Alma Mater Studiorum
Università di Bologna

Department of Statistical Sciences "Paolo Fortunati"

First Cycle Degree/Bachelor in Statistical Sciences
Curriculum Stats&Maths

Contextual Analysis of the Central Limit Theorem:

Exploring Inspirations, Significance, and Mathematical Developments

Presented by:
Sona Yavrumyan
ID: 0000967073

Supervisor:
Prof. Christian Martin Hennig

Academic Year 2022/2023
Session I

Contents

Chapter 1

Introduction

The Central Limit Theorem (CLT) refers to a group of theorems that describe the convergence of distributions, densities, or discrete probabilities. First labelled so by Hungarian mathematician George Pólya in 1920, it occupies a fundamental role within the fields of statistics and probability, bridging the gap between theoretical constructs and practical applications. Its influences span a broad array of disciplines, including finance, physics, computer science, data analysis, business and economics, etc. Despite its large influence, the evolution and historical context of this theorem tend to be overlooked in academic discourse. Through this thesis, my primary aim is to embark on a thorough exploration of the Central Limit Theorem, enabling a clear contextualization and understanding of the various forms of it. To promote a comprehensive understanding, this thesis commits to adopting a unified mathematical notation.

The commencing part of this work acknowledges the seminal contributions of early pioneers in the field, such as Jacob Bernoulli, Abraham de Moivre, and Pierre Simone Laplace. These individuals set the foundation with the formulation of the earliest limit theorems expanding upon each other's work and improving the newfound field of probability in the period of the 17-19th century. While these initial theoretical models served as a robust starting point to the formulation of what's known as "The Classical Central Limit Theorem" in the literature, they were critiqued for their lack of formal mathematical rigour by 20th century probabilists to come.

Succeeding mathematicians like Pafnuty Chebyshev and his students Aleksandr Lyapounov, and Andrey Markov took the lead in developing analytical and mathematically sound proofs using a newer formulation distinguishing the CLT as a proper limit theorem. Their significant contributions solidified the theorem's standing beyond classical probability. Other notable mathematicians, such as Jarl Lindeberg and Paul Lévy, presented further refinements extending the theorem beyond the Gaussian distribution and previous assumptions, introducing entirely new variations of the CLT.

Within this thesis, the Central Limit Theorem will also briefly be viewed from both Bayesian and frequentist frameworks. This dual examination aims to highlight the theorem's versatility and utility across different methodological approaches. Throughout this writing, I further aim to delve into the extensive implications and applications of the Central Limit Theorem, the theorem's integral role in shaping theoretical foundations and impact across various fields it is applicable in.

Ultimately, this thesis is an effort to underscore the enduring relevance of the Central Limit Theorem, not just as a mathematical statement about random variables, but as an indispensable tool in our understanding and interpretation of the world. Thus, this exploration extends beyond just understanding the theorem to appreciating the historical, philosophical, and practical contexts that define its significance.

1.1 Commonly Used Mathematical Notations

$X_1, X_2, \ldots, X_n$	Sequence of i.i.d. random variables
μ	Population mean or expected value of each random variable X_i
σ	Population standard deviation or standard deviation of each random variable X_i
$\overline{X}_n$	Sample mean, defined as the average of the n random variables $\{X_1, X_2, \ldots, X_i, \ldots, X_n\}$
$\sqrt{n}$	Square root of the sample size n
$\Phi(x)$	CDF of the standard normal distribution
$\phi(x)$	PDF of the standard normal distribution
Z	Standard normal random variable, with mean $\mu = 0$ and standard deviation $\sigma = 1$
Z_n	Standardized random variable, defined as $(\bar{X}_n - \mu)/(\sigma/\sqrt{n})$
$\lim_{n\to\infty}$	Limit as n approaches infinity
$\to$	Convergence in distribution symbol
$N(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
S_n	Sum of the n random variables $S_n = X_1 + X_2 + \ldots + X_n$
$Var(X)$	Variance of a random variable X
$E(X)$	Expected value or mean of a random variable X
$Pr(E)$	Probability of an event E

