§ 2.2. Central limit theorem.

The most ideal case of the CLT is that the random variables are iid with finite variance. Although it is a special case of the more general Lindeberg-Feller CLT, it is most standard and its proof contains the essential ingredients to establish more general CLT. Throughout the chapter, $\Phi(\cdot)$ is the cdf of standard normal distribution $N(0,1)$.

(i). Central limit theorem (CLT) for iid r.v.s.

The following lemma plays a key role in the proof of CLT.

Lemma 2.1 For any real x and $n \geq 1$,

$$|e^{ix} - \sum_{j=0}^{n} \frac{(ix)^j}{j!}| \leq \min\Big(\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\Big).$$

Consequently, for any r.v. X with characteristic function ψ and finite second moment,

$$\Big|\psi(t) - [1 + itE(X) - \frac{t^2}{2}E(X^2)]\Big| \leq \frac{|t|^2}{6}E(\min(|t||X|^3,\ 6|X|^2)). \tag{2.1}$$

Proof. The proof relies on the identity

$$e^{ix} - \sum_{j=0}^{n} \frac{(ix)^j}{j!} = \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is} ds = \frac{i^n}{(n-1)!}\int_0^x (x-s)^{n-1}(e^{is}-1)ds,$$

which can be shown by induction and by taking derivatives. The middle term is bounded by $|x|^{n+1}/(n+1)!$, and the last bounded by $2|x|^n/n!$. □

Theorem 2.2 *Suppose $X, X_1, ..., X_n, ...$ are iid with mean μ and finite variance $\sigma^2 > 0$. Then,*

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \to N(0,1) \qquad \textit{in distribution.}$$

Proof. Without loss of generality, let $\mu = 0$. Let ψ be the common characteristic function of X_i. Observe that, by dominated convergence

$$E(\min(|t_n||X|^3,\ 6|X|^2)) \to 0 \quad \text{as} \quad |t_n| \to 0$$

The characteristic function of $S_n/\sqrt{n\sigma^2}$ is, by applying the above lemma,

$$\begin{aligned} E(e^{itS_n/\sqrt{n\sigma^2}}) &= E(e^{it_nS_n}) = \prod_{j=1}^{n} E(e^{itX_j/\sqrt{n\sigma^2}}) = \psi^n(\frac{t}{\sqrt{n\sigma^2}}) \\ &= [1 + \frac{it}{\sqrt{n\sigma^2}}E(X) - \frac{t^2}{2n\sigma^2}E(X^2) + o(\frac{1}{n})]^n = [1 - \frac{t^2}{2n} + o(\frac{1}{n})]^n \\ &\to e^{-t^2/2}, \end{aligned}$$

which is the characteristic function of $N(0,1)$. Then, Levy's continuity theorem implies the above CLT. □

In the case the common variance is not finite, the partial sum, after proper normalization, may or may not converge to a normal distribution. The following theorem provides sufficient and necessary condition. The key point here is whether there exists appropriate truncation, which is a trick that we have used so many times before.

Theorem 2.3 *Suppose $X, X_1, X_2, ...$ are iid nondegenerate. Then, $(S_n - a_n)/b_n$ converges to a normal distribution for some constants a_n and $0 < b_n \to \infty$, if and only if*

$$\frac{x^2 P(|X| > x)}{E(X^2 1_{\{|X| \le x\}})} \to 0, \qquad \text{as } x \to \infty. \tag{2.2}$$

The proof is omitted. We note that (2.2) holds if X_i has finite variance $\sigma^2 > 0$, in which case CLT of Theorem 2.2 holds with $a_n = nE(X)$ and $b_n = \sqrt{n}\sigma$. Theorem 2.3 is of interest when $E(X^2) = \infty$. In this case, one can choose to truncate the X_is at

$$c_n = \sup\{c : nE(|X|^2 1_{\{|X| \le c\}})/c^2 \ge 1\}$$

With some calculation, condition (2.2) ensures

$$nP(|X| > c_n) \to 0 \qquad \text{and} \qquad nE(|X|^2 1_{\{|X| \le c_n\}})/c_n^2 \to 1.$$

