The Central Limit Theorem

As its name indicates, the Central Limit Theorem is one of the most important statements in mathematics. It shows that the Gaussian distribution is the universal attractor in fluctuations of a large number of random elements, sometimes called the “law of errors”.

It appears historically that the first central limit theorem was put forward by A. de Moivre for Bernoulli random variables in 1733 (cf. [4]), introducing at the same time the normal law, which thus should be called *de Moivrian*! (Note also that A. de Moivre was close to 60 at the time of the result, and thus would not have got the Fields Medal although he would have fully deserved it.)

The central limit theorem is extensively presented in all standard textbooks on probability theory or statistics. The post here is focused on its most classical form for independent identically distributed summands. The bibliography is limited to a few specific pointers.

Table of contents

1. The classical one-dimensional central limit theorem

2. A. de Moivre

3. Fourier analytic proof

4. The Lindeberg method

5. The Berry-Esseen inequality

Drafted by M. L. **v1** May 2024

6. The multivariate central limit theorem

7. Unlimited extensions...

References

1 The classical one-dimensional central limit theorem

On a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, let X_k, $k \geq 1$, be a sequence of independent and identically distributed random variables, with the same law as a real random variable X. Set $S_n = X_1 + \cdots + X_n$, $n \geq 1$. The standard law of large numbers expresses that if $\mathbb{E}(|X|) < \infty$, then

$$\frac{S_n}{n} \to \mathbb{E}(X) \quad \text{almost surely.}$$

(Conversely, if the sequence $\left(\frac{S_n}{n}\right)_{n\geq 1}$ is almost surely bounded, then $\mathbb{E}(|X|) < \infty$.) What is then the rate of convergence to 0 of the sequence $\frac{S_n}{n} - \mathbb{E}(X)$, $n \geq 1$? Under suitable moment hypotheses, the sequence $n^\beta\left(\frac{S_n}{n} - \mathbb{E}(X)\right)$ still converges almost surely provided that $\beta < \frac{1}{2}$ When $\beta = \frac{1}{2}$, the convergence mode changes, and a limiting de Moivrian distribution arises.

Theorem 1 (Central limit theorem)**.** *Under the preceding notation, if* $0 < \mathbb{E}(X^2) < \infty$*,*

$$\sqrt{n}\left[\frac{S_n}{n} - \mathbb{E}(X)\right] \to G \quad \textit{in distribution}$$

where G is a de Moivrian random variable with law $\mathcal{N}(0, \sigma^2)$ where $\sigma^2 = \mathrm{Var}(X)$.

(Conversely, it may be shown that if the sequence $\left(\frac{S_n}{\sqrt{n}}\right)_{n\geq 1}$ is stochastically bounded, then $\mathbb{E}(X^2) < \infty$, and $\mathbb{E}(X) = 0$.)

The central limit theorem is in particular a stability property of the variance, and de Moivrian variables are a fixed point of the statement: if X_k, $k \geq 1$, are independent with law $\mathcal{N}(0, \sigma^2)$ (thus centered for simplicity), then $\frac{S_n}{\sqrt{n}}$ has law $\mathcal{N}(0, \sigma^2)$ for every $n \geq 1$.

Note for the formulation that

$$\sqrt{n}\left[\frac{S_n}{n} - \mathbb{E}(X)\right] = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \big[X_k - \mathbb{E}(X_k)\big].$$

so that it is often convenient to center the variables. Also, it is useful to normalize the variance so to converge towards a standard normal $\mathcal{N}(0, 1)$.

In terms of distribution functions, the weak convergence in Theorem 1 thus indicates that, for every $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\bigg(\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}\big[X_k - \mathbb{E}(X_k)\big] \leq t\bigg) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}x^2}dx. \tag{1}$$

It is important to note that the weak convergence property in the central limit theorem cannot be strengthened into a stronger convergence mode. For simplicity, let X be such that $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^2) = 1$ so that $\frac{S_n}{\sqrt{n}}$ converges, by Theorem 1, to a de Moivrian variable G with law $\mathcal{N}(0,1)$. For every $n \geq 1$,

$$\sqrt{2}\,\frac{S_{2n}}{\sqrt{2n}} - \frac{S_n}{\sqrt{n}} = \frac{X_{n+1} + \cdots + X_{2n}}{\sqrt{n}}\,.$$

If the sequence $\frac{S_n}{\sqrt{n}}$, $n \geq 1$, would converge for example in probability to G, the left-hand side in this inequality would converge to $(\sqrt{2}-1)G$. On the other hand, the right-hand has the same law as $\frac{S_n}{\sqrt{n}}$, and thus converges weakly to G. Since G is de Moivrian, $(\sqrt{2}-1)G$ cannot have the same law than G.

