Central Limit Theorem

! Explore the Central Limit Theorem

This is a quick review of the Central Limit Theorem (CLT), trying to give a concrete meaning to the nature of the convergence (in distribution, *not* in probability).
This page contains (at its end) interactive content, feel free to explore and modify it. For more involved modifications, you are encouraged to run and edit the interactive Jupyter Notebook instead. You can either:

- Run it on a computer with a Python/Jupyter-lab installation (requires `ipywidgets`, `numpy`, `matplotlib`).
- Run it using an online service. The official Jupyter website provides a free and open service. If you need more computing power, you can import the notebook in Google Colab (requires a Google account).

Convergence Results

Let us review some basic convergence results for sums of centered random variables with finite variance.

The classical statement

Theorem 0.1 (Central Limit Theorem)**.** *Let* $(X_n)_{n\geq 1}$ *be an i.i.d. sequence of real-valued random variables with* $\mathbb{E}[|X_n|^2] < \infty$*. Denote* $m := \mathbb{E}[X_n]$*,* $\sigma := \sqrt{\mathrm{Var}[X_n]}$ *and*

$$S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - m}{\sigma} \tag{1}$$

*Then S_n converges **in distribution** to a standard normal random variable, say $Z \sim \mathcal{N}(0,1)$. In other words, for each bounded measurable function $f\colon \mathbb{R} \to \mathbb{R}$ which is continuous a.e., it holds*

$$\lim_{n\to\infty} \mathbb{E}[f(S_n)] = \frac{1}{\sqrt{2\pi}} \int f(x) e^{-x^2/2} dx$$

In particular one can deduce uniform convergence of the distribution function from the previous theorem

$$\lim_{n\to\infty} \sup_{a<b} |\mathbb{P}(a < S_n \leq b) - \mathbb{P}(a < Z \leq b)| = 0$$

A quantitative version

The previous Theorem 0.1 does not address the rate of convergence.

Theorem 0.2 (Quantitative Central Limit Theorem)**.** *Let $(Y_n)_{n\geq 1}$ be a sequence of **independent** random variables with $\mathbb{E}[Y_n] = 0$ and $\mathbb{E}[Y_n^2] = 1$, $\mathbb{E}[|Y_n|^3] < \infty$. Let*

$$S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$$

For $g \in C_b^3(\mathbb{R})$ with $C := \sup_x |g'''(x)|$, the following inequality holds

$$|\mathbb{E}[g(S_n)] - \mathbb{E}[g(Z)]| \leq \frac{C}{6\sqrt{n}} \left(\frac{2^{3/2}}{\sqrt{\pi}} + \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[|Y_k|^3] \right)$$

where $Z \sim \mathcal{N}(0,1)$; namely

$$\mathbb{E}[g(Z)] = \tfrac{1}{\sqrt{2\pi}} \int g(x) e^{-x^2/2} dx$$

Lemma 0.1. *Let V, Y, and Z be three random variables, such that*

- *V and Y are independent; V and Z are independent.*
- *Y and Z have a finite third moment.*
- *$\mathbb{E}[Y] = \mathbb{E}[Z]$ and $\mathbb{E}[Y^2] = \mathbb{E}[Z^2]$.*

Then for any $g \in C_b^3$, setting $C := \sup_{x\in\mathbb{R}} |g'''(x)|$, the following inequality holds:

$$|\mathbb{E}[g(V+Y)] - \mathbb{E}[g(V+Z)]| \leq \frac{C}{6} \left(\mathbb{E}[|Y|^3] + \mathbb{E}[|Z|^3] \right)$$

Lemma 0.1. By Taylor expansion, for three points $v, y, z \in \mathbb{R}$ it holds

$$g(v+y) - g(v+z) = g'(v)(y-z) + \tfrac{1}{2}g''(v)(y^2 - z^2) + R(v,y) - R(v,z)$$

where the remainder terms $R(v,\cdot)$ are bounded as $|R(v,x)| \leq C|x|^3/6$. Computing the last formula at $v = V(\omega)$, $y = Y(\omega)$ and $z = Z(\omega)$, then taking expectation, one gathers

$$|\mathbb{E}[g(V+Y) - g(V+Z)]| = |\mathbb{E}[R(V,Y) - R(V,Z)]| \leq \frac{C}{6}\left(\mathbb{E}[|Y|^3] + \mathbb{E}[|Z|^3]\right)$$

since (using the independence and equal expectation hypotheses) $\mathbb{E}[g'(V)(Y - Z)] = \mathbb{E}[g'(V)]\mathbb{E}[(Y-Z)] = 0$, and reasoning similarly $\mathbb{E}[g''(V)(Y^2 - Z^2)] = 0$. □

