A Proof of the Central Limit Theorem

Prof. Michael Mascagni

Department of Computer Science
Department of Mathematics
Department of Scientific Computing
Graduate Program in Molecular Biophysics
Florida State University, Tallahassee, FL 32306 **USA**
E-mail: mascagni@fsu.edu
URL: http://www.cs.fsu.edu/~mascagni

Why the CLT matters

- It explains why averages of many small random effects look approximately normal.
- It powers confidence intervals, hypothesis tests, error analysis, and simulation.
- In CS, it shows up in randomized algorithms, performance modeling, and Monte Carlo methods.

Statement of the theorem

Theorem (Lindeberg–Lëvy CLT)

Let $X_1, X_2, \ldots$ be i.i.d. random variables with

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2 < \infty.$$

Define

$$S_n = X_1 + \cdots + X_n.$$

Then

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

The proof in this talk uses moment generating functions, so we assume the MGF exists in a neighborhood of 0. That covers many common distributions, including Bernoulli, binomial, Poisson, and normal.

Idea of the proof

We prove convergence by studying moment generating functions.

- Standardize the variables so they have mean 0 and variance 1.
- Show the MGF of one standardized summand looks like

$$1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right).$$

- Raise this to the nth power and take the limit:

$$\left(1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{t^2/2}.$$

- Recognize $e^{t^2/2}$ as the MGF of $\mathcal{N}(0, 1)$.

Step 1: standardize

Define

$$Y_i = \frac{X_i - \mu}{\sigma}.$$

Then

$$\mathbb{E}[Y_i] = 0, \qquad \mathrm{Var}(Y_i) = 1.$$

Let

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i.$$

It is enough to show that $Z_n \xrightarrow{d} \mathcal{N}(0,1)$.

Moment generating functions

For a random variable Y, its moment generating function is

$$M_Y(t) = \mathbb{E}[e^{tY}],$$

whenever the expectation exists. This is a Laplace transform, and you can use techniques and information from that to do calculations with them.

- If $Y_1, \ldots, Y_n$ are independent, then

$$M_{Y_1+\cdots+Y_n}(t) = \prod_{j=1}^{n} M_{Y_j}(t).$$

- If a is a constant, then $M_{aY}(t) = M_Y(at)$.
- If MGFs exist on an interval around 0 and converge to the MGF of a distribution, then the distributions converge.

MGF of Z_n

Since the Y_i are i.i.d.,

$$M_{Z_n}(t) = \mathbb{E}\left[e^{t\sum_{i=1}^{n} Y_i/\sqrt{n}}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{tY_i/\sqrt{n}}\right] = \left(M_Y\left(\frac{t}{\sqrt{n}}\right)\right)^n.$$

So the whole proof reduces to understanding $M_Y(u)$ near $u = 0$.

Key lemma: local expansion near 0

Lemma

If $\mathbb{E}[Y] = 0$ *and* $\mathrm{Var}(Y) = 1$*, then as* $u \to 0$*,*

$$M_Y(u) = 1 + \frac{u^2}{2} + o(u^2).$$

Why this is true

Use the Taylor expansion

$$e^{uY} = 1 + uY + \frac{u^2 Y^2}{2} + r(u, Y),$$

where the remainder is small enough after taking expectation. The linear term vanishes because $\mathbb{E}[Y] = 0$, and the quadratic term becomes $u^2/2$ because $\mathbb{E}[Y^2] = 1$.

Applying the lemma

Set $u = t/\sqrt{n}$. Then

$$M_Y\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right).$$

Therefore

$$M_{Z_n}(t) = \left(1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n.$$

Now use the standard limit

$$\left(1 + \frac{a_n}{n}\right)^n \to e^a \quad \text{when } a_n \to a,$$

to get

$$M_{Z_n}(t) \to e^{t^2/2}.$$

Finish the proof

The function

$$e^{t^2/2}$$

is the moment generating function of the standard normal distribution $\mathcal{N}(0,1)$.
Since $M_{Z_n}(t) \to e^{t^2/2}$ in a neighborhood of 0, the MGF convergence theorem gives

$$Z_n \xrightarrow{d} \mathcal{N}(0,1).$$

Undoing the standardization,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1).$$

This completes the proof.

What the theorem does and does not say

- It says the normalized sum approaches a normal distribution as n grows.
- It does not say the original variables are normal.
- It requires finite variance.
- The MGF proof also needs an MGF in a neighborhood of 0.
- The speed of convergence depends on the distribution; heavy tails can slow it down a lot.

Simple example

Suppose X_i are Bernoulli(p) with

$$\mu = p, \qquad \sigma^2 = p(1-p).$$

Then for large n,

$$\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \approx \mathcal{N}(0,1).$$

This is the basis of the normal approximation to the binomial distribution.

Why CS students should care

- Randomized algorithms often analyze sums of independent trials.
- Performance measurements combine many small sources of noise.
- Monte Carlo estimates become approximately normal after rescaling.
- The CLT is the bridge between discrete randomness and continuous Gaussian models.

Summary

- Standardize the sum.
- Use MGFs and independence.
- Expand near zero to capture the first two moments.
- Show the limit is $e^{t^2/2}$.
- Conclude convergence to $\mathcal{N}(0, 1)$.

Questions?