CENTRAL LIMIT THEOREM

LUKE BULA

Abstract. This paper develops the mathematical background necessary to understand the Central Limit Theorem and presents a proof using moment generating functions. It then explores the intuition behind the bell-shaped curve, including the concentration of outcomes near the mean and the emergence of symmetry. Several applications are considered, including sums of dice, Monte Carlo option pricing, biological measurements, and the multivariate Central Limit Theorem. Finally, the Cauchy distribution is discussed as an example showing why the theorem's assumptions are essential.

1. Introduction

A fair die has a uniform distribution, a lottery has a heavily skewed distribution, and the daily movement of a stock price may follow a complicated and unpredictable pattern. Yet when many independent observations from these very different distributions are added or averaged, the same familiar shape repeatedly emerges: the bell curve. This remarkable phenomenon is described by the Central Limit Theorem.

Theorem 1.1 (The Central Limit Theorem). *If $X_1, X_2, ..., X_n$ is a sequence of independently distributed and identically distributed random variables, each having mean μ and variance σ^2, and $S_n = \sum_{i=1}^{n} X_i$, then as $n \to \infty$, $\frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$.*

The origins of the theorem can be traced to the study of games of chance by Abraham de Moivre and Pierre-Simon Laplace. These early results, involving fair and later skewed, coin flips formed what is now called the de Moivre–Laplace theorem. During the nineteenth and early twentieth centuries, mathematicians such as Chebyshev, Markov, and Lyapunov developed increasingly general versions of the result. The name "Central Limit Theorem" was introduced by George Pólya in 1920, reflecting the theorem's central role in probability theory.

The theorem is important because it turns complicated random systems into approximately normal ones. Instead of calculating the exact distribution of a large sum, which may require an enormous amount of counting or computation, one can often use the normal distribution to obtain a simple and highly accurate approximation. The purpose of this paper is to explain both why the Central Limit Theorem is mathematically true and why it is intuitively reasonable. After introducing the necessary background, we prove the theorem using moment generating functions and examine how repeated addition produces concentration, symmetry, and the characteristic bell-shaped curve. We then explore several practical applications and higher-dimensional generalizations, as well as examples such as the Cauchy distribution for which the classical theorem fails.

Date: July 12, 2026.

2. Background

Definition 2.1 (Random Variables)**.** A **random variable** is a quantity whose value depends on the outcome of a random experiment. More formally, a random variable is a function

$$X : \Omega \to \mathbb{R},$$

where Ω is the sample space of all possible outcomes. For each outcome $\omega \in \Omega$, the random variable assigns a real number $X(\Omega)$.

Definition 2.2 (Expected Values)**.**

$$E[|X|] < \infty,$$

then the **expected value** of X is defined by

$$E[X] = \int_\Omega X \, dP.$$

For a discrete random variable,

$$E[X] = \sum_x x P(X = x),$$

while for a continuous random variable with probability density function f_X,

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx.$$

Definition 2.3 (Cumulative Distribution Functions)**.** Let X be a random variable. The **cumulative distribution function**, or CDF, of X is the function

$$F_X : \mathbb{R} \to [0, 1]$$

defined by

$$F_X(x) = P(X \leq x).$$

Thus, $F_X(x)$ gives the probability that the random variable X takes a value less than or equal to x.

Definition 2.4 (Probability Density Functions)**.** Let X be a continuous random variable. A function

$$f_X : \mathbb{R} \to [0, \infty)$$

is called a **probability density function**, or PDF, for X if

$$P(a \leq X \leq b) = \int_a^b f_X(x) \, dx$$

for all real numbers $a \leq b$. Equivalently, for every Borel set $B \subseteq \mathbb{R}$,

$$P(X \in B) = \int_B f_X(x) \, dx.$$

The total area under the density is equal to 1:

$$\int_{-\infty}^{\infty} f_X(x) \, dx = 1.$$

Definition 2.5 (Moment Generating Functions)**.** Let X be a real-valued random variable. The **moment generating function** of X, abbreviated MGF, is the function

$$M_X(t) = E(e^{tX}),$$

provided that this expectation exists for t in some open interval containing 0. Equivalently, if f is the distribution function of X, then

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx.$$

The name comes from the fact that, when the MGF exists near 0, its derivatives at 0 give the moments of X:

$$M_X^{(k)}(0) = E(X^k).$$

Definition 2.6 (Stieltjes Integral Notation)**.** If F is the distribution function of a random variable X, then integrals of the form

$$\int_{-\infty}^{\infty} g(x)\, dF(x)$$

are called **Stieltjes integrals**. In probability, this notation means that we are integrating $g(x)$ with respect to the probability distribution of X. Thus,

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)\, dF(x).$$

For example, the moment generating function can be written as

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx}\, dF(x).$$

This notation is especially useful because it works for discrete, continuous, and mixed random variables.