2 A. de Moivre

As mentioned in the introduction, the first central limit theorem was put forward by A. de Moivre back in 1733 for sequences of Bernoulli random variables. If X is a Bernoulli random variable on $\{0,1\}$, with parameter $p \in]0,1[$, $\mathbb{E}(X) = p$ and $\mathrm{Var}(X) = p(1-p)$, so that the statement (1) takes the form

$$\lim_{n\to\infty} \mathbb{P}\bigg(\frac{1}{\sqrt{p(1-p)n}}\sum_{k=1}^{n}[X_k - p] \leq t\bigg) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}x^2}dx$$

for every $t \in \mathbb{R}$. Alternatively,

$$\lim_{n\to\infty} \mathbb{P}\bigg(\sum_{k=1}^{n} X_k \leq pn + t\sqrt{p(1-p)n}\bigg) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}x^2}dx.$$

But $\sum_{k=1}^{n} X_k$ has binomial distribution $\mathcal{B}(n,p)$, so that the preceding limit explicitly expresses that

$$\lim_{n\to\infty} \sum_{k\in I_n} \binom{n}{k} p^k(1-p)^{n-k} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}x^2}dx$$

where I_n is the set of integers k in $\{0,1,\ldots,n\}$ such that $k \leq pn + t\sqrt{p(1-p)n}$.

The proof by A. de Moivre for $p = \frac{1}{2}$ (the case of arbitrary $p \in (0,1)$ was settled later by P.-S. de Laplace, by other means) is then based on the equivalence

$$\binom{n}{k} p^k (1-p)^{n-k} \sim \frac{1}{\sqrt{2\pi np(1-p)}}\, e^{-\frac{(k-np)^2}{2p(1-p)n}}$$

as a consequence of Stirling's formula (just proved a few years before in 1730), justifying (a posteriori) the factor $\sqrt{2\pi}$ of the density of the normal law.

3 Fourier analytic proof

The standard proof of Theorem 1, presented in most textbooks, uses the Fourier transform and the Paul Lévy theorem. Assume, after translation and homothety, that $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^2) = 1$. By independence and identical distribution, the characteristic function of $\frac{S_n}{\sqrt{n}}$ is given by

$$\varphi_{\frac{S_n}{\sqrt{n}}}(u) = \mathbb{E}\big(e^{iu\frac{S_n}{\sqrt{n}}}\big) = \big[\mathbb{E}\big(e^{i\frac{u}{\sqrt{n}}X}\big)\big]^n = \big[\varphi_X\big(\tfrac{u}{\sqrt{n}}\big)\big]^n$$

for every $u \in \mathbb{R}$, where $\varphi_X(u) = \mathbb{E}(e^{iuX})$ is the characteristic function of the law of X. Since $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^2) = 1$, φ_X is twice differentiable, and $\varphi_X(0) = 1$, $\varphi_X'(0) = 0$, $\varphi_X''(0) = -1$. Hence, a Taylor expansion around 0 expresses that

$$\varphi_X(v) = 1 - \tfrac{1}{2}\, v^2 + v^2 \varepsilon(v)$$

where $\varepsilon(v) \in \mathbb{C}$ tends to 0 as $v \to 0$. Therefore, for $u \in \mathbb{R}$ fixed, as $n \to \infty$,

$$\varphi_{\frac{S_n}{\sqrt{n}}}(u) = \Big(1 - \tfrac{1}{2n}\, u^2 + \tfrac{1}{n}\, u^2\, \varepsilon\big(\tfrac{u}{\sqrt{n}}\big)\Big)^n.$$

It is easy to check that for every sequence z_n, $n \in \mathbb{N}$, of complex numbers converging to some $z \in \mathbb{C}$,

$$\lim_{n\to\infty} \Big(1 + \frac{z_n}{n}\Big)^n = e^z.$$

This observation, applied, for $u \in \mathbb{R}$ fixed, to $z_n = -\frac{1}{2}u^2 + \varepsilon(\frac{u}{\sqrt{n}})$, $n \geq 1$, which converges to $z = -\frac{1}{2}u^2$, ensures that

$$\lim_{n\to\infty} \varphi_{\frac{S_n}{\sqrt{n}}}(u) = e^{-\frac{1}{2}u^2}.$$

Since the right-hand side is precisely the Fourier transform of the standard normal distribution $\mathcal{N}(0,1)$, the central limit theorem follows.

4 The Lindeberg method

An alternate, most interesting, idea of proof of the central limit theorem is the Lindeberg replacement method. Assuming again that $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^2) = 1$, denote by G_k, $k \geq 1$, a sequence of independent standard normal variables, independent of the sequence X_k, $k \geq 1$. The principle is to replace one by one the variables X_k in the sum S_n by the G_k's so to obtain at the end the sum $T_n = G_1 + \cdots + G_n$ with is de Moivrian (of variance n). The procedure may be performed along smooth test functions $\psi : \mathbb{R} \to \mathbb{R}$ to evaluate the differences, $k = 1, \ldots, n$,