Separate S_n into two parts, one with X_i beyond $\pm c_n$ and the other bounded by $\pm c_n$. The former takes value 0 with chance going to 1. The latter, when standardized by

$$a_n = nE(X 1_{\{|X| \le c_n\}}) \qquad \text{and} \qquad b_n = \sqrt{nE(X^2 1_{\{|X| \le c_n\}})} \approx c_n.$$

converges to $N(0, 1)$, which can be shown by repeating the proof of Theorem 2.2 or by citing Lindeberg-Feller CLT. We note that $b_n \approx \sqrt{n\mathrm{var}(X 1_{\{|X| \le c_n\}})}$ by (2.2).

Example 2.5 Recall Example 1.13, in which $X, X_1, X_2, ...$ are iid symmetric such that $P(|X| > x) = x^{-\alpha}$ for some $\alpha > 0$ all large x. Then, Theorem 2.3 implies $(S_n - a_n)/b_n \to N(0, 1)$ if and only if $\alpha \ge 2$. Indeed, when $\alpha > 2$, the common variance is finite and CLT applies. When $\alpha = 2$,

$$S_n/(n \log n)^{1/2} \to N(0, \sigma^2)$$

for some σ^2.

When $\alpha < 2$, the condition in Theorem cannot hold. As to be seen in Section 1.3, S_n when properly normalized shall converge to non-normal distribution.

(ii). *The Lindeberg-Feller CLT.*

Theorem 2.4 Lindeberg-Feller CLT. *Suppose $X_1, ..., X_n, ...$ are independent r.v.s with mean 0 and variance σ_n^2. Let $s_n^2 = \sum_{j=1}^n \sigma_j^2$ denote the variance of partial sum $S_n = X_1 + \cdots + X_n$. If, for every $\epsilon > 0$,*

$$\frac{1}{s_n^2} \sum_{j=1}^n E(X_j^2 1_{\{|X_j| > \epsilon s_n\}}) \to 0, \tag{2.3}$$

then $S_n/s_n \to N(0, 1)$. Conversely, if $\max_{j \le n} \sigma_j^2/s_n^2 \to 0$ and $S_n/s_n \to N(0, 1)$, then (2.3) holds.

Proof. "$\Longleftarrow$" The Lindeberg condition (2.3) implies

$$\max_{1 \le j \le n} \left(\frac{\sigma_j^2}{s_n^2}\right) \le \epsilon^2 + \frac{1}{s_n^2} \max_{1 \le j \le n} E(X_j^2 1_{\{|X_j| > \epsilon s_n\}}) \to 0, \tag{2.4}$$

by letting $n \to \infty$ and then $\epsilon \downarrow 0$. Observe that for every real $x > 0$, $|e^{-x} - 1 + x| \le x^2/2$. Moreover, for complex z_j and w_j with $|z_j| \le 1$ and $|w_j| \le 1$,

$$|\prod_{j=1}^n z_j - \prod_{j=1}^n w_j| \le \sum_{j=1}^n |z_j - w_j|, \tag{2.5}$$

which can be proved by induction. With Lemma 2.1, it follows that, for any $\epsilon > 0$,

$$\begin{aligned}
& |E(e^{itX_j/s_n}) - e^{-t^2\sigma_j^2/2s_n^2}| \\
\leq\ & |E\Big(1 + itX_j - \frac{(tX_j)^2}{2s_n^2}\Big) - \Big(1 - \frac{t^2\sigma_j^2}{2s_n^2}\Big)| + E\Big[\min\Big(\frac{t^2X_j^2}{s_n^2}, \frac{|tX_j|^3}{6s_n^3}\Big)\Big] + \frac{t^4\sigma_j^4}{8s_n^4} \\
\leq\ & E\Big(\frac{t^2X_j^2}{s_n^2}1_{\{|X_j|>\epsilon s_n\}}\Big) + E\Big(\frac{|tX_j|^3}{6s_n^3}1_{\{|X_j|\leq\epsilon s_n\}}\Big) + \frac{t^4\sigma_j^4}{8s_n^4} \\
\leq\ & \frac{t^2}{s_n^2}E(X_j^2 1_{\{|X_j|>\epsilon s_n\}}) + \frac{|t|^3\epsilon}{s_n^2}E(X_j^2) + \frac{t^4\sigma_j^2}{s_n^2}\max_{1\leq k\leq n}\frac{\sigma_k^2}{s_n^2}
\end{aligned}$$