Chapter 2

The Classical Central Limit Theorem

2.1 Evolution of Limit Theorems: Foundations and Predecessors

The history of the Central Limit Theorem can be traced back to the 18th century, in the works of Swiss mathematician Jacob Bernoulli and French mathematician Abraham de Moivre. Although in their exploration of probability theory they were both motivated by using it to solve real-world problems and analyze uncertainties in fields such as insurance, finance, and games of chance, it was through their pioneering research that the foundation for the CLT was laid.

Jacob Bernoulli's Law of Large Numbers (LLN), introduced in his *Ars Conjectandi* ("The Art of Conjecturing") in 1713, was a significant development in the field of probability theory. The LLN, in its simplest form, states that as the number of independent, identically distributed random trials increases, the sample mean converges to the population mean. Mathematically, this is represented as: [1]

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\mu, \tag{2.1}$$

where X_i represents the i-th observation, n is the number of observations, and μ is the population mean. [11]

In other words, it emphasizes that the more observations you have, the closer your estimate of the population's true parameters (like the mean) will be. Bernoulli's LLN was a robust theorem, but its applicability was somewhat restricted due to the lack of precision in estimating the minimum number of observations required for the law to hold true.

Abraham de Moivre sought to refine Bernoulli's Law of Large Numbers. His 1733 paper, *Approximatio ad summam terminorum binomii* $(a + b)^n$ *in seriem expansi* ("Approximation to the Sum of the Terms of the Binomial $(a + b)^n$") which he shared with a close circle, aimed to offer a more precise approximation to the binomial distribution than what Bernoulli had achieved. This paper, while focused on a specific case where the probability of success was 0.5, was a significant leap forward. He developed an approach for approximating the binomial distribution, which offered insights into the underlying regularities in large sets of numbers. Specifically, for large n, he provided the approximations: [1]

$$P\left(Z = \left[\frac{n}{2}\right] + i\right) \approx \frac{2}{\sqrt{2\pi n}} e^{-2i^2/n}, \tag{2.2}$$

where $P(Z = \frac{n}{2} + i)$ represents the probability of getting $\frac{n}{2} + i$ successes in n trials, i is the number of deviations from the expected number of successes $\frac{n}{2}$, and n is the number of trials.
This statement, as seen in Equation 2.3 can be viewed as a rudimentary form of the modern "local" limit theorem:

$$P(|n \cdot p - k| < \sqrt{n}) \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-(k-np)^2/(2np(1-p))}, \tag{2.3}$$

where $P(|n \cdot p - k| < \sqrt{n})$ represents the probability that the number of successes is within $\sqrt{n}$ of its expected value np, p is the probability of success, k is the actual number of successes.
De Moivre's method was heavy, requiring approximations and series expansions, but it marked a significant step towards the formulation of the CLT. His observations were not just limited to the special case but also had implications for a general probability of success.

These foundational theorems by Bernoulli and de Moivre set the stage for future developments in probability theory. It should be noted that probabilities for sums of independent random variables already had a significant role in 18th century probability theory, especially in relation to games of chance and the emerging field of error theory. [6] [10]

Even though these early approaches were far from the final form of the CLT, they provided an essential foundation. De Moivre's "limit" theorems and Bernoulli's Law of Large Numbers were stepping stones, serving as a precursor to Laplace's work, which represented a substantial leap forward in understanding the underlying universality of statistical distributions.

