A Proof of the Central Limit Theorem via Moment Generating Functions

Pranav Krishnapuram

July 13, 2025

Abstract

The Central Limit Theorem (CLT) is one of the most profound and useful results in probability theory and statistics. It describes the remarkable tendency for the sum of a large number of independent random variables to become approximately normally distributed, irrespective of the original distributions. This paper seeks to explain and prove 2 forms of CLT, starting with the more foundational Lindeburg-Levy CLT, and then the more general Lyapunov CLT. Normal distributions are found throughout nature, in places where one would never expect such a nice function. This paper explains why that happens and how the order arises from chaos via CLT's various forms.

1 Moments and Moment Generating Functions

To analyze the behavior of probability distributions, we use moments to describe their shape and characteristics. This acts similarly to taking derivatives to find information about a function indirectly, such as a taylor series. It also has parallels to fourier transforms, by creating a new function in order to more easily describe the original.

Definition 1.1 (Moment)**.** The n^{th} moment of a random variable X about the origin is defined as the expected value of X^n, denoted by:

$$\mu'_n = E[X^n]$$

The first moment, $\mu'_1 = E[X]$, is the mean of the distribution. Higher degree moments give information on other features of the distribution, such as the variance, which is given by $\text{Var}(X) = E[X^2] - (E[X])^2$. We use moments to characterize distributions, as we will see in the final proof of CLT.

A very powerful tool for working with moments is the Moment Generating Function (MGF). They act very similarly to a taylor series expansion or fourier series in their ability to describe a given distribution.

Definition 1.2 (MGF)**.** The moment generating function of a random variable X, denoted $M_X(t)$, is defined as:

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

given that this expectation exists for t in some open interval containing zero.

The utility of the MGF lies in its ability to calculate moments through differentiation. By taking the Taylor series expansion of e^{tX} around $t = 0$, we see that:

$$M_X(t) = E\left[\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right] = \sum_{n=0}^{\infty} \frac{t^n}{n!} E[X^n]$$

This shows that the MGF is a power series where the coefficient of $\frac{t^n}{n!}$ is the n^{th} moment, $E[X^n]$. Consequently, we can find any moment by differentiating the MGF and evaluating it at $t = 0$:

$$\frac{d^n}{dt^n} M_X(t)\bigg|_{t=0} = E[X^n]$$

Because of this, assuming that we can take the MGF of a distribution, we can get information about its general characteristics.

2 Moments and Moment Generating Functions

To analyze and compare probability distributions, we need tools to describe their properties. Moments provide a systematic way to characterize the shape of a distribution, moving from basic location and spread to more subtle features, like asymmetry. There are also functions that use the moments, and they function like Taylor series or Fourier transforms do, at least in terms of how they describe the original distribution.

2.1 Moments: The Descriptors of a Distribution

Definition 2.1 (Moment about the Origin)**.** The n^{th} moment of a random variable X about the origin is defined as the expected value of X^n, denoted by:

$$\mu'_n = E[X^n]$$

The first few moments about the origin are used to define the most common descriptive statistics.

- The 1st moment, $\mu'_1 = E[X]$, is the mean (μ), representing the distribution's center of mass.
- The 2nd moment, $\mu'_2 = E[X^2]$, is used to calculate the variance, $\text{Var}(X) = E[X^2] - (E[X])^2$, which measures the spread of the distribution.

While moments about the origin are fundamental, it is often more insightful to describe a distribution's shape relative to its mean. This leads to the concept of central moments.

Definition 2.2 (Central Moment)**.** The n^{th} central moment of a random variable X with mean μ is defined as:

$$\mu_n = E[(X - \mu)^n]$$

The second central moment, $\mu_2 = E[(X-\mu)^2]$, is the variance. Higher-order standardized moments describe shape, although they are not used nearly as much as the first two:

- **Skewness:** The third standardized moment, $\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$, measures the asymmetry of the distribution. A positive value indicates a longer tail to the right, while a negative value indicates a longer tail to the left.
- **Kurtosis:** The fourth standardized moment, $\gamma_2 = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3$, measures the "tailedness" of the distribution compared to a normal distribution (for which this value is 0). A positive kurtosis (leptokurtic) indicates heavier tails and a sharper peak, while a negative kurtosis (platykurtic) indicates lighter tails and a flatter peak.

