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 $0 < \sigma_n^2 = \mathrm{Var}(\sum_{j=1}^{k_n} X_{nj}) < \infty$, $n = 1, 2, ...$, and $k_n \to \infty$ as $n \to \infty$. If

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

then

$$\frac{1}{\sigma_n}\sum_{j=1}^{k_n}(X_{nj} - EX_{nj}) \to_d N(0, 1). \tag{2}$$

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 $t \in \mathcal{R}$ be given. From the inequality $|e^{\sqrt{-1}tx} - (1 + \sqrt{-1}tx - t^2x^2/2)| \leq \min\{|tx|^2, |tx|^3\}$, the ch.f. of X_{nj} satisfies

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

where $\sigma_{nj}^2 = \mathrm{Var}(X_{nj})$. For any $\epsilon > 0$, the right-hand side of (3) is bounded by

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

which is bounded by

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

Summing over j and using condition (1), we obtain that

$$\sum_{j=1}^{k_n}\left|\phi_{X_{nj}}(t) - \left(1 - t^2\sigma_{nj}^2/2\right)\right| \to 0. \tag{4}$$

By condition (1), $\max_{j\leq k_n}\sigma_{nj}^2 \leq \epsilon^2 + \max_{j\leq k_n} E(X_{nj}^2 I_{\{|X_{nj}|>\epsilon\}}) \to \epsilon^2$ for arbitrary $\epsilon > 0$. Hence

$$\lim_{n\to\infty}\max_{j\leq k_n}\frac{\sigma_{nj}^2}{\sigma_n^2} = 0. \tag{5}$$

(Note that $\sigma_n^2 = 1$ is assumed for convenience.) This implies that $1 - t^2\sigma_{nj}^2$ are all between 0 and 1 for large enough n. Using the inequality

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

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

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

which is bounded by $t^4 \sum_{j=1}^{k_n} \sigma_{nj}^4 \leq t^4 \max_{j \leq k_n} \sigma_{nj}^2 \to 0$, since $|e^x - 1 - x| \leq x^2/2$ if $|x| \leq \frac{1}{2}$ and $\sum_{j=1}^{k_n} \sigma_{nj}^2 = \sigma_n^2 = 1$. Also,

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

is bounded by the quantity on the left-hand side of (4) and, hence, converges to 0 by (4). Thus,

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

This shows that 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.

Condition (1) is called Lindeberg's condition.
From the proof, Lindeberg's condition implies (5), which is called Feller's condition.
Feller's condition (5) 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 result (2), which is the well-known Lindeberg-Feller CLT.
A proof can be found in Billingsley (1986, pp. 373-375).
Note that neither Lindeberg's condition nor Feller's condition is necessary for result (2) (Exercise 158).

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

$$\frac{1}{\sigma_n^{2+\delta}} \sum_{j=1}^{k_n} E|X_{nj} - EX_{nj}|^{2+\delta} \to 0 \quad \text{for some } \delta > 0. \tag{6}$$

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 $E|X_i - EX_i|^3 = (1 - p_i)^3 p_i + p_i^3(1 - p_i) \leq 2p_i(1 - p_i)$. Hence $\sum_{i=1}^n E|X_i - EX_i|^3 \leq 2\sigma_n^2$, i.e., Liapounov's condition (6) holds with $\delta = 1$. Thus, by Theorem 1.15,

$$\frac{1}{\sigma_n} \sum_{i=1}^n (X_i - p_i) \to_d N(0, 1). \tag{7}$$

It can be shown (exercise) that the condition $\sigma_n \to \infty$ is also necessary for result (7).

Useful corollaries of Theorem 1.15 (and Theorem 1.9(iii))

Corollary 1.2 (Multivariate CLT). Let $X_1, ..., X_n$ be i.i.d. random k-vectors with a finite $\Sigma = \mathrm{Var}(X_1)$. Then

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

Corollary 1.3. Let $X_{ni} \in \mathcal{R}^{m_i}$, $i = 1, ..., k_n$, be independent random vectors with $m_i \leq m$ (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 $c_{ni} \in \mathcal{R}^{m_i}$ be vectors such that

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

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

$$\sum_{i=1}^{k_n} c_{ni}^\tau (X_{ni} - EX_{ni}) \Big/ \left[\sum_{i=1}^{k_n} \mathrm{Var}(c_{ni}^\tau X_{ni}) \right]^{1/2} \to_d N(0, 1). \qquad (8)$$

(ii) Suppose that whenever $m_i = m_j$, $1 \leq i < j \leq k_n$, $n = 1, 2, ...$, X_{ni} and X_{nj} have the same distribution with $E\|X_{ni}\|^2 < \infty$. Then (8) holds.

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).

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,

$$\lim_{n\to\infty} \sup_x |F_{(Y_n-\mu_n)/\sigma_n}(x) - \Phi(x)| = 0, \qquad (9)$$

where Φ is the c.d.f. of $N(0, 1)$.
This implies that for any sequence of real numbers $\{c_n\}$, $\lim_{n\to\infty} |P(Y_n \leq c_n) - \Phi(\frac{c_n-\mu_n}{\sigma_n})| = 0$, i.e., $P(Y_n \leq c_n)$ can be approximated by $\Phi(\frac{c_n-\mu_n}{\sigma_n})$, regardless of whether $\{c_n\}$ has a limit.
Since $\Phi(\frac{t-\mu_n}{\sigma_n})$ 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*.
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 $N_k(nEX_1, n\Sigma)$.