Lemma 0.2. *Let $g \in C_b^3(\mathbb{R})$, let $Y_1, \dots, Y_n$ be independent random variables, and $Z_1, \dots, Z_n$ be another set of independent random variables. Assume $\mathbb{E}[Y_i] = \mathbb{E}[Z_i]$, and $\mathbb{E}[Y_i^2] = \mathbb{E}[Z_i^2] < \infty$. Let C be as in Lemma 0.1. Then*

$$\left|\mathbb{E}\left[g\left(\tfrac{Y_1+\ldots+Y_n}{\sqrt{n}}\right) - g\left(\tfrac{Z_1+\ldots+Z_n}{\sqrt{n}}\right)\right]\right| \leq \frac{C}{6n^{3/2}}\sum_{k=1}^{n}\left(\mathbb{E}[|Y_k|^3] + \mathbb{E}[|Z_k|^3]\right)$$

In particular if the Y_i are i.i.d. and the Z_i are i.i.d. (in general with a different distribution)

$$\left|\mathbb{E}\left[g\left(\tfrac{Y_1+\ldots+Y_n}{\sqrt{n}}\right) - g\left(\tfrac{Z_1+\ldots+Z_n}{\sqrt{n}}\right)\right]\right| \leq \frac{C\left(\mathbb{E}[|Y_1|^3] + \mathbb{E}[|Z_1|^3]\right)}{6\sqrt{n}}$$

Lemma 0.2. With no loss of generality, one can assume that all the $Y_1, \dots, Y_n, Z_1, \dots, Z_n$ are independent random variables. Then write $V_k = (Y_1 + \ldots + Y_{k-1} + Z_{k+1} + \ldots + Z_n)/\sqrt{n}$ to get

$$\begin{aligned}\left|\mathbb{E}\left[g\left(\tfrac{Y_1+\ldots+Y_n}{\sqrt{n}}\right) - g\left(\tfrac{Z_1+\ldots+Z_n}{\sqrt{n}}\right)\right]\right| &= \left|\sum_{k=1}^{n}\mathbb{E}\left[g\left(V_k + \tfrac{Y_k}{\sqrt{n}}\right) - g\left(V_k + \tfrac{Z_k}{\sqrt{n}}\right)\right]\right| \\ &\leq \sum_{k=1}^{n}\frac{C}{6}\left(\mathbb{E}[|Y_k/\sqrt{n}|^3] + \mathbb{E}[|Z_k/\sqrt{n}|^3]\right)\end{aligned}$$

where in the last inequality we used Lemma 0.1 n times. □

Theorem 0.2. If the Z_i are i.i.d. standard normal, then $(Z_1 + \ldots + Z_n)/\sqrt{n}$ is also a standard normal and thus its distribution does not depend on n. Theorem 0.2 is therefore a consequence of Lemma 0.2 and the identity $\mathbb{E}[|Z_i|^3] = 2^{3/2}/\sqrt{\pi}$. □

Exercise 0.1. Let $Z \sim \mathcal{N}(0,1)$ be a standard normal random variable. Let $(Y_n)_{n\geq 1}$ be an i.i.d. sequence with $\mathbb{E}[Y_i^k] = \mathbb{E}[Z^k]$ for $k = 1, \dots, \ell$. Let $g \in C_b^\ell(\mathbb{R})$. Prove that there exists a constant C (depending on g and the distribution of the Y_i) such that

$$|\mathbb{E}[g(S_n)] - \mathbb{E}[g(Z)]| \leq Cn^{-(\ell-1)/2}$$

A martingale version

It is worth mentioning that the Central Limit Theorem extends far beyond the scope of independent random variables.Ultimately, this type of result does not even need the variables to be defined on a linear space (e.g. making small random steps on a manifold will, as the steps decrease, converge to a distribution over continuous curves on the manifold, called Brownian Motion). So there are strong local versions of the CLT, metric-space versions, ergodic versions, and so on. An interesting example that only requires elementary hypotheses covers the case of **martingales**.

Theorem 0.3 (Martingale Central Limit Theorem)**.** *Let $(X_n)_{n\geq 1}$ be a sequence of real-valued random variables and let*

$$M_n := X_1 + \dots + X_n$$

Assume that

- $\mathbb{E}[X_n | M_{n-1}] = 0$.
- *For $Q_n := \mathbb{E}[X_n^2 | M_{n-1}]$, it holds $\sum_{n=1}^{\infty} Q_n = \infty$ a.s..*
- $\mathbb{E}\left[\sup_n \mathbb{E}[|X_n|^3 | M_{n-1}]\right] < \infty$.