2.2 The Moment Generating Function (MGF)

A more sophisticated tool for working with moments is the Moment Generating Function (MGF). It encodes all of a distribution's moments into a single function.

Definition 2.3 (MGF)**.** The moment generating function of a random variable X, denoted $M_X(t)$, is defined as:

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

provided that this expectation exists for t in some open interval containing zero.

The MGF's name comes from its ability to "generate" moments. By taking the Taylor series expansion of e^{tX} around $t = 0$, we see this:

$$M_X(t) = E\left[\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right] = \sum_{n=0}^{\infty} \frac{t^n}{n!} E[X^n]$$

This shows the MGF is a power series where the coefficient of $\frac{t^n}{n!}$ is the n^{th} moment, $E[X^n]$. We can therefore find any moment by differentiating the MGF n times and evaluating the result at $t = 0$:

$$\frac{d^n}{dt^n} M_X(t)\bigg|_{t=0} = E[X^n]$$

Example 1 (Continuous): Let X be an exponential random variable with rate λ, so its pdf is $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. Its MGF is:

$$M_X(t) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{(t-\lambda)x} dx = \frac{\lambda}{\lambda - t}, \quad \text{for } t < \lambda$$

To find the mean, we differentiate once: $M'_X(t) = \frac{\lambda}{(\lambda - t)^2}$. Evaluating at $t = 0$ gives $E[X] = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$.

Example 2 (Bernoulli, Discrete): Let Y be a Bernoulli random variable with probability of success p. So $P(Y = 1) = p$ and $P(Y = 0) = 1 - p$. Its MGF is:

$$M_Y(t) = E[e^{tY}] = e^{t \cdot 1}(p) + e^{t \cdot 0}(1 - p) = pe^t + 1 - p$$

Example 3 (Binomial, Discrete): Let $X \sim \text{Bin}(n, p)$. A binomial random variable can be seen as the sum of n independent and identically distributed Bernoulli random variables,

$Y_1, \ldots, Y_n$. Since $X = Y_1 + \cdots + Y_n$ and the trials are independent, we can use the property for sums of independent variables:

$$M_X(t) = \prod_{i=1}^{n} M_{Y_i}(t) = (pe^t + 1 - p)^n$$

These are some of the most important moment generating functions, and as an exercise to the reader, we can check the first few derivatives and can derive the moments, successfully giving the same formulas we already are familiar with for mean and standard deviation.

---

3 The Normal Distribution

The Central Limit Theorem is interesting precisely because the distribution it converges to, the normal distribution, appears very commonly and has many useful properties. The most important and simple ones are stated below, but the CLT gives another property, about how it appears in so many places.

3.1 Properties of the Normal Distribution

A random variable X following a normal distribution with mean μ and variance σ^2, denoted $X \sim N(\mu, \sigma^2)$, has the following properties:

- **Symmetry:** Its probability density function (pdf) is a symmetric, bell-shaped curve centered at its mean, μ.
- **Parameters:** The distribution is completely defined by its mean μ (which controls location) and its standard deviation σ (which controls spread).
- **Formula:** For any normal distribution, the pdf can be modeled by the function $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, where σ is the standard deviation and μ is the mean.

The special case where $\mu = 0$ and $\sigma^2 = 1$ is called the standard normal distribution, $Z \sim N(0, 1)$. Any normal variable X can be standardized by the transformation $Z = (X - \mu)/\sigma$.

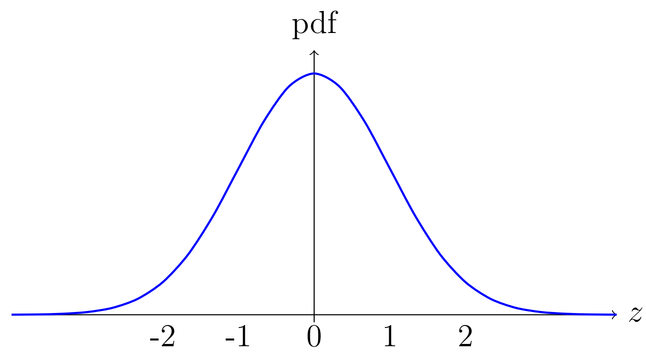

Figure 1: The probability density function of the standard normal distribution $N(0,1)$.