Definition 2.7 (Analytic Functions)**.** A complex-valued function f is called **analytic** on a region $D \subseteq \mathbb{C}$ if it is complex differentiable at every point of D. Analytic functions have strong rigidity properties. For example, if two analytic functions agree on a set of points with a limit point inside their domain, then they agree everywhere on the connected part of the domain.

Definition 2.8. Uniform Boundedness A sequence of functions $\{f_n\}$ is said to be **uniformly bounded** on a set D if there exists a constant $C > 0$ such that

$$|f_n(x)| \le C$$

for every n and every $x \in D$.

Definition 2.9 (Compactness)**.** A subset K of a metric space is called **compact** if every open cover of K has a finite subcover. That is, whenever

$$K \subseteq \bigcup_{\alpha \in A} U_\alpha$$

where each U_α is open, there exist finitely many indices

$$\alpha_1, \alpha_2, \ldots, \alpha_n$$

such that

$$K \subseteq U_{\alpha_1} \cup U_{\alpha_2} \cup \cdots \cup U_{\alpha_n}.$$

In $\mathbb{R}^n$, a set is compact if and only if it is closed and bounded.

Lemma 2.10 (Vitali's Theorem)**.** *Let $\{f_n\}$ be a sequence of analytic functions on a region $D \subseteq \mathbb{C}$. Suppose that $\{f_n\}$ is uniformly bounded on every compact subset of D. Suppose also that $f_n(z)$ converges for every z in some subset of D having a limit point in D. Then there exists an analytic function f on D such that*

$$f_n \to f$$

uniformly on every compact subset of D.

Definition 2.11 (Characteristic Functions)**.** Let X be a random variable. The **characteristic function** (of X is the function

$$\phi_X : \mathbb{R} \to \mathbb{C}$$

defined by

$$\phi_X(t) = E\left[e^{itX}\right],$$

where $i = \sqrt{-1}$. Equivalently,

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx}\, dF_X(x),$$

where F_X is the cumulative distribution function of X.

Lemma 2.12 (Lévy's Continuity Theorem)**.** *Let $X_1, X_2, \ldots$ be random variables with characteristic functions $\phi_1, \phi_2, \ldots$. Suppose that*

$$\lim_{n\to\infty} \phi_n(t) = \phi(t)$$

for every real t, and suppose that ϕ is continuous at $t = 0$. Then ϕ is the characteristic function of some random variable X, and

$$X_n \Rightarrow X.$$

That is, if F_n and F are the distribution functions of X_n and X, then

$$\lim_{n\to\infty} F_n(x) = F(x)$$

at every continuity point of F.

Definition 2.13 (Leibniz's Rule for Differentiating Under the Integral)**.** Leibniz's rule gives conditions under which one may differentiate an integral depending on a parameter by differentiating the integrand. In a simple form, if $g(x,t)$ is sufficiently well-behaved, then

$$\frac{d}{dt}\int g(x,t)\, dF(x) = \int \frac{\partial}{\partial t} g(x,t)\, dF(x).$$

Lemma 2.14. *Let $F_n(x)$ and $M_n(a)$ be respectively the distribution function and moment generating function of a random variable X_n and $F(x)$ and (a) be respectively the distribution functions and moment generating function of a variable X. Suppose that $M_n(a)$ exists for*

$$|a| < a_1$$

and for all $n \geq n_0$. Suppose also that $M(a)$ is a finite valued function, defined for

$$|a| \leq a_2 < a_1, \qquad a_2 > 0,$$

such that

$$\lim_{n\to\infty} M_n(a) = M(a), \qquad |a| \le a_2.$$

Then

$$\lim_{n\to\infty} F_n(x) = F(x)$$

at each continuity point of F, and uniformly in each finite or infinite interval of continuity of F.

Appending the proof in the cited paper [Cur42], we will now prove this theorem.

Proof. In proving this theorem, we first introduce the Laplace transform

$$\varphi_n(s) = E(e^{sX_n})$$

where $s = a + it$ and $a, t \in \mathbb{R}$ and i is imaginary. This allows us to capture both the characteristic function, $\varphi_n(it)$, and the moment generating function of X_n, $\varphi_n(a)$, in a transform that has the potential to converge to an analytic function. Moving forward, this will allow us to leverage powerful theorems and show that convergence of MGFs leads to convergence in characteristic functions.