2.2 Laplace's Formulation, Assumptions, and Shortcomings

Pierre-Simon Laplace, a prominent French mathematician and astronomer of the 18th and 19th centuries, was a pivotal figure in the development of the Central Limit Theorem. His work on the theorem was marked by a keen application of mathematical rigour on astronomical questions about planetary orbits that were earlier of captivating academic interest to John Bernoulli. In his investigations into the orbits of comets, Laplace applied probability theory to propose that the orbital inclinations of these celestial bodies occurred randomly and were equally likely. His work, however, was not confined to this specific issue. He further extended his research to the question of determining the probability that the mean of a large number of observational errors falls within specific limits. [1]

Laplace's main contributions to the Central Limit Theorem date back to 1810, when he published a memoir titled *Mémoire sur les approximations des formules qui sont fonctions de très grands nombres, et sur leur application aux probabilités* ("Memoir on Approximations of Formulas that are Functions of Very Large Numbers, and Their Application to Probabilities"). In this memoir, he used integrals of type e^{-x^2} [1] to approximate certain probabilities, which signaled a significant step towards the development of the Central Limit Theorem.

His work in probability theory didn't stop there. In 1812, he issued the *Théorie analytique des probabilités* ("Analytical Theory of Probabilities"), a seminal work in which he established many fundamental results in statistics. The treatise was divided into two parts: the first half dealt with methods and problems in probability, and the second half focused on statistical methods and applications. Despite his perspective sometimes oscillating between Bayesian and non-Bayesian frameworks (not clearly distinguishable at the time), his conclusions have remained fundamentally sound. [1]

Later, in 1819, Laplace published a popular account of his work on probability, which served as an accessible introduction to the complex ideas presented in his *Théorie analytique* ("Analytical Theory"). Despite the contemporary view that focused on the practical applicability of probabilistic problems, Laplace emphasized the analytical importance of such problems, especially in the context of "approximations of formula functions of large numbers".

His approach to approximation was based on transforming terms dependent on a large number, n, into series expansions. The quality of the approximation was measured by the number of calculated terms and the rate at which these terms decreased with increasing n.

Laplace's work[6] focused on identically distributed random variables $(X_1, ..., X_n)$ with zero means. These variables took on values $\frac{k}{m}$ (where m is a natural number, and k ranges from $-m$ to m) with respective probabilities p_k.
He used the generating function $T(t) = \sum_{k=-m}^{m} p_k t^k$. Because the X_l's are independent, the probability P_j (that the sum of the X_l's equals $\frac{j}{m}$) is equal to the coefficient of t^j in $(T(t))^n$ after expanding this expression.

To simplify the computation, Laplace substituted t with e^{ix} (where $i = \sqrt{-1}$), thereby introducing what we now call the characteristic function. Laplace then showed that the probability $P(j)$ can be calculated as a certain integral involving the characteristic function raised to the power n. To approximate this integral, he expanded the integrand around its maximum at $x = 0$, but only the part involving the power with exponent n in the following way: [1]

Starting with the characteristic function

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}tx}\mathrm{e}^{\mathrm{i}sx}dx = \delta_{ts} \quad (t, s \in \mathbb{Z})^{2} \tag{2.4}$$

the probability function can be written as

$$P(j) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}jx}\left[\sum_{k=-m}^{m} p_k \mathrm{e}^{\mathrm{i}kx}\right]^n dx. \tag{2.5}$$

Next, a Taylor series expansion can be performed on the complex exponential term e^{ikx}, which gives us:

$$e^{ikx} = 1 + ikx - \frac{k^2x^2}{2} - \frac{ik^3x^3}{6} + \cdots$$

Substituting this expansion into the original expression for $P(j)$, we get the following integral:

$$\begin{aligned} P(j) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}jx}\left[\sum_{k=-m}^{m} p_k \mathrm{e}^{\mathrm{i}kx}\right]^n dx \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}jx}\left[\sum_{k=-m}^{m} p_k \left(1 + \mathrm{i}kx - \frac{k^2x^2}{2} - \frac{\mathrm{i}k^3x^3}{6} + \cdots\right]^n dx. \end{aligned}$$

Taking into account the condition that the sum of the probabilities multiplied by their respective values, from $-m$ to m, equals zero. $\sum_{k=-m}^{m} p_k k = 0$, and performing the substitution $m^2\sigma^2 = \sum_{k=-m}^{m} p_k k^2$ relating the mean square deviation to the sum of the probabilities multiplied by the square of their respective values, we deduce the following result:

$$P(j) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}jx}\left[1 - \frac{m^2\sigma^2x^2}{2} - \mathrm{i}Ax^3 + \cdots\right]^n dx,$$

where A is an introduced constant determined by $\sum_{k=-m}^{m} p_k k^3$ [3].