3.2 The MGF of the Standard Normal Distribution

A critical component of the CLT proof is the MGF of its target distribution, $Z \sim N(0,1)$, which has a pdf of $\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$.

Derivation

We compute the MGF by its definition:

$$M_Z(t) = E[e^{tZ}] = \int_{-\infty}^{\infty} e^{tz}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tz-z^2/2}dz$$

To solve the integral, we complete the square in the exponent:

$$tz - \frac{z^2}{2} = -\frac{1}{2}(z^2 - 2tz) = -\frac{1}{2}(z^2 - 2tz + t^2 - t^2) = -\frac{1}{2}(z-t)^2 + \frac{t^2}{2}$$

Substituting this back into the integral:

$$M_Z(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t)^2+\frac{t^2}{2}}dz = e^{t^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z-t)^2}dz$$

The integral term is the integral of the pdf of a normal distribution with mean t and variance 1, which evaluates to 1. Thus, the MGF of the standard normal distribution is:

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

4 Proof of the Central Limit Theorem

We now prove the Lindeberg-Lévy CLT.

Theorem 4.1 (CLT). *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean $E[X_i] = 0$ and variance $Var(X_i) = 1$. Let $S_n = \sum_{i=1}^{n} X_i$. Then the distribution of the random variable $Z_n = \frac{S_n}{\sqrt{n}}$ converges to the standard normal distribution as $n \to \infty$.*

Proof. Our goal is to show that the MGF of Z_n converges to the MGF of the standard normal distribution, $e^{t^2/2}$.

First, we find the MGF of Z_n in terms of the MGF of a single X_i (which we denote $M_X(t)$). Using the properties from Section 2:

$$M_{Z_n}(t) = M_{\frac{1}{\sqrt{n}}\sum X_i}(t) = M_{\sum X_i}\left(\frac{t}{\sqrt{n}}\right) = \left[M_X\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

Next, we expand $M_X(s)$ as a Taylor series around $s = 0$:

$$M_X(s) = E[e^{sX}] = E\left[1 + sX + \frac{s^2X^2}{2!} + \frac{s^3X^3}{3!} + \ldots\right]$$

$$M_X(s) = 1 + sE[X] + \frac{s^2}{2}E[X^2] + r$$

By the theorem's premise, $E[X] = 0$ and $E[X^2] = \text{Var}(X) + (E[X])^2 = 1 + 0^2 = 1$. Substituting these values gives:

$$M_X(s) = 1 + \frac{s^2}{2} + r$$

Now, we let $s = \frac{t}{\sqrt{n}}$ and substitute this into our expression for $M_{Z_n}(t)$:

$$M_X\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + r = 1 + \frac{t^2}{2n} + r$$

Plugging this back into the expression for $M_{Z_n}(t)$:

$$M_{Z_n}(t) = \left[1 + \frac{t^2}{2n} + r\right]^n$$

Finally, we take the limit as $n \to \infty$. We use the well-known limit identity $\lim_{n\to\infty}(1+\frac{x}{n})^n = e^x$. As $n \to \infty$, the higher-order term r vanishes faster than the $\frac{1}{n}$ term, leaving:

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty}\left[1 + \frac{t^2/2}{n}\right]^n = e^{t^2/2}$$

Since the MGF of Z_n converges to the MGF of a standard normal random variable, we conclude by the Uniqueness Property that the distribution of Z_n converges to the standard normal distribution. □

5 The Lyapunov Central Limit Theorem

The true power of the CLT is realized when we relax the "identically distributed" condition. The Lyapunov CLT does this, requiring only that the variables be independent and that no single variable contributes too much to the total variance.

5.1 Statement of the Lyapunov CLT

Consider a sequence of independent random variables $\{X_k\}_{k=1}^n$, each with a finite expected value $E[X_k] = \mu_k$ and a finite variance $\mathrm{Var}(X_k) = \sigma_k^2$. Let $s_n^2 = \sum_{k=1}^n \sigma_k^2$ be the total variance of the sum.