Noting that $|e^{itX}| = 1$, we see that $\varphi(|s|) = E(|e^{(a+it)X_n}|) = E(|e^{itX_n}e^{aX_n}|) = E(e^{aX_n}) = M_n(a)$. Thus, we also find that

$$|\varphi_n(s)| \le \varphi_n(|s|) = M_n(a), \qquad n > n_0,$$

for any s in the strip

$$-a_1 < \Re(s) < a_1.$$

Writing $M_n(a)$ in Stieltjes integral notation, we can now write

$$M_n(a) = \int_{-\infty}^{+\infty} e^{ax}\, dF_n(x)$$

for both continuous and discrete cases of X_n Differentiating by Leibniz's rule under the integral sign, we find that

$$M_n''(a) = \int_{-\infty}^{+\infty} x^2 e^{ax}\, dF_n(x), \qquad |a| < a_1,$$

from which it appears that

$$M_n''(a) > 0, \qquad |a| < a_1.$$

Acknowledging the convexity of $M_n(a)$ we can now see that $M_n(a) \le \max\{M_n(-a_2), M_n(a_2)\}, \forall a \in [-a_2, a_2]$. By our assumptions, $M_n(_2)$ approach finite limits, $M(\pm a_2)$, as n becomes infinite. Therefore the sequence

$$\{M_n(a)\}, \qquad n \ge n_0,$$

is uniformly bounded in the interval $|a| \le a_2$.

Thus the sequence

$$\{\varphi_n(s)\}, \qquad n \ge n_0,$$

is uniformly bounded in the strip

$$-a_2 \le \Re s \le a_2,$$

and moreover has a limit at each point of an infinite set possessing a limit point in the strip, namely at each point of the interval

$$-a_2 < s < a_2.$$

So by Vitali's theorem, there exists an analytic function $\varphi^*(s)$ such that

$$\lim_{n\to\infty} \varphi_n(s) \to \varphi^*(s)$$

uniformly in each bounded closed subregion of the strip

$$-a_2 < \Re s < a_2.$$

Noting that

$$\lim_{n\to\infty} \varphi_n(it) \to \varphi^*(it),$$

we have now shown convergence of characteristic functions.

Finally, using Levy's continuity theorem, we can see that

$$\lim_{n\to\infty} F_n(x) \to F(x),$$

and, thus, the proof is complete.

Definition 2.15 (The Standard Normal Distribution)**.** A random variable Z has the **standard normal distribution**, written

$$Z \sim N(0,1),$$

if it has probability density function

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty.$$

The standard normal distribution has mean 0 and variance 1:

$$E(Z) = 0, \qquad \mathrm{Var}(Z) = 1.$$

Its moment generating function is

$$M_Z(t) = E(e^{tZ}) = e^{t^2/2}.$$

Proposition 2.16. *If X and Y are independent random variables, then*

$$M_{X+Y}(t) = M_X(t)\, M_Y(t),$$

Proof. By the definition of the moment generating function,

$$M_{X+Y}(t) = \mathbb{E}\left[e^{t(X+Y)}\right].$$

Using the exponent rule $e^{a+b} = e^a e^b$, we obtain

$$M_{X+Y}(t) = \mathbb{E}\left[e^{tX} e^{tY}\right].$$

Since X and Y are independent, the random variables e^{tX} and e^{tY} are also independent. Therefore, provided the relevant expectations are finite,

$$\mathbb{E}\left[e^{tX} e^{tY}\right] = \mathbb{E}\left[e^{tX}\right] \mathbb{E}\left[e^{tY}\right].$$

Recognizing these expectations as the moment generating functions of X and Y, we conclude that

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

Proposition 2.17. *Let Y be a random variable with moment generating function $M_Y(t)$. If $b \in \mathbb{R}$ and $b \neq 0$, then the moment generating function of $\frac{Y}{b}$ is*

$$M_{Y/b}(t) = M_Y\left(\frac{t}{b}\right),$$

for every t for which the expectation is finite.

Proof. By the definition of the moment generating function,

$$M_{Y/b}(t) = \mathbb{E}\left[e^{t(Y/b)}\right].$$

Since

$$t\left(\frac{Y}{b}\right) = \frac{t}{b}Y,$$

we may rewrite this as

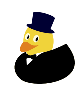
$$M_{Y/b}(t) = \mathbb{E}\left[e^{(t/b)Y}\right].$$

By the definition of M_Y, we have

$$\mathbb{E}\left[e^{(t/b)Y}\right] = M_Y\left(\frac{t}{b}\right).$$

Therefore,

$$\boxed{M_{Y/b}(t) = M_Y\left(\frac{t}{b}\right)}.$$



3. Proof

We closely follow the proof of the Central Limit Theorem given in Sheldon Ross's textbook *A First Course In Probability* [Ros14].

