# Ma140a Probability The Central Limit Theorem

Juan Pablo Vigneaux

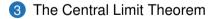
### Caltech

Department of Mathematics California Institute of Technology Pasadena, CA, USA vigneaux@caltech.edu

August 3, 2024

### 1 Experimental results





## **Table of Contents**

## 1 Experimental results

2 De Moivre-Laplace Theorem









#### Theorem (De Moivre-Laplace)

Let  $(X_i)_{i\geq 1}$  be a sequence of iid random variables, each with distribution Ber(p), defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Set  $S_n = X_1 + \cdots + X_n$  and q = 1 - p. Then for any constants a and b such that  $-\infty < a < b < +\infty$ , we have

$$\lim_{n \to \infty} \mathbb{P}\left(a < \frac{S_n - np}{\sqrt{npq}} \le b\right) = \frac{1}{2\pi} \int_a^b e^{-x^2/2} \,\mathrm{d}x.$$
 (1)

There is a very interesting discussion about the intuition behind this theorem here (whuber's answer): https://stats.stackexchange.com/questions/3734/what-intuitive-explanation-is-there-for-the-central-limit-theorem

### Proof.

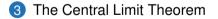
Denoting  $x_k = (k - np) / \sqrt{npq}$ , we rewrite the probability on the LHS of (1) as  $\sum_{a < x_k < b} {n \choose k} p^k q^{n-k}.$ Then use that within that sum  $k \sim np$  and  $n - k \sim nq$ .<sup>1</sup> Together with Stirling's approximation (which should also be credited to De Moivre!) and a Taylor expansion of the logarithm of  $(\frac{np}{k})^k (\frac{nq}{m-k})^{n-k}$ , one proves that for  $|x_k|$  bounded,  $\binom{n}{k} p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi n n q}} e^{-x_k^2/2}.$ By observing that  $x_{k+1} - x_k = 1/\sqrt{npq}$ , the sum  $\sum_{a < x_k \le b} \binom{n}{k} p^k q^{n-k} = \frac{1}{\sqrt{2\pi}} \sum_{a < x_k \le b} e^{-x_k^2/2} (x_{k+1} - x_k)$  can be seen as a Riemann sum approximating  $(2\pi)^{-1/2} \int_{a}^{b} e^{-x^2/2} dx$ .

For details, see Chung and AitSahlia, *Elementary Probability Theory*, Springer, 2003.

 ${}^{1}f_{n} \sim g_{n}$  means that  $\lim_{n} f_{n}/g_{n} = 1$ .

### 1 Experimental results

2 De Moivre-Laplace Theorem



## Generalization

Can we expect the same limiting behavior for an arbitrary sequence of random variables?

- We need some conditions. For instance, if  $(X_i)_i$  is a sequence of iid variables following a Cauchy distribution, with p.d.f.  $\frac{1}{\pi(1+x^2)}$ , then the means  $n^{-1}\sum_{i=1}^{n}$  are also Cauchy distributed. (You see that for these variables not even the LLN holds.)
- However, we shall prove that the theorem at least holds for any sequence  $(X_i)_i$  of iid variables with finite nonzero variance  $\sigma^2$ . (The Cauchy distribution has infinite variance.)
- The  $\sqrt{n}$  denominator gives the good scaling to get a limiting "shape" that's independent of *n*. This is because

$$\mathbb{V}\left(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}}\right) = \mathbb{V}(X_1) = \sigma^2.$$

Suppose for simplicity that  $(X_i)_i$  is a sequence of iid random variables with mean  $\mu = 0$  and variance  $\sigma^2 = 1$ .

Suppose  $\lim \frac{X_1 + \dots + X_{2n}}{\sqrt{2n}} \sim Z$  (i.e. that in the limit the distribution is that of a random variable *Z*). Then it's also the case that

$$\lim_{n} \frac{X_1 + \dots + X_n}{\sqrt{n}} + \frac{X_{n+1} + \dots + X_{2n}}{\sqrt{n}} \sim Z_1 + Z_2.$$

Here  $(Z, Z_1, Z_2)$  are iid, and must have variance 1. It follows that  $\sqrt{2}Z = Z_1 + Z_2$ . If  $Z \sim \mathcal{N}(0, 1)$  this equation is satisfied. It turns out that's the only solution.

# Weak convergence to a normal

#### Definition

We say that a sequence  $(\mu_n)_n \subset \operatorname{Prob}(\mathbb{R})$ , the probability distributions on  $\mathbb{R}$ , converges weakly to  $\mu \in \operatorname{Prob}(\mathbb{R})$  if for all  $h \in C_b(\mathbb{R})$  (bounded function) one has  $\mu_n(h) \to \mu(h)$  as  $n \to \infty$ . This is denoted  $\mu_n \stackrel{w}{\to} \mu$ .