Condition 5.1 (Lyapunov's Condition). *There exists some $\delta > 0$ such that the $(2+\delta)^{th}$ moments of the variables exist ($E[|X_k - \mu_k|^{2+\delta}] < \infty$) and the following limit holds:*

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^n E\left[|X_k - \mu_k|^{2+\delta}\right] = 0$$

This may seem overly technical to use, but this limit effectively determines if the higher order moments grow slower than the total standard deviation, and thus determine if they would stay bounded. The $\frac{1}{s_n^{2+\delta}}$ term holds the sum of the variances of the X_ks, and the $\sum_{k=1}^n E\left[|X_k - \mu_k|^{2+\delta}\right]$ term holds the higher order moments.

Theorem 5.1 (Lyapunov CLT). *Let $\{X_k\}$ be a sequence of independent random variables. If Lyapunov's condition holds for some $\delta > 0$, then the standardized sum*

$$Z_n = \frac{\sum_{k=1}^n (X_k - \mu_k)}{s_n}$$

converges in distribution to a standard normal random variable, $Z \sim N(0,1)$, as $n \to \infty$.

6 Proof of the Lyapunov Central Limit Theorem

The proof of the Lyapunov CLT is more technical than the i.i.d. case, but the strategy remains the same: we show that the MGF of the standardized sum converges to the MGF of a standard normal distribution. For simplicity, we assume without loss of generality that $\mu_k = 0$ for all k.

Proof. Let $Z_n = \frac{1}{s_n}\sum_{k=1}^n X_k$. The MGF of Z_n is the product of the MGFs of the independent terms X_k/s_n:

$$M_{Z_n}(t) = E[e^{tZ_n}] = \prod_{k=1}^n E[e^{tX_k/s_n}] = \prod_{k=1}^n M_{X_k}\left(\frac{t}{s_n}\right)$$

To handle the product, we take the logarithm:

$$\ln M_{Z_n}(t) = \sum_{k=1}^n \ln\left(M_{X_k}\left(\frac{t}{s_n}\right)\right)$$

Now, we expand each MGF term using a Taylor series. For a variable X_k with mean 0, $M_{X_k}(u) = 1 + \frac{1}{2}E[X_k^2]u^2 + R_k(u)$, where $R_k(u)$ is the remainder. Let $u = t/s_n$:

$$M_{X_k}\left(\frac{t}{s_n}\right) = 1 + \frac{\sigma_k^2 t^2}{2s_n^2} + R_k\left(\frac{t}{s_n}\right)$$

Next, we use the Taylor expansion for the logarithm, $\ln(1+x) = x - \frac{x^2}{2} + \dots$. For small x, we can write $\ln(1+x) = x + O(x^2)$. Let $x_k = \frac{\sigma_k^2 t^2}{2s_n^2} + R_k(\frac{t}{s_n})$. As $n \to \infty$, $s_n \to \infty$, so $x_k \to 0$.

$$\ln M_{Z_n}(t) = \sum_{k=1}^n \left[\left(\frac{\sigma_k^2 t^2}{2s_n^2} + R_k\left(\frac{t}{s_n}\right)\right) + O\left(x_k^2\right)\right]$$

Let's analyze the sum term by term.

1. **The Main Term:** The sum of the first part gives the desired result.

$$\sum_{k=1}^n \frac{\sigma_k^2 t^2}{2s_n^2} = \frac{t^2}{2s_n^2}\sum_{k=1}^n \sigma_k^2 = \frac{t^2}{2s_n^2}(s_n^2) = \frac{t^2}{2}$$

2. **The Remainder Terms:** We must show that the sums of all error terms from the Taylor expansions vanish as $n \to \infty$. This is the most technical part of the proof, and it is where Lyapunov's condition is essential.