Proof. Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean $\mu = 0$ and variance $\sigma^2 = 1$, and let

$$S_n = \sum_{i=1}^{n} X_i.$$

We want to show that the moment generating function of

$$\frac{S_n}{\sqrt{n}}$$

converges to the moment generating function of the standard normal distribution, $\mathcal{N}(0,1)$, which is given by

$$M_{\mathcal{N}(0,1)}(t) = e^{\frac{t^2}{2}}.$$

Now, by the definition of the MGF, we see that the MGF of X_i, for some $i \in \mathbb{N}$, is given by

$$M_{X_i}(t) = \mathbb{E}\left[e^{tX_i}\right].$$

We use Proposition 2.16 to show that

$$M_{S_n}(t) = M_{X_1 + \cdots + X_n}(t) = \left(M_{X_i}(t)\right)^n.$$

We also use Proposition 2.17 to see that

$$M_{S_n/\sqrt{n}}(t) = [M_{X^i}(\frac{t}{\sqrt{n}})]^n.$$

We have now obtained an expression for the MGF of $\frac{S_n}{\sqrt{n}}$. To show the convergence of our MGFs and ultimately utilize Lemma 2.14, we first suppose $L(t) = \log M_{X_i}(t)$. We take note of the fact that

$$\begin{aligned}
L(0) &= 0, \\
L'(0) &= \frac{M'_{X_i}(0)}{M_{X_i}(0)} \\
&= \mu \\
&= 0, \\
L''(0) &= \frac{M_{X_i}(0)M''_{X_i}(0) - \left[M'_{X_i}(0)\right]^2}{\left[M_{X_i}(0)\right]^2} \\
&= E[X^2] \\
&= 1.
\end{aligned}$$

Now, using the identity $a^n = e^{n \log a}$, we rewrite $[M_{X^i}(\frac{t}{\sqrt{n}})]^n$ as $e^{nL(\frac{t}{\sqrt{n}})}$ where $L(\frac{t}{\sqrt{n}}) = \log{(M_{X^i}(\frac{t}{\sqrt{n}}))}$.

We will now conclude the proof and show that

$$\lim_{n\to\infty} [M_{X^i}(\frac{t}{\sqrt{n}})]^n = e^{\frac{t^2}{2}}$$

or equivalently

$$\lim_{n\to\infty} nL(\frac{t}{\sqrt{n}}) = \frac{t^2}{2}.$$

First, we rewrite the limit as

$$= \lim_{n\to\infty} \frac{L(t/\sqrt{n})}{n^{-1}}.$$

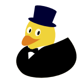
Applying L'Hopital's rule, we get

$$\begin{aligned}
&= \lim_{n\to\infty} \frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}} \\
&= \lim_{n\to\infty} \left[\frac{L'(t/\sqrt{n})t}{2n^{-1/2}}\right].
\end{aligned}$$

Applying L'Hopital's rule again, we get

$$\begin{aligned}
&= \lim_{n\to\infty} \left[\frac{-L''(t/\sqrt{n})n^{-3/2}t^2}{-2n^{-3/2}}\right] \\
&= \lim_{n\to\infty} \left[L''\left(\frac{t}{\sqrt{n}}\right)\frac{t^2}{2}\right]
\end{aligned}$$

$$= \frac{t^2}{2}$$

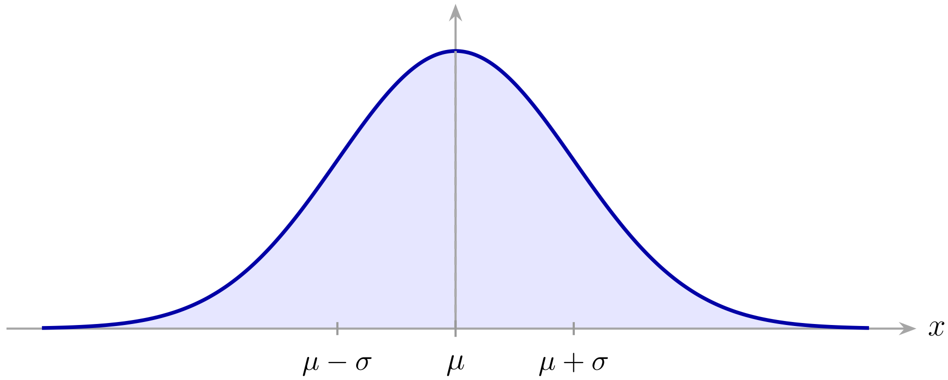


4. Intuition

Now, one may realize that the given proof of the CLT does not provide much intuition behind the formation of the normal distribution. Although it shows us that the normal distribution must appear, it does not tell us why: why that shape and why that equation? Unfortunately, there is not much concrete intuitive reasoning behind the CLT that is between the elementary explanation that "the random variables cancel out" and something that involves full-on analysis. However, it is worthwhile to spend time explaining the phenomena in a qualitative way.

The intuition of the central limit theorem is best exemplified by the Galton Board (demonstration), where the final shape of data emerges from thousands of tiny, independent, random decisions canceling each other out. When a bead drops through the grid of pegs, each peg forces it to bounce left or right with equal probability. For a bead to end up on the far left or far right edge, it must choose the exact same direction at every single peg, a rare streak of extreme luck. Instead, the vast majority of beads take a zigzagging path of mixed left and right bounces that naturally pull them back toward the center. This accumulation of random, balancing choices is why thousands of individual beads inevitably stack up into a perfect, predictable bell curve at the bottom, mimicking how random sampling tames the chaos of any population.