$$\begin{aligned} \Delta_k \,&=\, \psi\bigg(\frac{1}{\sqrt{n}}\big[X_1 + \cdots + X_k + G_{k+1} + \cdots + G_n\big]\bigg) \\ &\quad - \psi\bigg(\frac{1}{\sqrt{n}}\big[X_1 + \cdots + X_{k-1} + G_k + \cdots + G_n\big]\bigg) \end{aligned}$$

by a Taylor expansion at the third order. Assuming that $\mathbb{E}(|X|^3) < \infty$ and $\|\psi'''\|_\infty \leq 1$, due to the moment normalization, the expectation of the preceding expression may be bounded in absolute value by

$$\big|\mathbb{E}(\Delta_k)\big| \,\leq\, \frac{1}{6n^{3/2}}\,\Big(\mathbb{E}\big(|X|^3\big) + \mathbb{E}\big(|G|^3\big)\Big).$$

Now

$$\sum_{k=1}^n \mathbb{E}(\Delta_k) \,=\, \mathbb{E}\bigg(\psi\Big(\frac{S_n}{\sqrt{n}}\Big)\bigg) - \mathbb{E}\bigg(\psi\Big(\frac{T_n}{\sqrt{n}}\Big)\bigg) \,=\, \mathbb{E}\bigg(\psi\Big(\frac{S_n}{\sqrt{n}}\Big)\bigg) - \mathbb{E}\big(\psi(G_1)\big)$$

so that

$$\bigg|\mathbb{E}\bigg(\psi\Big(\frac{S_n}{\sqrt{n}}\Big)\bigg) - \mathbb{E}\big(\psi(G_1)\big)\bigg| \,\leq\, \frac{1}{6\sqrt{n}}\,\Big(\mathbb{E}\big(|X|^3\big) + \mathbb{E}\big(|G|^3\big)\Big).$$

Since weak convergence may be reduced along families of smooth functions, the central limit theorem follows. The stronger third moment assumption may be then be weakened to the second moment condition after a suitable truncation argument, not developed here.

The Lindeberg proof is closely related to Stein's method [6, 3, 5, 2, 1].

5 The Berry-Esseen inequality

The Berry-Esseen inequality quantifies the convergence of the distribution functions (1) in the central limit theorem uniformly over $t \in \mathbb{R}$ (that is in the Kolmogorov distance between probability laws).

Theorem 2. *Let X be a real random variable such that $\mathbb{E}(|X|^3) < \infty$, and let X_k, $k \geq 1$, be a sequence of independent copies of X. Then, with σ^2 the variance of X, for every $n \geq 1$,*

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} \left[X_k - \mathbb{E}(X_k) \right] \leq t \right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}x^2} dx \right| \leq \frac{C}{\sqrt{n}} \mathbb{E}\big(|X|^3\big)$$

where $C > 0$ is a numerical constant.

The currently best possible value for the constant C is 0.469 in the works of I. G. Shevtsova.

6 The multivariate central limit theorem

If X is a random vector on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with values in $\mathbb{R}^d$, the central limit theorem for a sequence of independent copies of X is equivalent to the central limit theorem for $\langle u, X \rangle$ for any vector $u \in \mathbb{R}^d$. The following statement is then just a reformulation of the standard Theorem 1.

Let $(X_k)_{k\geq 1}$ be a sequence of independent copies of X, and $S_n = X_1 + \cdots + X_n$, $n \geq 1$.

Theorem 3 (Multivariate central limit theorem)**.** *Under the preceding notation, if* $0 < \mathbb{E}(|X|^2) < \infty$,

$$\sqrt{n} \left[\frac{S_n}{n} - \mathbb{E}(X) \right] \to G \quad \textit{in distribution}$$

where G is a de Moivrian random vector in $\mathbb{R}^d$ with law $\mathcal{N}(0, \Sigma)$ where $\mathrm{Cov}(X) = \Sigma$.

7 Unlimited extensions...

It would be an unattainable task to try to describe the (almost!) infinite number illustrations, occurrences, applications of de Moivrian fluctuation-type results and central limit theorems for numerous probabilistic and statistical models and instances. This profusion is the best witness of the *Central* character of the *Central* Limit Theorem!

References

[1] Stein's method for normal approximation. *The Gaussian Blog*.

[2] S. Chatterjee. A short survey of Stein's method. *Kyung Moon Sa*, 1–24. Seoul (2014).

[3] L. Chen, L. Goldstein, Q.-M. Shao. *Normal approximation by Stein's method.* Probability and its Applications. Springer (2011).

[4] R. Dudley. *Real analysis and probability*. Cambridge Studies in Advanced Mathematics 74. Cambridge University Press (2010).

[5] I. Nourdin, G. Peccati. *Normal approximations with Malliavin calculus: from Stein's method to universality*. Cambridge Tracts in Mathematics. Cambridge University Press (2012).

[6] C. Stein. *Approximate computation of expectations.* Institute of Mathematical Statistics Lecture Notes – Monograph Series 7. Institute of Mathematical Statistics (1986).