   Let's first focus on the remainder from the MGF expansion, $\sum R_k(t/s_n)$. A more rigorous version of Taylor's theorem provides a bound on the remainder term. For a variable W with $E[W] = 0$ and $E[|W|^{2+\delta}] < \infty$, the remainder $R(u)$ in the expansion $M_W(u) = 1 + \frac{1}{2}E[W^2]u^2 + R(u)$ can be bounded. A standard inequality shows that for some constant C:

$$|R(u)| \leq C \cdot E[|W|^{2+\delta}] \cdot |u|^{2+\delta}$$

Applying this to our variable X_k with $u = t/s_n$:

$$|R_k(t/s_n)| \leq C \cdot E[|X_k|^{2+\delta}] \cdot \left| \frac{t}{s_n} \right|^{2+\delta}$$

Now we can bound the sum of these remainders:

$$\left| \sum_{k=1}^{n} R_k(t/s_n) \right| \leq \sum_{k=1}^{n} |R_k(t/s_n)| \leq \sum_{k=1}^{n} C \cdot E[|X_k|^{2+\delta}] \cdot \frac{|t|^{2+\delta}}{s_n^{2+\delta}}$$

Factoring out the constants that do not depend on the summation index k:

$$= C|t|^{2+\delta} \left(\frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} E[|X_k|^{2+\delta}] \right)$$

The term in the parenthesis is precisely the expression from Lyapunov's condition. As $n \to \infty$, this term goes to 0. Therefore, the entire sum of the MGF remainder terms vanishes.

A similar, though more involved, argument is used to show that the error terms from the logarithm expansion, $\sum O(x_k^2)$, also converge to zero. Because Lyapunov's condition forces the individual variances σ_k^2 to be small relative to the total variance s_n^2, each x_k term becomes small enough for the sum of their squares to vanish.

With the main term converging to $t^2/2$ and all error terms converging to zero, we have:

$$\lim_{n \to \infty} \ln M_{Z_n}(t) = \frac{t^2}{2}$$

Exponentiating both sides gives the final result:

$$\lim_{n \to \infty} M_{Z_n}(t) = e^{t^2/2}$$

By the Uniqueness Property, since the MGF of Z_n converges to the MGF of a standard normal distribution, the distribution of Z_n must also converge to the standard normal distribution.

$\square$

6.1 Applications where Lyapunov CLT is Necessary

- **Econometrics: Portfolio Returns.** Consider a portfolio of n different assets. The return of each asset, X_k, has its own mean μ_k and variance σ_k^2. The Lindeberg-Lévy CLT does not apply. The Lyapunov CLT can apply if the portfolio is well-diversified, ensuring no single asset's variance dominates. If the condition holds, the total portfolio return can be modeled as approximately normal.

- **Engineering: Sensor Fusion.** Combining measurements from different sensors (e.g., a precise laser and a less precise sonar) results in a sum of independent but not identically distributed errors. The Lyapunov CLT allows us to model the total error as normal, which is crucial for algorithms like the Kalman filter.

7 Applications of the Central Limit Theorem

7.1 The Z-Test for a Population Proportion

One of the most common applications of the CLT is in hypothesis testing for proportions. A sample proportion, $\hat{p}$, is the ratio of successes k to a sample of size n, so $\hat{p} = k/n$. This can be viewed as the mean of n i.i.d. Bernoulli random variables (where success=1, failure=0). The CLT tells us that for a large enough sample size, the sampling distribution of $\hat{p}$ will be approximately normal.

This allows us to construct a z-test. Suppose we want to test the null hypothesis $H_0 : p = p_0$ against an alternative $H_a : p \neq p_0$.

- The mean of the sampling distribution of $\hat{p}$ is p_0.
- The standard deviation (standard error) is $\sqrt{\frac{p_0(1-p_0)}{n}}$.

The CLT justifies using the normal distribution to calculate a p-value. We compute the test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Under the null hypothesis, this z statistic follows an approximate standard normal distribution, provided the sample is large enough (typically checked with $np_0 \geq 10$ and $n(1-p_0) \geq 10$). We can then find the probability of observing a test statistic as extreme as ours to determine whether to reject the null hypothesis.

7.2 The Normal Approximation to the Binomial Distribution

A direct consequence of the CLT is the De Moivre-Laplace theorem, which states that a binomial distribution can be approximated by a normal distribution. A binomial random variable $X \sim \text{Bin}(n, p)$ can be expressed as the sum of n independent Bernoulli trials. Since X is a sum of i.i.d. variables, the CLT applies.

For large n, the distribution of X is approximately normal with the same mean and variance:

- Mean: $\mu = np$
- Variance: $\sigma^2 = np(1-p)$

Thus, we can approximate probabilities for $X \sim \text{Bin}(n, p)$ using a normal distribution $Y \sim N(np, np(1-p))$. This is particularly useful when n is large, as calculating binomial probabilities directly can be computationally intensive. A common rule of thumb is that the approximation is valid when $np \geq 10$ and $n(1-p) \geq 10$.