The logarithm of the expression in the integral is then expanded into a series, which is exponentiated to give:

[1] It is important to note that the notation and mathematical logic do not exactly correspond to those of Laplace's in "Théorie analytique des probabilités", and are rather simplified for ease of understanding.

[2] The notation δ_{ts} in Equation 2.4 refers to the Dirac delta function (in this context also - Kronecker delta function), which is a function of two variables, t and s, equal to 1 if the variables are equal and 0 otherwise. The equation is essentially stating that the integral of the product of two complex exponential functions (which are each a function of x, t, and s) over the interval $[-\pi, \pi]$ is equal to 1 if $t = s$ and 0 if $t \neq s$.

[3] The expansion can be continued by applying the binomial theorem to the expression inside the integral, and introducing new constants for each following term.

$$\log\left[1-\frac{m^2\sigma^2x^2}{2}-\mathrm{i}Ax^3+\cdots\right]^n =: \log z.$$

Followed by

$$\log z = -\frac{m^2\sigma^2nx^2}{2}-\mathrm{i}Anx^3+\cdots$$

further transformed to

$$z=\mathrm{e}^{-\frac{m^2\sigma^2nx^2}{2}-\mathrm{i}Anx^3+\cdots}=\mathrm{e}^{-\frac{m^2\sigma^2nx^2}{2}}\left(1-\mathrm{i}Anx^3+\cdots\right).$$

This allowed Laplace to then express $P(j)$ in the following way:

$$P(j)=\frac{1}{2\pi\sqrt{n}}\int_{-\pi\sqrt{n}}^{\pi\sqrt{n}}\mathrm{e}^{-\mathrm{i}j\frac{y}{\sqrt{n}}}\mathrm{e}^{-\frac{m^2\sigma^2y^2}{2}}\left(1-\frac{\mathrm{i}Ay^3}{\sqrt{n}}+\cdots\right)dy.$$

This process simplifies the integral and allows for change of variables. By substituting $x=y/\sqrt{n}$, the integral can be approximated for large n, as demonstrated in Equation 2.6:

$$P(j)\approx\frac{1}{2\pi\sqrt{n}}\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}j\frac{y}{\sqrt{n}}}\mathrm{e}^{-\frac{m^2\sigma^2y^2}{2}}dy, \tag{2.6}$$

which with further algebraic steps can be shown to be equivalent to

$$\frac{1}{m\sigma\sqrt{2\pi n}}\mathrm{e}^{-\frac{j^2}{2m^2\sigma^2n}}.$$

Summing up the above result for $\frac{j}{m}\in\left[r_1\sqrt{n};r_2\sqrt{n}\right]$ and considering the values r_1 and r_2 that define the integration interval for Laplace's integral approximation, results in the following:

$$\begin{aligned}P\left(r_1\sqrt{n}\le\sum X_l\le r_2\sqrt{n}\right)&\approx\sum_{j\in\left[mr_1\sqrt{n};mr_2\sqrt{n}\right]}\frac{1}{m\sigma\sqrt{2\pi n}}\mathrm{e}^{-\frac{j^2}{2m^2\sigma^2n}}\\&\approx\int_{mr_1}^{mr_2}\frac{1}{m\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2m^2\sigma^2}}dx=\int_{r_1}^{r_2}\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2\sigma^2}}dx.\end{aligned}\tag{2.7}$$

After several transformations and approximations, Laplace arrived at the integral form of the CLT as shown in Equation 2.7, which states that the sum of independent and identically distributed random variables (when suitably normalized) converges in distribution to a standard normal distribution.