Bell Curve



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

As seen in the equation of the bell curve, the normal distribution has an exponential quadratic shape. To make sense of this, it is useful to count events. One definition of probability is given in terms of events: $P(A) = \frac{\text{outcomes in A}}{\text{total outcomes}}$. It portrays probability as a ratio between successful events and total events. Now, suppose the case where we flip 50 coins simultaneosly and consider a flip a success $\binom{50}{25}$ ways for this to happen. Then, we consider the case where 26 heads and 24 tails is a success. This has $\binom{50}{26}$ ways to happen. Because of the way combinations are calculated, the relationship between the number of successful outcomes and the distance from the mean is exponential quadratic at very large amounts. This is what leads to the exponential quadratic shape of the curve.

Another point of confusion or disbelief behind the CLT is the symmetry of the resulting distribution. Given many skewed random variables, for example, it can be difficult to see why the resulting distribution is symmetrical (try this out in the following simulation). Wnen thinking about this, it useful to think about the mean of the sum of the random variables and the distribution's behavior around this value. We will illustrate this with an example. Suppose we have a heavily-skewed distribution, like a lottery where 99% of people win \$0, and 1% win \$1,000. If you pull a large sample of 100 people at a time and average their winnings most samples will contain 1 winner, it is slightly less likely you get 0 or 2 winners by the exact same margin of chance. In this case, the distribution is very skewed since the chance that 100 people win the lottery is astronomically smaller than no one winning. However, as the number of people in our sample increases, the expected value grows farther and farther from the both lower and upper boundaries in our distribution, which evens out the tails. This is what leads to symmetry in the distribution.

5. Application: The Sum of Thirty Dice

One of the most useful applications of the Central Limit Theorem is the approximation of probabilities involving sums of many independent random variables. As an example, suppose that thirty fair six-sided dice are rolled. We wish to determine the probability that the sum of the dice is greater than 100.

Let X_i denote the outcome of the i-th die, and let

$$S_{30} = X_1 + X_2 + \cdots + X_{30}$$

denote the sum of the thirty dice. Each X_i is uniformly distributed on the set

$$\{1, 2, 3, 4, 5, 6\}.$$

5.1. The Direct Counting Approach.

In principle, this probability could be calculated using combinatorics. There are

$$6^{30} = 221073919720733357899776$$

equally likely possible outcomes when thirty dice are rolled. We would have to count how many of these outcomes have a sum greater than 100.

Using generating functions, the exact probability can be written as

$$\mathbb{P}(S_{30} > 100) = \frac{1}{6^{30}} \sum_{s=101}^{180} [x^s] \left(x + x^2 + x^3 + x^4 + x^5 + x^6\right)^{30},$$

where $[x^s]f(x)$ denotes the coefficient of x^s in the polynomial $f(x)$.

Although this formula is exact, expanding a polynomial of degree 180 and adding all of the coefficients corresponding to sums from 101 to 180 would be extremely tedious to perform by hand. The Central Limit Theorem gives a much simpler approximation.

5.2. **Applying the Central Limit Theorem.** We first calculate the mean and variance of a single die. Since the die is fair,

$$\mathbb{E}[X_i] = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}.$$

Also,

$$\mathbb{E}[X_i^2] = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} = \frac{91}{6}.$$

Therefore,

$$\begin{aligned}\operatorname{Var}(X_i) &= \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 \\ &= \frac{91}{6} - \left(\frac{7}{2}\right)^2 \\ &= \frac{35}{12}.\end{aligned}$$

Because the dice are independent, the mean of their sum is

$$\mu_{S_{30}} = 30\left(\frac{7}{2}\right) = 105,$$

and its variance is

$$\sigma_{S_{30}}^2 = 30\left(\frac{35}{12}\right) = \frac{175}{2} = 87.5.$$

Thus, the standard deviation is

$$\sigma_{S_{30}} = \sqrt{\frac{175}{2}} \approx 9.354.$$

The Central Limit Theorem tells us that the standardized sum

$$\frac{S_{30} - 105}{\sqrt{175/2}}$$

is approximately distributed as a standard normal random variable. In other words,

$$S_{30} \approx N\left(105, \frac{175}{2}\right).$$

Since S_{30} can take only integer values,

$$S_{30} > 100 \quad \Longleftrightarrow \quad S_{30} \geq 101.$$

When replacing the discrete distribution by a continuous normal distribution, we use a continuity correction. The boundary between 100 and 101 is 100.5, so

$$\mathbb{P}(S_{30} > 100) = \mathbb{P}(S_{30} \geq 101) \approx \mathbb{P}(N > 100.5),$$

where

$$N \sim N\left(105, \frac{175}{2}\right).$$

Standardizing gives

$$\begin{aligned}\mathbb{P}(S_{30} > 100) &\approx \mathbb{P}\left(Z > \frac{100.5 - 105}{\sqrt{175/2}}\right)\\ &= \mathbb{P}\left(Z > -\frac{9}{\sqrt{350}}\right)\\ &\approx \mathbb{P}(Z > -0.4811),\end{aligned}$$

where $Z \sim N(0,1)$. By symmetry of the standard normal distribution,

$$\begin{aligned}\mathbb{P}(Z > -0.4811) &= \Phi(0.4811)\\ &\approx 0.68477.\end{aligned}$$

Therefore, the Central Limit Theorem gives the approximation

$$\boxed{\mathbb{P}(S_{30} > 100) \approx 0.68477}$$

or approximately a 68.48% chance.