8 Applications of the Central Limit Theorem

8.1 Confidence Intervals for a Population Mean

One of the most fundamental tasks in statistics is to estimate a population mean μ from a sample mean $\bar{X}$. The CLT is the theoretical foundation for this. It tells us that for a large sample size n, the sampling distribution of $\bar{X}$ is approximately normal, with mean μ and standard deviation $\sigma/\sqrt{n}$ (the standard error). This allows us to construct a confidence interval. A 95% confidence interval, for example, is calculated as:

$$\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

The value 1.96 comes from the standard normal distribution: 95% of the area under the curve lies between -1.96 and +1.96. The CLT guarantees that this formula is valid for large samples even if the original population distribution is not normal.

8.2 The Z-Test for a Population Proportion

A sample proportion, $\hat{p} = k/n$, can be viewed as the mean of n i.i.d. Bernoulli random variables. The CLT tells us that for large n, the sampling distribution of $\hat{p}$ is approximately normal, allowing us to construct a z-test. We compute the test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Under the null hypothesis $H_0 : p = p_0$, this statistic follows an approximate $N(0, 1)$ distribution, allowing for p-value calculation.

8.3 The Normal Approximation to the Binomial Distribution

A binomial variable $X \sim \text{Bin}(n, p)$ is a sum of n i.i.d. Bernoulli trials. By the CLT, for large n, its distribution is approximately normal with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$. This is valid when $np \geq 10$ and $n(1-p) \geq 10$. Because we approximate a discrete distribution with a continuous one, a continuity correction is used. For example, $P(X \leq k)$ is approximated by the area under the normal curve to the left of $k + 0.5$.

Example: Suppose a fair coin is flipped 100 times. What is the probability of getting 60 or more heads?

- Here, $X \sim \text{Bin}(100, 0.5)$. The exact calculation is tedious.
- We check the conditions: $np = 100(0.5) = 50 \geq 10$ and $n(1-p) = 50 \geq 10$. The approximation is valid.
- Mean: $\mu = 50$. Standard deviation: $\sigma = \sqrt{100(0.5)(0.5)} = \sqrt{25} = 5$.
- We want to find $P(X \geq 60)$. Applying the continuity correction, we find the area under the normal curve for $X \geq 59.5$.

- Standardize the value: $z = \frac{59.5-50}{5} = \frac{9.5}{5} = 1.9$.
- Find the probability: $P(Z \geq 1.9) \approx 1 - 0.9713 = 0.0287$.

The normal approximation gives a probability of about 2.87%.

Example 1: Moments and MGFs for a Discrete Distribution

This example demonstrates how to find the MGF of a discrete distribution and use it to calculate the first two moments, and subsequently, the variance.

Let X be a Poisson random variable with rate λ, so its probability mass function (pmf) is $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$. Find its MGF, mean, and variance.

Step 1: Find the MGF. By the definition of the MGF for a discrete variable:

$$M_X(t) = E[e^{tX}] = \sum_{k=0}^{\infty} e^{tk} P(X = k) = \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\lambda}\lambda^k}{k!}$$

We can factor out the constant term $e^{-\lambda}$ and combine the terms with exponent k:

$$M_X(t) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t\lambda)^k}{k!}$$

The sum is the Taylor series expansion for e^z, where $z = \lambda e^t$. Therefore:

$$M_X(t) = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)}$$

Step 2: Find the first moment (Mean). We differentiate the MGF with respect to t and evaluate at $t = 0$.

$$M'_X(t) = \frac{d}{dt}\left(e^{\lambda(e^t-1)}\right) = e^{\lambda(e^t-1)} \cdot (\lambda e^t)$$

$$E[X] = M'_X(0) = e^{\lambda(e^0-1)} \cdot (\lambda e^0) = e^{\lambda(1-1)} \cdot \lambda = e^0 \cdot \lambda = \lambda$$

The mean of the Poisson distribution is λ.