Then, for any fixed t,

$$\begin{aligned}
& |E(e^{itS_n/s_n}) - e^{-t^2/2}| \\
=\ & |\prod_{j=1}^n E(e^{itX_j/s_n}) - \prod_{j=1}^n e^{-t^2\sigma_j^2/2s_n^2}| \\
\leq\ & \sum_{j=1}^n |E(e^{itX_j/s_n}) - e^{-t^2\sigma_j^2/2s_n^2}| \qquad \text{by (2.5)} \\
\leq\ & \sum_{j=1}^n \Big(\frac{t^2}{s_n^2}E(X_j^2 1_{\{|X_j|>\epsilon s_n\}}) + \frac{|t|^3\epsilon}{s_n^2}E(X_j^2) + \frac{t^4\sigma_j^2}{s_n^2}\max_{1\leq j\leq n}\frac{\sigma_j^2}{s_n^2}\Big) \\
\leq\ & \Big(\frac{t^2}{s_n^2}\sum_{j=1}^n E(X_j^2 1_{\{|X_j|>\epsilon s_n\}}) + \epsilon|t|^3 + t^4\max_{1\leq j\leq n}\frac{\sigma_j^2}{s_n^2}\Big) \\
\to\ & \epsilon|t|^3, \qquad \text{as } n\to\infty, \qquad \text{by (2.3) and (2.4).}
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, it follows that $E(e^{itS_n/s_n}) \to e^{-t^2/2}$ for all t. Levy's continuity theorem implies $S_n/s_n \to N(0,1)$.

"$\Longleftarrow$" Let ψ_j be the moment generating function of X_j. The asymptotic normality is equivalent to $\prod_{j=1}^n \psi_j(t/s_n) \to e^{-t^2/2}$. Notice that (2.1) implies

$$|\psi_j(t/s_n) - 1| \leq 2\frac{t^2\sigma_j^2}{s_n} \tag{2.6}$$

Write, as $n\to\infty$,

$$\begin{aligned}
& \sum_{j=1}^n [\psi_j(t/s_n) - 1] + t^2/2 \\
=\ & \sum_{j=1}^n [\psi_j(t/s_n) - 1 - \log\psi_j(t/s_n)] + \sum_{j=1}^n [\log\psi_j(t/s_n)] + t^2/2 \\
\leq\ & \sum_{j=1}^n |\psi_j(t/s_n) - 1 - \log\psi_j(t/s_n)| + +o(1) \\
\leq\ & \sum_{j=1}^n |\psi_j(t/s_n) - 1|^2 + o(1) \\
\leq\ & \max_{1\leq k\leq n}|\psi_k(t/s_n) - 1| \times \sum_{j=1}^n |\psi_j(t/s_n) - 1| + o(1) \\
\leq\ & 4\max_{1\leq k\leq n}\frac{t^2\sigma_k^2}{s_n} \times \sum_{j=1}^n \frac{t^2\sigma_j^2}{s_n} + o(1) \qquad \text{by (2.6)} \\
=\ & o(1), \qquad \text{by the assumption } \max_{j\leq n}\sigma_j^2/s_n^2 \to 0.
\end{aligned}$$

On the other hand, by definition of characteristic function, the above expression is, as $n \to \infty$,

$$
\begin{aligned}
o(1) &= \sum_{j=1}^{n}[\psi_j(t/s_n) - 1] + t^2/2 \\
&= \sum_{j=1}^{n} E(e^{itX_j/s_n} - 1) + t^2/2 = \sum_{j=1}^{n} E(cos(tX_j/s_n) - 1) + t^2/2 + i\sum_{j=1}^{n} E(sin(tX_j/s_n)) \\
&= \sum_{j=1}^{n} E\{(cos(tX_j/s_n) - 1)1_{\{|X_j|>\epsilon s_n\}}\} + \sum_{j=1}^{n} E\{(cos(tX_j/s_n) - 1)1_{\{|X_j|\le\epsilon s_n\}}\} + t^2/2 \\
&\qquad +\text{imaginary part (immaterial)}.
\end{aligned}
$$