This derivation was initially carried out for discrete random variables, but Laplace then extended it to the continuous case by letting m approach infinity.

Since Laplace's interest in the CLT was purely sparked by his further studies in astronomy, there is not any proper formulation of a generalized CLT that would correspond to his version. However a modernized notation can be applied, which states that if we have a sequence of independent and identically distributed random variables, denoted as $X_1, X_2, \ldots, X_n$, with a finite mean m and a finite variance s^2 (where s is greater than zero), then as the sample size n increases, the standardized sum of these variables, given by [1]

$$Z_n = \frac{(X_1 + X_2 + \ldots + X_n) - nm}{s\sqrt{n}}, \tag{2.8}$$

converges in distribution. In particular, the probability of Z_n being less than or equal to a given value t approaches the standard normal distribution function as n tends to infinity. [9]

Despite how large of a milestone this marked, Laplace's proof technique involved creating the idea of the characteristic function and using it to get the CLT in a special case which meant his proof was more of a proof sketch. He assumed without comment that all the moments of a distribution exist and are finite, allowing him to expand the characteristic function. His work was primarily focused on discrete random variables with finite variance. [4]

Laplace also attempted to extend his result to continuous random variables using further approximation argument, but this part of his proof sketch couldn't be made rigorous. This gap was later filled by Simeon de Poisson - a French mathematician and astronomer, who established the first CLT for continuous random variables with finite range of support.

Hald (1998) [4] summarized three main ways in which the work on the CLT, including Laplace's contributions, fell short of contemporary understanding:

- It failed to extend the CLT to continuous individual observations with infinite support.
- It didn't identify precise moment conditions under which the CLT is true.
- It didn't evaluate the remainder terms in their series expansions with sufficient accuracy to pin down the rate of convergence of the distribution of the sample mean to the Normal distribution.

2.3 Developments in the Classical Framework on the Pathway to Abstract Probability

As the 19th century drew to a close, the rise of mathematics in the West as a distinct profession led to a clear division between it and the physical and actuarial sciences. This shift had significant implications for the field of mathematics. No longer did it lean on its practical applications as a measure of its validity. Consequently, there was a substantial increase in the rigour of mathematical proof and argumentation, with the field developing remarkably higher standards of precision and thoroughness. [8]

In the context of the CLT, which had so far been approached as a mere tool for practical approximations and derivations, the forthcoming generation of mathematicians would provide a fresh perspective and a clear shift to a more abstract approach. Pierre-Simon Laplace, for instance, who worked on his own version of the CLT, viewed exceptions to the normal distribution as singularities that wouldn't occur "in practice". In contrast, other notable mathematicians like Augustin-Louis Cauchy and Johann Dirichlet, representatives of this new mathematical conception, would see these as boundaries defining the limits of the concept. [1]

The Central Limit Theorem therefore would signify its unique place in probability theory. The first stride into a more abstract formulation was noted in Russian mathematician Pafnuty Chebyshev's work. His approach was unique as his research did not focus on further improving on what Laplace had already developed, but to rigorously reformulate probability theories, which resulted in an entirely novel direction. [1]

Chapter 3

The Modern Approach and Formulation of The Central Limit Theorem

3.1 Historical Context for 20th Century Foundations of the Central Limit Theorem

The formulation of the Central Limit Theorem as a distinct and vital component of probability theory owes a great deal to the pioneering work of Pafnuty Chebyshev as noted in *Chapter 2*. His insightful contributions significantly shaped the theorem's development and inspired a new generation of mathematicians. Two of Chebyshev's most gifted students - Andrey Markov and Aleksandr Lyapunov, were amongst those. Carrying forward Chebyshev's intellectual legacy, they pursued research along similar lines, making significant strides not only in the advancement of the Central Limit Theorem but also in the broader field of probability theory, recognized as some of the greatest mathematicians of the 20th century. The efforts of the St. Petersburg School of mathematicians [14] brought the theorem to completion with Lyapunov's third and final formulation of it. [1]