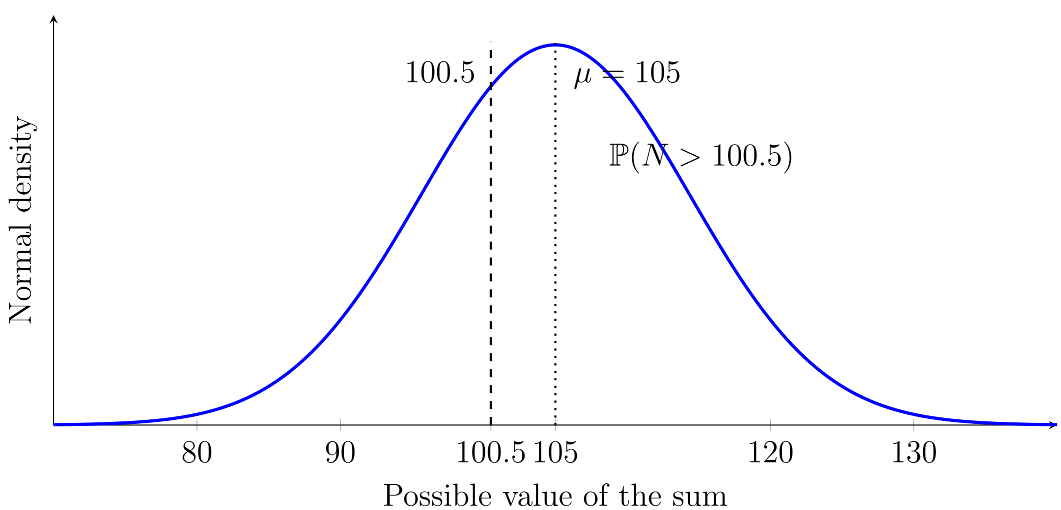


Figure 1. The shaded region represents the normal approximation to $\mathbb{P}(S_{30} > 100)$. The boundary 100.5 is used for the continuity correction.

5.3. **Accuracy of the Approximation.** The exact probability, calculated by extracting the appropriate coefficients of the generating function, is

$$\mathbb{P}(S_{30} > 100) = 0.6840091634\ldots.$$

The continuity-corrected normal approximation gives

$$\mathbb{P}(S_{30} > 100) \approx 0.6847667092.$$

Therefore, the absolute error is

$$|0.6847667092 - 0.6840091634| = 0.0007575458.$$

The relative error is approximately

$$\frac{0.0007575458}{0.6840091634} \times 100\% \approx 0.111\%.$$

The results can be summarized as follows:

Method	Probability
Exact counting probability	0.6840092
Normal approximation with continuity correction	0.6847667
Absolute error	0.0007575
Relative error	0.111%

Thus, the normal approximation differs from the exact probability by only about 0.076 percentage points. This example demonstrates how the Central Limit Theorem can replace a difficult counting problem with a relatively simple calculation involving the normal distribution.

It is also important to note the effect of the continuity correction. Without it, one would calculate

$$\mathbb{P}(N > 100) \approx 0.70351,$$

which is considerably less accurate. The continuity correction accounts for the fact that the dice sum is discrete and substantially improves the normal approximation.

6. Application: Monte Carlo Option Pricing

An important application of the Central Limit Theorem in finance is the numerical pricing of options.

6.1. **What Is an Option?**

An *option* is a financial contract that gives its owner the right, but not the obligation, to buy or sell an asset at a predetermined price. A *European call option* gives its owner the right to buy a stock at a fixed price K, called the *strike price*, at a future time T.

Let S_T denote the stock price at time T. If $S_T > K$, the owner can buy the stock for K dollars and obtain an asset worth S_T dollars. If $S_T \leq K$, the owner simply does not exercise the option. Therefore, the payoff is

$$X = \max(S_T - K, 0) = (S_T - K)^+.$$

For example, if $K = 100$ and $S_T = 120$, the payoff is 20. If $S_T = 90$, the payoff is 0.

6.2. **The Difficulty of Pricing an Option.** The difficulty is that the future stock price S_T is unknown. Since the payoff depends on S_T, the payoff is also random.