Since $cos(x) - 1 \geq -x^2/2$ for all real x,

$$
\begin{aligned}
\frac{1}{s_n^2}\sum_{j=1}^{n} E(X_j^2 1_{\{X_j|>\epsilon s_n\}}) &= 1 - \frac{2}{t^2}\sum_{j=1}^{n} E(\frac{t^2 X_j^2}{2s_n^2} 1_{\{X_j|\le\epsilon s_n\}}) \\
&\le \frac{2}{t^2}\Big(\frac{t^2}{2} + \sum_{j=1}^{n} E\{(cos(tX_j/s_n) - 1)1_{\{|X_j|\le\epsilon s_n\}}\}\Big) \\
&\le \frac{2}{t^2}\Big(|\sum_{j=1}^{n} E\{(cos(tX_j/s_n) - 1)1_{\{|X_j|>\epsilon s_n\}}\}| + o(1)\Big) \\
&\le \frac{2}{t^2}\sum_{j=1}^{n} 2P(|X_j| > \epsilon s_n) + o(1) \\
&\le \frac{4}{t^2}\sum_{j=1}^{n} \frac{\sigma_j^2}{(\epsilon s_n)^2} + o(1) \qquad \text{by Chebyshev inequality} \\
&\le \frac{4}{t^2\epsilon^2} + o(1).
\end{aligned}
$$

Since t can be chosen arbitrarily large, Lindeberg condition holds. □

Remark. Sufficiency is proved by Lindeberg in 1922 and necessity by Feller in 1935. Lindeberg-Feller CLT is one of the most far-reaching results in probability theory. Nearly all generalizations of various types of central limit theorems spin from Lindeberg-Feller CLT, such as, for example, CLT for martingales, for renewal proceses, or for weakly dependent processes. The insights of the Lindeberg condition (2.3) are that the "wild" values of the random variables, compared with s_n, the standard deviation of S_n as the normalizing constant, are insignificant and can be truncated off without affecting the general behavior of the partial sum S_n.

Example 2.6. Suppose X_n are independent and

$$P(X_n = n) = P(X_n = -n) = n^{-\alpha}/4 \qquad \text{and} \qquad P(X_n = 0) = 1 - n^{-\alpha}/2,$$

with $0 < \alpha < 3$. Then, $\sigma_n^2 = E(X_n^2) = n^{2-\alpha}/2$ and $s_n^2 = \sum_{j=1}^{n} j^{2-\alpha}/2$, which increases to ∞ at the order of $n^{3-\alpha}$. Note that Lindeberg condition (2.3) is equivalent to $n^2/n^{3-\alpha} \to 0$, i.e., $0 < \alpha < 1$. On the other hand, $\max_{1\le j\le n} \sigma_j^2/s_n^2 \to 0$. Therefore, it follows from Theorem 2.4 that $S_n/s_n \to N(0,1)$ if and only if $0 < \alpha < 1$. □

Example 2.7 Suppose X_n are independent and $P(X_n = 1) = 1/n = 1 - P(X_n = 0)$. Then,

$$[S_n - \log(n)]/\sqrt{\log(n)} \to N(0,1) \qquad \text{in distribution.}$$

It's clear that $E(X_n) = 1/n$ and $\text{var}(X_n) = (1 - 1/n)/n$. So, $E(S_n) = \sum_{i=1}^{n} = \sum_{i=1}^{n} 1/i$, and $\text{var}(S_n) = \sum_{i=1}^{n}(1 - 1/i)/i \approx \log(n)$. As X_n are all bounded by 1 and $\text{var}(S_n) \uparrow \infty$, the Lindeberg condition is satisfied. Therefore, by the CLT,

$$\frac{S_n - \sum_{i=1}^n 1/i}{[\sum_{i=1}^n (1-1/i)/i]^{1/2}} \to N(0,1), \qquad \text{in distribution.}$$

Then, $[S_n - \log(n)]/\sqrt{\log(n)} \to N(0,1)$ in distribution since $|\log(n) - \sum_{i=1}^n 1/i| \le 1$ and $\mathrm{var}(S_n)/\log(n) \to 1$. □

Theorem 2.2 as well as the following Lyapunov CLT are both special cases of the Lindeberg-Feller CLT. Nevertheles they are convenient for application.

