A proof of the De Moivre–Laplace Central Limit Theorem

Alex Fu

2023-01-03

Theorem 1 (De Moivre–Laplace Central Limit Theorem, or CLT).

Let $X_1, X_2, \dots$ be i.i.d. Rademacher random variables, which are ± 1 with probability $\frac{1}{2}$ each. Then $\frac{1}{\sqrt{n}} S_n \coloneqq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ converges in distribution to a standard normal $Z \sim \mathcal{N}(0, 1)$.

The CLT gets its name as a limit theorem central to probability and statistics, not because it is a theorem about central limits. The De Moivre–Laplace CLT was historically the first CLT to be shown. The sums $(S_n)_{n \in \mathbb{N}}$ above form a *simple random walk* on the integers $\mathbb{Z}$ — one of the first random processes to be studied in probability theory.

Lemma 1 (Standard normal PDF).

The probability density function of a standard normal random variable $Z \sim \mathcal{N}(0, 1)$ is

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Lemma 2 (Stirling's approximation).

The following quantities are asymptotic: their ratio tends to 1 as $n \to \infty$.

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$

Lemma 3 (Continuization in distribution).

Let $(Z_n)_{n=1}^{\infty}$ be a sequence of $\mathbb{Z}$-valued random variables, $\delta_n \downarrow 0$ a sequence of positive constants in $\mathbb{R}$, and f a probability density function. If for all x in a Lebesgue-almost sure set,

$$\mathbb{P}(Z_n = z_n)/\delta_n \to f(x)$$

for any sequence of integers $(z_n)_{n=1}^{\infty}$ for which $z_n \delta_n \to x$, then $Z_n \delta_n$ converges in distribution to a random variable X with density f.

Lemma 4 (Generalized definition of e).

Let $x_n \to 0$ and $y_n \to \infty$ be such that $x_n y_n \to c$. Then $(1 + x_n)^{y_n} \to e^c$.

Lemma 5 (Slutsky's theorem).

Suppose that $a_n \to a$, $b_n \to b$, and $X_n \xrightarrow{\text{d}} X$. Then $a_n X_n + b_n$ converges in distribution to $aX + b$.

We will omit the proofs of the lemmas above.

Proof of Theorem 1. We first observe that S_n and n have the same *parity*, as $S_0 = 0$ and $S_{n+1} = S_n \pm 1$. Let us work with the distribution of S_{2n}. By counting the number of possible configurations,

$$\begin{aligned}\mathbb{P}(S_{2n} = 2k) &= \binom{2n}{n-k} 2^{-2n} \\ &= \frac{(2n)!}{(n+k)!(n-k)!} 2^{-2n}\end{aligned}$$

By Lemma 2 (Stirling's approximation),

$$\begin{aligned}&\sim \frac{(2n)^{2n} e^{-2n} \sqrt{2\pi \cdot 2n}}{(n+k)^{n+k} e^{-n-k} \sqrt{2\pi(n+k)} \cdot (n-k)^{n-k} e^{-n+k} \sqrt{2\pi(n-k)}} \cdot 2^{-2n} \\ &= \frac{1}{\sqrt{2\pi}} \frac{n^{2n} \sqrt{2n}}{(n+k)^{n+k}(n-k)^{n-k}} \frac{1}{\sqrt{(n+k)(n-k)}} \\ &= \frac{1}{\sqrt{\pi n}} \frac{1}{(1+\frac{k}{n})^{n+k}(1-\frac{k}{n})^{n-k}} \frac{1}{\sqrt{(1+\frac{k}{n})(1-\frac{k}{n})}}\end{aligned}$$

Stirling's approximation above requires $n - k \to \infty$, so let us choose $k = x\sqrt{\frac{n}{2}}$ so that $\frac{2k}{\sqrt{2n}} \to x$, with the intention of invoking Lemma 3.

$$\begin{aligned}&= \frac{1}{\sqrt{\pi n}} \frac{(1-\frac{k}{n})^k}{(1-\frac{k^2}{n^2})^{n+\frac{1}{2}}(1+\frac{k}{n})^k} \\ &\sim \frac{1}{\sqrt{\pi n}} \left(1 - \frac{x^2}{2n}\right)^{-n} \left(1 - \frac{x}{\sqrt{2n}}\right)^{x\sqrt{2n}} \left(1 + \frac{x}{\sqrt{2n}}\right)^{-x\sqrt{2n}}\end{aligned}$$

By Lemma 4, as $n \to \infty$, this tends to

$$= \frac{1}{\sqrt{\pi n}} e^{-x^2/2}.$$

Invoking Lemma 3 for $Z_n = \frac{1}{2}S_{2n}$ and $\delta_n = \frac{2}{\sqrt{2n}}$, we find that S_{2n} converges in distribution to $\mathcal{N}(0, 1)$. To extend this result to the full sequence $(S_n)_{n=1}^{\infty}$, we write

$$\frac{S_{2n+1}}{\sqrt{2n+1}} = \frac{\sqrt{2n}}{\sqrt{2n+1}} \cdot \frac{S_{2n}}{\sqrt{2n}} + \frac{X_{2n+1}}{\sqrt{2n+1}}$$

and use Lemma 5 to conclude our proof. □

This proof was quite combinatorial in nature: it leveraged the symmetry in the situation to find $\mathbb{P}(S_{2n} = 2k)$ by counting, then related it to $\Phi(x)$ by asymptotic analysis, finishing with the key Lemma 3. A more general CLT for i.i.d. random variables with finite variance can be proven using the method of Fourier transforms or **characteristic functions** $\phi_X(t) = \mathbb{E}(e^{itX})$. Further generalizations of the CLT to weakly dependent random variables make use of other techniques, such as the *Lindeberg exchange trick*, which we will not introduce here.

This note was adapted largely from Professor Shirshendu Ganguly's fall 2022 offering of Math C218A / Stat C205A at UC Berkeley.

■