He proved that the CLT holds if the random variables under consideration have a finite second moment and their characteristic functions converge to that of the standard normal distribution, which allowed for the inclusion of a broader range of distributions than earlier versions. For instance, the Lyapunov CLT applies even when the random variables aren't identically distributed, provided they satisfy certain conditions (the Lyapunov condition) [8] related to their moments and moment-generating functions. This was a departure from the classical versions of the theorem, which required the variables to be identically distributed and have finite mean and variance. His approach was especially unique in a sense that across his three formulations of the CLT he focused on relaxing the conditions under which it holds as much as possible, paving a way for the modern probability theory.

After the First World War, several notable Western mathematicians, each developed different versions of the Central Limit Theorem. A remarkable trait shared by these influential probabilists of the modern era, such as Lévy, Lindeberg, and Sergei Bernstein, was that their primary training was not specifically in probability theory. Instead, they brought their extensive experience in analytical research into this emergent field. They incorporated analytical methods from various domains, including differential and integral equations, Fourier analysis, the analytical theory of numbers, and measure and integration theory. [3]

The quest to find the necessary and sufficient conditions for the approximation of laws of sums of random variables by Gaussian distributions had been a long-standing problem in probability theory by the mid 20th century and significant advances were made in this search which will briefly be discussed in the following sections.

3.2 Lindeberg's Contributions and Lindeberg's Condition

In 1922, mathematician Harald Cramér's student F. Lindeberg, presented a proof of the CLT that extended the theorem beyond the case of identically distributed random variables. Lindeberg's proof introduced the concept of Lindeberg's condition, which focused on the variances and tail behaviors of the individual random variables. This provided a sufficient condition for the convergence in distribution of the sum of independent random variables to a normal distribution. A simplified outline of the proof, maintaining Laplace's notation (*Chapter 2*) notation is highlighted below. [3] [6]

Theorem 3.2.1. *Consider a sequence of independent random variables $X_1, X_2, \ldots, X_n$ with mean zero and finite variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$. Our goal is to show that the sum of these variables, denoted as $S_n = X_1 + X_2 + \ldots + X_n$, converges to a normal distribution as n approaches infinity.*

Proof outline. To begin, let's define the characteristic function of a random variable X as $\phi_X(t) = \mathrm{E}\left[e^{itX}\right]$, where i represents the imaginary unit. The characteristic function uniquely determines the distribution of X.

We want to analyze the characteristic function of the standardized sum $\frac{S_n}{\sqrt{s_n^2}}$, where $s_n^2 = \sum_{i=1}^{n} \sigma_i^2$. Using the independence of the variables, we can express the characteristic function of S_n as the product of the characteristic functions of the individual variables as shown in Equation 3.1:

$$\phi_{S_n}(t) = \prod_{i=1}^{n} \phi_{X_i}\left(\frac{t}{\sqrt{s_n^2}}\right). \tag{3.1}$$

Now, let's consider the logarithm of the characteristic function:

$$\log \phi_{S_n}(t) = \sum_{i=1}^{n} \log \phi_{X_i}\left(\frac{t}{\sqrt{s_n^2}}\right).$$

We can expand the logarithm using the Taylor series:

$$\log \phi_{S_n}(t) = \sum_{i=1}^{n} \frac{t}{\sqrt{s_n^2}} \mathrm{E}\left[X_i\right] - \frac{t^2}{2s_n^2} \mathrm{E}\left[X_i^2\right] + R_n(t), \tag{3.2}$$

where $R_n(t)$ represents the remainder term.