Let $\tau_\ell := \inf\{N \in \mathbb{N} : \sum_{n=1}^{N} Q_n \geq \ell\}$. Then $M_{\tau_\ell}/\sqrt{\ell}$ converges to a standard normal in distribution as $\ell \to \infty$.

Non-Convergence Results

So far so good. If the Y_n are centered i.i.d. random variables with finite variance S_n converges to a normal limit. **In distribution**. Will this convergence hold in probability or even a.s.?

Proposition 0.1. *Regardless of the probability space and the distribution of the Y_n, the Central Limit Theorem* ***does not*** *hold in probability, not even along subsequences.*

The point is that if two sequences $(S_n), (S'_n)$ converge in probability, then $S_n + S'_n$ converges in probability (by triangular inequality). The same statement does not hold in distribution, since convergence in distribution does not concern the random variables, but only their distribution. Thus the convergence of S_n or S'_n says nothing about their joint distribution.

Proposition 0.1. Any limit point (along some subsequence) in probability S of S_n, will have a standard normal distribution. In particular $\sqrt{2}S_{2n} - S_n$ would converge to $(\sqrt{2}-1)S \sim \mathcal{N}(0, 3-2\sqrt{2})$ in probability (along the same subsequence). But

$$S'_n := \sqrt{2}S_{2n} - S_n = \frac{1}{\sqrt{n}} \sum_{i=n+1}^{2n} Y_i$$

is a sum of n i.i.d. divided by $\sqrt{n}$, thus the Theorem 0.1 applies to S'_n. Namely for any limit point S, $(\sqrt{2}-1)S$ should also have $\mathcal{N}(0,1)$ law. Therefore there are no limit points. □

Visualizing the Convergence

What does it mean that the sequence converges in distribution? Let's fix μ, a centered probability measure on $\mathbb{R}$, and some value n 'large enough' (as we have seen, how large depends on μ, for instance in its third moment) and let us consider i.i.d. $X_1, \dots, X_n$ with law μ and the ensuing S_n as in Equation 1. We can sample many times, say N, $(X_1, \dots, X_n)$ and thus S_n independently. The Central Limit Theorem tells us that, with large probability, the fraction of samples for which S_n falls in a given interval $[a, b]$ is *roughly* equal to the gaussian integral over $[a, b]$. Here *roughly* means with a probability converging to 1 as N and n grow.

Here we take N samples, plot how many of them fall in each interval, and compare the result against the theoretical Gaussian density.

On the other hand, the fact that S_n is *not* converging a.s., means that if we fix a sample (an ω so to speak) and we follow the value of S_n as depending on n, it will not converge to any value.

Each individual sample does not converge as a function of n: we extract the X_i i.i.d., and plot S_n as a function of n. Even for n large, the plot oscillates and convergence does not set in.