We say that  $X_n$  converges to X in distribution, denoted  $X_n \xrightarrow{D} X$  if  $\Lambda_{X_n} \xrightarrow{w} \Lambda_X$ 

#### Theorem (CLT)

Let  $X_1, X_2, ...$  be a sequence of iid random variables with finite mean  $\mu$  and finite nonzero variance  $\sigma^2$ , and Z a normal random variable with mean 0 and variance 1. Set  $S_n = X_1 + \cdots + X_n$ . Then,

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{D} Z \tag{2}$$

There are several ways of proving the CLT, and all of them are involved in their own way.

We're going to follow the most "classic" path, via *characteristic functions*. The characteristic function of a real-valued random variable X is the complex-valued function  $\phi_X(t) = \mathbb{E}(e^{itX})$  i.e. it's Fourier transform (although this was already considered by Laplace).

#### Theorem

If *X* and *Y* are independent, then  $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$ .

# Laplace's idea for a proof

Supposing that *X* that *X* is discrete and bounded, one would write  $\phi_X(t) = \sum_{k=-m}^m p_k e^{ikt}$ , so that  $\phi_{X_1+\dots+X_n}(t) = (\phi_X(t))^n$ . By using the fact that  $\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itx} e^{isx} dx = \delta_{ts}$ , it follows that  $P(j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ijx} (\sum_{k=-m}^m p_k e^{ikt})^n dx$ . Here P(j) is the probability that  $X_1 + \dots + X_n$  equals *j*. This is an expression of what we would call today an *inversion formula*.

Then the idea is to expand  $(\sum_{k=-m}^{m} p_k e^{ikt})^n$  as a power series in k, and somehow justify that as  $n \to \infty$  all terms of degree greater than 2 become negligible, and that  $P(j) \approx \frac{1}{2\pi\sqrt{n}} \int_{-\infty}^{\infty} e^{-ij\frac{y}{\sqrt{n}}} e^{-\frac{m^2\sigma^2y^2}{2}} dy$ . To this integral one can apply a *saddle-point approximation*, then summing for different values of j one gets again a Riemann sum as in De Moivre-Laplace theorem.

See Hans Fischer, *A History of the Central Limit Theorem*, Springer, 2011, Sec. 2.1.3

# The modern proof, cf. Williams

Facts about characteristic functions (briefly: CFs):

- Taylor expansion: If  $\mathbb{E}|X|^k < \infty$ , then  $\phi_X(t) = \sum_{j=0}^k \frac{\mathbb{E}(X^j)}{j!}(it)^j + o(t^k)$ . See G& S, 5.7.4]
- A cumulative distribution function might be reconstructed from  $\phi_X$  (Levy's inversion formula).
- Weak convergence of probability laws corresponds to pointwise convergence of characteristic functions (Levy's theorem).

## Proof of the CLT.

Suppose  $(X_i)_i$  have mean zero and variance 1. (Otherwise, start by introducing the new variables  $X' = (X - \mu)/\sigma$ .) Then

$$\phi_{S_n/\sqrt{n}}(t) = \phi_X(t/\sqrt{n}) = \left(1 - \frac{1}{2}\frac{t^2}{n} + o(1)\right)^n \to \exp(-\frac{1}{2}t^2)$$

as  $n \to \infty$ . The RHS is the CF of a standard normal distribution  $\mathcal{N}(0,1)$ .

#### Theorem (Levy's inversion formula; Williams 16.6)

Let  $\phi_X$  be the CF of a random variable X that has mean  $\mu$  and distribution function *F*. Then, for a < b,

$$\lim_{T\uparrow\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{ita} - e^{itb}}{it} \phi(t) \, \mathrm{d}t = \frac{1}{2} (\mu(\{a\}) + \mu(\{b\})) + \mu(a, b)$$
(3)  
$$= \frac{1}{2} (F(b) - F(b-)) - \frac{1}{2} (F(a) - F(a-)).$$
(4)

Moreover, if  $\int_{\mathbb{R}} |\phi(t)| dt < \infty$ , then *X* has continuous probability density function *f* and

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi(t) \, \mathrm{d}t.$$

#### Theorem (Levy's convergence theorem; Williams 18.1)

Let  $(F_n)_n$  be a sequence of distribution functions, and  $\phi_n$  denote the characteristic function of  $F_n$ . Suppose that

 $g(t) = \lim_{n} \phi_n(t)$ 

exists for all  $t \in \mathbb{R}$  and that g is continuous at 0. Then  $g = \phi_F$  for some distribution function F and  $F_n \xrightarrow{w} F$ .