Under the standard no-arbitrage framework, the current price of a European call option is

$$C_0 = e^{-rT}\mathbb{E}^{\mathbb{Q}}\left[(S_T - K)^+\right],$$

where r is the risk-free interest rate and $\mathbb{E}^{\mathbb{Q}}$ denotes expectation under the risk-neutral model.

Thus, the option price is the expected future payoff, discounted to its present value. For simple options, this expectation can sometimes be calculated exactly. However, many options depend on several stocks, the average stock price, or the entire path followed by the stock. In these cases, a simple formula may not exist.

6.3. **Monte Carlo Pricing.** Monte Carlo simulation approximates the expected payoff by generating many possible future stock prices.

Suppose that

$$S_T^{(1)}, S_T^{(2)}, \ldots, S_T^{(N)}$$

are N independently simulated future stock prices. The Monte Carlo estimate of the option price is

$$\widehat{C}_N = e^{-rT}\frac{1}{N}\sum_{i=1}^{N}\left(S_T^{(i)} - K\right)^+.$$

The procedure is simple:

(1) Simulate a possible future stock price.
(2) Calculate the option payoff.
(3) Repeat the process many times.
(4) Average and discount the simulated payoffs.

In the Black–Scholes model, future stock prices may be simulated using

$$S_T^{(i)} = S_0\exp\left(\left(r - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}Z_i\right),$$

where S_0 is the current stock price, σ is the volatility, and

$$Z_i \sim N(0,1).$$

Each independently generated value of Z_i produces a different possible future stock price.

6.4. **The Role of the Central Limit Theorem.** Define the discounted payoff from the i-th simulation by

$$Y_i = e^{-rT}\left(S_T^{(i)} - K\right)^+.$$

Then

$$\widehat{C}_N = \frac{1}{N}\sum_{i=1}^{N}Y_i.$$

The random variables $Y_1, \ldots, Y_N$ are independent and identically distributed, and their expected value is the true option price:

$$\mathbb{E}[Y_i] = C_0.$$

The Law of Large Numbers guarantees that

$$\widehat{C}_N \longrightarrow C_0$$

as $N \to \infty$. Therefore, increasing the number of simulations causes the estimate to approach the true option price.

The Central Limit Theorem provides more information. If

$$\mathrm{Var}(Y_i) = \sigma_Y^2 < \infty,$$

then

$$\frac{\sqrt{N}\left(\widehat{C}_N - C_0\right)}{\sigma_Y} \xrightarrow{d} N(0,1).$$

Hence, for large N,

$$\widehat{C}_N \approx N\left(C_0, \frac{\sigma_Y^2}{N}\right).$$

The standard error of the estimate is approximately

$$\mathrm{SE}(\widehat{C}_N) = \frac{s_Y}{\sqrt{N}},$$

where s_Y is the sample standard deviation of the simulated discounted payoffs. An approximate 95% confidence interval is

$$\boxed{\widehat{C}_N \pm 1.96\frac{s_Y}{\sqrt{N}}}.$$

Thus, the Central Limit Theorem allows analysts to measure the uncertainty in a Monte Carlo option-price estimate.

6.5. **Rate of Convergence.** The standard error decreases at the rate

$$\frac{1}{\sqrt{N}}.$$

Therefore, reducing the error by a factor of 2 requires approximately four times as many simulations, while reducing it by a factor of 10 requires approximately one hundred times as many.

The Law of Large Numbers explains why Monte Carlo pricing converges, while the Central Limit Theorem describes the size and distribution of the remaining error. This makes Monte Carlo option pricing a direct and practical application of the Central Limit Theorem.

7. A Higher-Dimensional Application: Joint Health Measurements

The Central Limit Theorem can be generalized to random variables that take values in higher-dimensional spaces. This is useful when several related quantities are measured at the same time. For example, suppose researchers record the systolic blood pressure and cholesterol level of each patient in a study. For the ith patient, these measurements can be represented by the random vector

$$X_i = \begin{pmatrix} B_i \\ C_i \end{pmatrix},$$

where B_i denotes systolic blood pressure and C_i denotes cholesterol level.

Let the population mean vector be

$$\boldsymbol{\mu} = \mathbb{E}[X_i] = \begin{pmatrix} \mu_B \\ \mu_C \end{pmatrix},$$

where μ_B and μ_C are the population mean blood pressure and cholesterol levels. The variability and relationship between the two measurements are described by the covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_B^2 & \mathrm{Cov}(B, C) \\ \mathrm{Cov}(B, C) & \sigma_C^2 \end{pmatrix}.$$

The diagonal entries give the variances of blood pressure and cholesterol, while the off-diagonal entries give their covariance. A positive covariance means that patients with above-average blood pressure also tend to have above-average cholesterol levels.