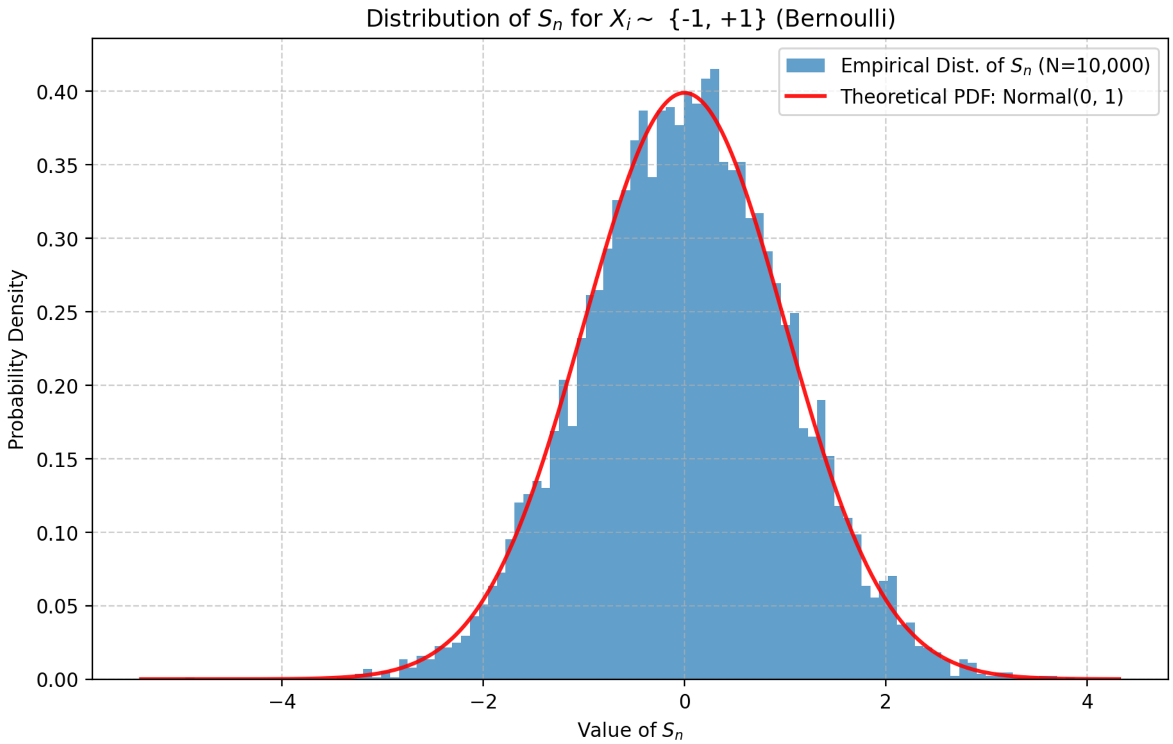

Figure 1: As n is large, the probability of finding a sample in a given interval converges to the Gaussian integral over that interval. This is the content of the Central Limit Theorem. Here we take N samples, plot how many of them fall in each interval, and compare the result against the theoretical Gaussian density.

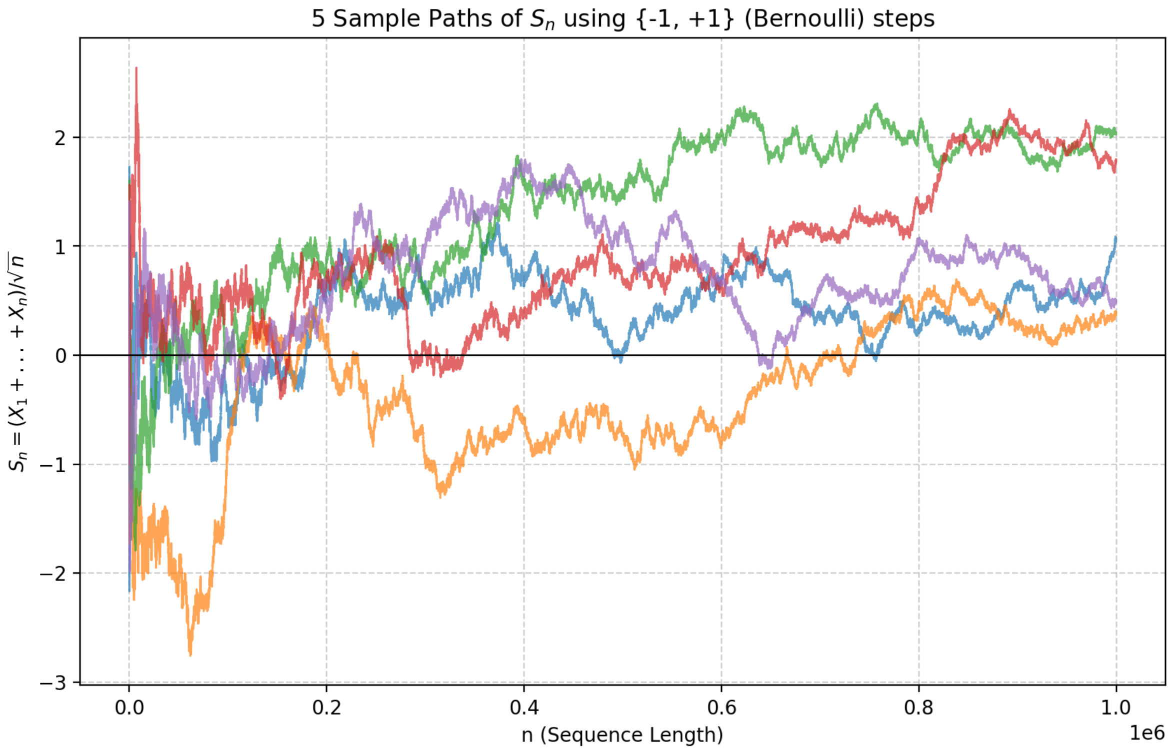

Figure 2: Each individual sample does not converge as a function of n. We extract the X_i i.i.d., and plot S_n as a function of n. Even for n large, the plot oscillates and convergence does not set in. Here we plot $N = 5$ different samples up to $n = 10^6$.