Step 3: Find the second moment. We differentiate a second time and evaluate at $t = 0$. Using the product rule:

$$M''_X(t) = \frac{d}{dt}\left((e^{\lambda(e^t-1)}\lambda e^t)\right) = (e^{\lambda(e^t-1)}\lambda e^t)(\lambda e^t) + (e^{\lambda(e^t-1)})(\lambda e^t)$$

$$E[X^2] = M''_X(0) = (e^0 \cdot \lambda e^0)(\lambda e^0) + (e^0)(\lambda e^0) = (\lambda)(\lambda) + \lambda = \lambda^2 + \lambda$$

Step 4: Calculate the variance. Using the computational formula for variance:

$$\text{Var}(X) = E[X^2] - (E[X])^2 = (\lambda^2 + \lambda) - (\lambda)^2 = \lambda$$

This confirms the well-known result that both the mean and variance of a Poisson distribution are equal to λ.

Example 2: Applying the Lindeberg-Lévy CLT

This example shows how to use the classical CLT to approximate a probability for a sum of i.i.d. non-normal random variables.

Let $X_1, X_2, \ldots, X_{48}$ be independent random variables, each following a Uniform distribution on the interval $[0, 2]$. Approximate the probability that their sum is greater than 58.

Step 1: Find the mean and variance of a single variable. For a single $X_i \sim \text{Uniform}(0, 2)$:

$$E[X_i] = \mu = \frac{0+2}{2} = 1$$

$$\text{Var}(X_i) = \sigma^2 = \frac{(2-0)^2}{12} = \frac{4}{12} = \frac{1}{3}$$

Step 2: Define the sum and find its parameters. Let $S_{48} = \sum_{i=1}^{48} X_i$. Since the variables are i.i.d., the mean and variance of the sum are:

$$E[S_{48}] = n\mu = 48 \cdot 1 = 48$$

$$\text{Var}(S_{48}) = n\sigma^2 = 48 \cdot \frac{1}{3} = 16$$

The standard deviation of the sum is $\sigma_{S_{48}} = \sqrt{16} = 4$.

Step 3: Apply the Central Limit Theorem. The CLT states that for a large n (48¿30, so 48 is generally considered large enough), the distribution of the sum S_{48} is approximately normal:

$$S_{48} \approx N(\mu = 48, \sigma^2 = 16)$$

Step 4: Calculate the desired probability. We want to find $P(S_{48} > 58)$. We standardize this value by converting it to a Z-score:

$$Z = \frac{\text{Value} - \text{Mean}}{\text{Standard Deviation}} = \frac{58 - 48}{4} = \frac{10}{4} = 2.5$$

So, $P(S_{48} > 58) \approx P(Z > 2.5)$. Using a standard normal table or calculator:

$$P(Z > 2.5) = 1 - P(Z \leq 2.5) \approx 1 - 0.9938 = 0.0062$$

Therefore, the probability that the sum of 48 such uniform random variables exceeds 58 is approximately 0.62%.

Example 3: A Scenario Requiring the Lyapunov CLT

This example illustrates a situation where variables are independent but not identically distributed, making the classical CLT inapplicable but allowing for the use of the Lyapunov CLT.

An electronics system receives a sequence of n signals. The k^{th} signal, X_k, is a random variable representing noise. The sensors improve over time, so the variance of the noise decreases. Suppose X_k are independent with $E[X_k] = 0$ and $\text{Var}(X_k) = \sigma_k^2 = \frac{1}{k}$. Can we describe the distribution of the total noise, $S_n = \sum_{k=1}^{n} X_k$, for large n?

Step 1: Check the conditions for Lindeberg-Lévy CLT. The variables are independent but not identically distributed, since their variances are different ($\sigma_1^2 = 1, \sigma_2^2 = 1/2, \dots$). Therefore, the classical CLT does not apply.