Suppose measurements are collected from n independent patients. The sample mean vector is

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \begin{pmatrix} \overline{B}_n \\ \overline{C}_n \end{pmatrix},$$

where $\overline{B}_n$ and $\overline{C}_n$ are the sample mean blood pressure and cholesterol levels.

The multivariate Central Limit Theorem states that, provided the random vectors have finite means and covariance matrices,

$$\sqrt{n}\left(\overline{X}_n - \boldsymbol{\mu}\right) \xrightarrow{d} N_2(\mathbf{0}, \Sigma).$$

Equivalently, for sufficiently large n,

$$\overline{X}_n \approx N_2\left(\boldsymbol{\mu}, \frac{\Sigma}{n}\right).$$

Thus, the joint distribution of the two sample means is approximately a two-dimensional normal distribution, even if the individual measurements are not normally distributed.

Although this is called a two-dimensional normal distribution because it involves two random variables, its probability density can be visualized as a bell-shaped surface in three-dimensional space. The two horizontal axes represent blood pressure and cholesterol, while the vertical axis represents probability density. Viewed from above, the contours of this surface form ellipses centered at

(μ_B, μ_C).

The shape and orientation of these ellipses depend on the covariance. If blood pressure and cholesterol are positively correlated, the ellipses are tilted upward, reflecting the tendency of high values of one measurement to occur with high values of the other.

As a numerical example, suppose

$$\boldsymbol{\mu} = \begin{pmatrix} 120 \\ 190 \end{pmatrix}$$

and

$$\Sigma = \begin{pmatrix} 100 & 60 \\ 60 & 400 \end{pmatrix}.$$

Here, the population standard deviations are 10 for blood pressure and 20 for cholesterol, and the positive covariance indicates that the two measurements tend to increase together. If a sample of $n = 100$ patients is taken, then

$$\overline{X}_{100} \approx N_2 \left(\begin{pmatrix} 120 \\ 190 \end{pmatrix}, \begin{pmatrix} 1 & 0.6 \\ 0.6 & 4 \end{pmatrix} \right).$$

Therefore, the standard deviations of the sample means are

$$\sigma_{\overline{B}} = 1 \qquad \text{and} \qquad \sigma_{\overline{C}} = 2.$$

The covariance 0.6 shows that samples with an unusually high average blood pressure also tend to have an unusually high average cholesterol level.

This approximation allows researchers to study both measurements simultaneously. For example, they may approximate probabilities such as

$$P\left(\overline{B}_n > 130 \text{ and } \overline{C}_n > 200\right),$$

or construct a confidence region for the pair (μ_B, μ_C). In two dimensions, such a region is typically an ellipse rather than an interval.

This illustrates the main advantage of the multivariate Central Limit Theorem. Applying the ordinary Central Limit Theorem separately to blood pressure and cholesterol would describe the distribution of each sample mean, but it would not fully describe how the two sample means vary together. The multivariate theorem preserves this dependence through the covariance matrix and therefore provides a joint approximation for several related quantities.

8. Interesting Things

8.1. **The Central Limit Theorem in Nature.** The Central Limit Theorem helps explain why many characteristics in nature are approximately normally distributed. Human height, for example, is influenced by a large number of genetic and environmental factors. Each individual gene may contribute only a small amount to a person's final height, while nutrition, illness, sleep, and other environmental conditions introduce additional variation. When many small, roughly independent effects are added together, the Central Limit Theorem suggests that their total effect will be approximately normally distributed. This produces the familiar bell-shaped distribution in which most people have heights near the population mean, while very short and very tall individuals are less common. Similar reasoning applies to other biological measurements, including birth weight, blood pressure, cholesterol levels, and the sizes of leaves, fruits, or animals within a species. These quantities are not determined by a single random cause, but by the combined influence of many small factors. Although the assumptions of the Central Limit Theorem are not always satisfied perfectly in biological systems, it provides an important explanation for why normal distributions appear so frequently in nature.

8.2. **The Cauchy Distribution.** The Cauchy distribution is an important example of a distribution for which the usual Central Limit Theorem does not apply. Its probability density function is

$$f(x) = \frac{1}{\pi(1 + x^2)}.$$

Unlike the normal distribution, the Cauchy distribution has extremely heavy tails and has neither a finite mean nor a finite variance. Since the classical Central Limit Theorem requires the random variables to have a finite mean and variance, its assumptions are not satisfied. In fact, if $X_1, \ldots, X_n$ are independent Cauchy random variables, then their sample mean

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

is also Cauchy distributed. Thus, the sample mean does not become more concentrated or approach a normal distribution as the sample size increases. This demonstrates why the finite-moment assumptions in the Central Limit Theorem are essential.

References

[Cur42] J. H. Curtiss. A note on the theory of moment generating functions. *The Annals of Mathematical Statistics*, 13(4):430–433, 1942.

[Ros14] Sheldon M. Ross. *A First Course in Probability*. Pearson, 9 edition, 2014.

Email address: lukejb0001@gmail.com