Since the variables have mean zero, the first term in the expansion disappears. Additionally, since the variables have finite variances, we can express the second term as:

$$\frac{t^2}{2s_n^2} \sum_{i=1}^{n} \sigma_i^2 = \frac{t^2}{2}.$$

Hence, Equation 3.2 will look like:

$$\log \phi_{S_n}(t) = -\frac{t^2}{2} + R_n(t).$$

Now, let's analyze the remainder term $R_n(t)$. By applying Lindeberg's condition[1], we can show that as n approaches infinity:

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \mathrm{E}\left[X_i^2 \cdot \mathbf{1}_{\{|X_i| > \epsilon s_n\}}\right] = 0, \tag{3.3}$$

for any small $\epsilon > 0$.

Using this condition, we can demonstrate that the remainder term $R_n(t)$ converges to zero as n grows. This convergence ensures that the characteristic function of S_n approaches:

$$\phi_{S_n}(t) = e^{-\frac{t^2}{2}}. \tag{3.4}$$

The characteristic function $e^{-\frac{t^2}{2}}$ corresponds to the standard normal distribution.

[1]The Lindeberg condition states:

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \int_{|x-\mu_i| > \epsilon s_n} (x - \mu_i)^2 f_i(x) dx = 0,$$

where n is the number of terms in the sum, $f_i(x)$ is the probability density function of the i-th random variable, μ_i is the mean of the i-th random variable, s_n^2 is the variance of the sum of the random variables.

This means that as n grows large, the sum of the variances of those random variables which deviate significantly from their means (specifically, those that deviate more than ε times the standard deviation of the sum away from their means) becomes negligible compared to the total variance.

So, if Lindeberg's condition holds, then the central limit theorem guarantees that the distribution of the sum of these random variables approaches a normal distribution as n becomes large. [8]

Therefore, as n approaches infinity, the distribution of $\frac{S_n}{\sqrt{s_n^2}}$ converges to the standard normal distribution. □

Lindeberg's proof was revolutionary in its simplicity and generality, noted by many mathematicians, including Lévy who admitted that [3] Lindeberg's proof and usage of characteristic equations were superior to his.

3.3 Relaxation of the Identical Distribution Requirement

Up until the 1930's, all formulations of the Central Limit Theorem illustrated how the sum of a large number of independent and identically distributed random variables converges to a normal distribution. However, the French mathematician Paul Lévy offered a more general perspective of this theorem, extending its scope beyond the constraints of identical distribution and normality.

Lévy built upon the work of Pólya, who had previously established that the characteristic function of a probability distribution could be expressed as the exponential of another function denoted as $\psi(x)$. However, Pólya's method was constrained to the cases where a parameter $\alpha < 1$. Lévy's genius was to extend Pólya's work to all values of a second parameter β and α excluding 1 and 2. He accomplished this by establishing the existence of a probability density function f with a characteristic function φ that satisfies an interesting convergence property. As the sample size increases, the characteristic function of the scaled variable, raised to the power of n, approaches the exponential of $\psi(t)$. Lévy showed that it happens uniformly within any finite interval of t values. This means the rate at which φ approaches $e^{\psi(t)}$ is consistent across the entire t value range within a given interval.[2]

Lévy's work also dealt with the restrictions on the parameters α and β. An error in Lévy's original proof claimed that the absolute value of β was less than or equal to 1 for values of α excluding 1 and 2. This issue was finally resolved in a later paper.

Building up on these results, he devised a generalized theorem that allows for the convergence of the distributions of normalized sums of independent, but not necessarily identically distributed,

[2]Fischer (2011) includes Lévy's general theorem concerning the convergence of the distributions of appropriately scaled sums of independent random variables. These random variables belonged to the same set of reduced laws, denoted as $\mathscr{L}\alpha, \beta$. The reduced laws were characterized by the logarithm $\psi(t)$ of their characteristic functions, which satisfied a specific condition near $t = 0$ outlined in the equation $\psi(t) = -(c_0 + c_1 \operatorname{sgn}(t)i)|t|^\alpha[1 + \omega(t)]$. Here, c_0, c_1, and $\omega(t)$ had specific properties, such as $\lim_{t\to 0} \omega(t) = 0$. Furthermore $\operatorname{sgn}(t)$ takes a real number t as input and returns -1 if t is negative, 0 if t is zero, and 1 if t is positive. A reduced law was defined as one where $c_0\Gamma(\alpha + 1) = 1$.

