N(0,1). \tag{4}$$

It can be shown (exercise) that the condition $\sigma_n \to \infty$ is also necessary for result (4).

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

$$\sum_{i=1}^{k_n} c_{ni}^\tau (X_{ni} - EX_{ni}) \Big/ \left[\sum_{i=1}^{k_n}\text{var}(c_{ni}^\tau X_{ni})\right]^{1/2} \to_d N(0,1). \qquad (5)$$

(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 (5) holds.

Corollary 1.2 (Multivariate CLT)

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

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

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 $\inf_{i,n} \lambda_-[\mathrm{var}(X_{ni})] > 0$, 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

$$\sum_{i=1}^{k_n} c_{ni}^{\tau}(X_{ni} - EX_{ni}) \Big/ \left[\sum_{i=1}^{k_n} \mathrm{var}(c_{ni}^{\tau} X_{ni}) \right]^{1/2} \to_d N(0,1). \qquad (5)$$

(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 (5) 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,

$$\lim_{n\to\infty} \sup_x \left| F_{(Y_n-\mu_n)/\sigma_n}(x) - \Phi(x) \right| = 0, \tag{6}$$

where Φ is the c.d.f. of $N(0,1)$.

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,

$$\lim_{n\to\infty} \sup_x |F_{(Y_n-\mu_n)/\sigma_n}(x) - \Phi(x)| = 0, \qquad (6)$$

where Φ is the c.d.f. of $N(0,1)$.

Asymptotic normality

(6) 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 $N_k(nEX_1, n\Sigma)$.

Asymptotic normality

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

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

i.e., $P(Y_n \leq c_n)$ 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 $N_k(nEX_1, n\Sigma)$.