Step 2: Check Lyapunov's Condition. We must check if Lyapunov's condition holds. Let's choose $\delta = 1$ and assume that the third absolute moments exist and are bounded, say $E[|X_k|^3] \leq M$ for some constant M. First, we need the total variance, $s_n^2 = \sum_{k=1}^{n} \sigma_k^2 = \sum_{k=1}^{n} \frac{1}{k}$. This is the harmonic series, which for large n is approximately $\ln(n)$. So, $s_n^2 \approx \ln(n)$. Now, we check the limit from Lyapunov's condition with $\delta = 1$:

$$\lim_{n \to \infty} \frac{1}{s_n^3} \sum_{k=1}^{n} E[|X_k|^3]$$

Substituting our values:

$$\lim_{n\to\infty} \frac{1}{(\ln(n))^{3/2}} \sum_{k=1}^{n} E[|X_k|^3] \leq \lim_{n\to\infty} \frac{1}{(\ln(n))^{3/2}} \sum_{k=1}^{n} M = \lim_{n\to\infty} \frac{nM}{(\ln(n))^{3/2}}$$

This limit diverges to infinity, so the condition is not met with this assumption. However, if the third moments decreased sufficiently fast (e.g., $E[|X_k|^3] \propto 1/k^2$), the sum would converge, and the overall limit would be zero.

Conclusion: This demonstrates the role of Lyapunov's condition. It provides a specific mathematical check to ensure that the variances of the later terms in the sequence are not so insignificant that the earlier, more variable terms dominate the sum entirely. If the condition were met, we could conclude that the standardized total noise, $\frac{S_n}{\sqrt{\ln(n)}}$, is approximately standard normal.

Example 4: Application - Z-Test for Proportions

This example shows a full hypothesis testing problem that relies on the CLT.

A pharmaceutical company claims that its new drug is effective in 80% of cases. A clinical trial is conducted with 200 patients, and the drug is found to be effective for 150 of them. Is there significant evidence at the $\alpha = 0.05$ level to suggest the company's claim is inaccurate?

Step 1: State the hypotheses. The null hypothesis is that the company's claim is true. The alternative is that it is not.

$$H_0 : p = 0.80$$

$$H_a : p \neq 0.80$$

Step 2: Calculate the sample proportion and check conditions. The sample proportion is $\hat{p} = \frac{150}{200} = 0.75$. We check the conditions for the CLT to apply: $np_0 = 200(0.80) = 160 \geq 10$ and $n(1 - p_0) = 200(0.20) = 40 \geq 10$. The conditions are met.

Step 3: Calculate the test statistic. The standard error under the null hypothesis is $\sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.80(0.20)}{200}} = \sqrt{\frac{0.16}{200}} \approx 0.0283$. The z-statistic is:

$$z = \frac{\hat{p} - p_0}{\text{SE}} = \frac{0.75 - 0.80}{0.0283} = \frac{-0.05}{0.0283} \approx -1.77$$

Step 4: Find the p-value and make a conclusion. We are performing a two-tailed test, so we need to find the probability of getting a result as extreme as -1.77 or more extreme in either direction: $P(Z \leq -1.77) + P(Z \geq 1.77)$.

$$P(Z \leq -1.77) \approx 0.0384$$

The total p-value is $2 \times 0.0384 = 0.0768$.

Since the p-value (0.0768) is greater than the significance level α (0.05), we fail to reject the null hypothesis. There is not sufficient statistical evidence to conclude that the company's claim of 80% effectiveness is inaccurate.

Conclusion

The method of using moment generating functions provides an elegant pathway to proving the Central Limit Theorem, given that moments fully define their distributions. We study MGFs to indirectly compare distributions, and by applying them correctly, we have proved one of the most famous results in statistics and probability theory.

A Standard Normal Distribution Table

The table below provides the cumulative probability for the standard normal distribution, $P(Z \leq z)$, for positive values of z. For negative values, use the identity $P(Z \leq -z) = 1 - P(Z \leq z)$.

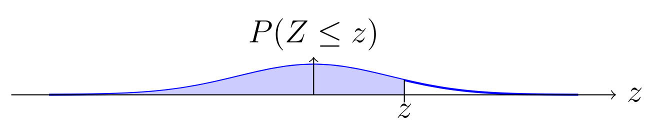

Figure 2: Area under the curve for a given z-score.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

References

[1] Yuval Filmus 2010. *Two Proofs of the central limit theorem.*

[2] Wikipedia (2025, July 12). *Central Limit Theorem.* Retrieved from en.wikipedia.org/wiki/Central_limit_theorem

[3] Wikipedia (2025, July 12). *Moment (Mathematics).* Retrieved from en.wikipedia.org/wiki/Moment_(mathematics)