Corollary (Lyapunov CLT) *Suppose X_n are indendent with mean 0 and $\sum_{j=1}^n E(|X_j|^\delta)/s_n^\delta \to 0$ for some $\delta > 2$, then $S_n/s_n \to N(0,1)$.*

Proof. For any $\epsilon > 0$, as $n \to \infty$,

$$\frac{1}{s_n^2}\sum_{j=1}^n E(X_j^2 1_{\{|X_j|>\epsilon s_n\}}) = \sum_{j=1}^n E(\frac{X_j^2}{s_n^2} 1_{\{|X_j|/s_n>\epsilon\}}) \le \frac{1}{\epsilon^{\delta-2}}\sum_{j=1}^n E(\frac{X_j^\delta}{s_n^\delta}) \to 0.$$

Lindeberg condition (2.3) holds and hence CLT holds. □

In Example 2.6, for any $\delta > 2$, $\sum_{j=1}^n E|X_j|^\delta = \sum_{j=1}^n j^\delta j^{-\alpha}/2$ which increasing at the order $n^{\delta-\alpha+1}$, while s_n^δ increases at the order of $n^{(3-\alpha)\delta/2}$. Simple calculation shows, when $0 < \alpha < 1$, Lypunov CLT holds.

(iii). *CLT for arrays of random variables.*

Very often Lindeberg-Feller CLT is presented in the form of arrays of random variables as given in the textbook.

Theorem 2.5 (CLT for arrays of r.v.s) *Let $X_{n,1}, ..., X_{n,n}$ be n independent random variables with mean 0 such that, as $n \to \infty$,*

$$\sum_{j=1}^n \mathrm{var}(X_{n,j}) \to 1 \qquad \textit{and} \qquad \sum_{j=1}^n E(X_{n,j}^2 1_{\{|X_{n,j}|>\epsilon\}}) \to 0, \qquad \textit{for any } \epsilon > 0.$$

Then, $S_n \equiv X_{n,1} + \cdots + X_{n,n} \to N(0,1)$.

This theorem is slightly more general than Lindeberg-Feller CLT, although the proof is identical to that of the first part of Theorem 2.4. Theorem 2.4. is a special case of Theorem 2.5 by letting $X_{n,i} = X_i/s_n$. Thus $X_{n,k}$ are undertood as the usual r.v.s normalized by the standard deviation of the partial sums. Thus S_n in this theorem is already standardized.

DIY Exercises

Exercise 2.5 Suppse X_n are independent with

$$P(X_n = n^\alpha) = P(X_n = -n^\alpha) = \frac{1}{2n^\beta} \qquad \text{and} \qquad P(X_n = 0) = 1 - \frac{1}{n^\beta}$$

with $2\alpha > \beta - 1$. Show that the Lindeberg condition holds if and only if $0 \le \beta < 1$.

Exercise 2.6 Suppose X_n are iid with mean 0 and variance 1. Let $a_n > 0$ be such that $s_n^2 = \sum_{j=1}^n a_i^2 \to \infty$ and $a_n/s_n \to 0$. Show that $\sum_{i=1}^n a_i X_i/s_n \to N(0,1)$.

Exercise 2.7 Suppose $X_1, X_2...$ are independent and $X_n = Y_n + Z_n$, where Y_n takes values 1 and -1 with chance 1/2 each, and $P(Z_n = \pm n) = 1/(2n^2) = (1 - P(Z_n = 0))/2$ Show that Lindeberg condition does not hold, yet $S_n/\sqrt{n} \to N(0,1)$.

Exercise 2.8 Suppse $X_1, X_2, ...$ are iid nonnegative r.v.s with mean 1 and finite variance $\sigma^2 > 0$. Show that $2(\sqrt{S_n} - \sqrt{n}) \to N(0,1)$.