Lévy's major theorem stated that if we consider a set of independent random variables, each belonging to the set of reduced laws $\mathscr{L}\alpha, \beta$, and appropriately scale their sums using positive numbers $a_1, a_2, \ldots, a_n$, where $A^\alpha = \sum_{k=1}^n a_k^\alpha$ then the normalized sum $X/A = (a_1\xi_1 + a_2\xi_2 + \cdots + a_n\xi_n)/A$ obeys a probability law that approaches the reduced law $L_{\alpha,\beta}$ as the "very small" number η becomes smaller, where η represents a small positive number,ξ_1 is a random variable in the sum, and $\Gamma(\alpha + 1)$ is the value of the gamma function evaluated at $\alpha + 1$

random variables to a stable distribution. This allows for a much wider array of real-world applications as it doesn't require the stringent condition of identical distributions.

Lévy's work extended further, as he continued developing sufficient and necessary conditions under which the CLT holds, alongside the Croatian-American mathematician William Feller. [8]

3.4 Uses in the Bayesian Framework

Earlier, I mentioned how Laplace's proofs and general approach seemed to alternate between Bayesian and non-Bayesian, making his work somewhat incomplete and incoherent. This section aims to briefly touch on the differences of the Central Limit Theorem, in the Bayesian and frequentist statistical frameworks.

The Central Limit Theorem (CLT) and the Bayesian Central Limit Theorem (BCLT), also known as the Bernstein-von Mises Theorem [15] [3], are similar in that they both provide insights into the behavior of a large set of random variables. However, they approach this from two different perspectives, and they are used in different contexts.

The Bayesian theorem bears the names of mathematicians Richard von Mises and S. N. Bernstein, although the initial formal proof was presented by Joseph L. Doob in 1949 [15], specifically for random variables within a finite probability space. Later on, L. Le Cam, together with others, further extended the proof to encompass broader assumptions.

Regarding the difference between the usual Central Limit Theorem (CLT) and the Bayesian CLT (BCLT), the normal CLT is concerned with the distribution of sums of independent, identically distributed random variables, stating that they tend towards a normal distribution as the number of variables increases. On the other hand, the BCLT is a statement about the convergence of the posterior distribution in a Bayesian setting. It says that under certain conditions, the posterior distribution of the parameter of interest will be approximately normally distributed, centered at the true parameter value, with a certain variance, when the sample size grows large. Therefore,

[3]The Bernstein-von Mises Theorem in its simples form states the following: The normal distribution is centered at the maximum likelihood estimate and its covariance matrix is given by the inverse of the Fisher information matrix, scaled by the inverse of the sample size. Here, the true population parameter is denoted by θ_0 and $I(\theta_0)$ represents the Fisher information matrix evaluated at θ_0. Mathematically, this looks like:

$$\lim_{n\to\infty} P(\theta|x_1, \dots x_n) = \mathcal{N}(\theta_0, n^{-1}I(\theta_0)^{-1})$$

This suggests that, as the sample size grows, the posterior distribution for the parameter θ given the observed data $(x_1, \dots, x_n)$ approaches a normal distribution with mean θ_0 and variance given by the inverse of the Fisher information matrix, scaled by $1/n$.

the BCLT is used in a Bayesian context to justify the use of normal approximations to the posterior distribution. [12]

To simply illustrate the difference in approaches, I created a simple example using R[7] (a statistical programming software), where I simulated the heights of adult males in a population from a normal distribution, then applied both the frequentist and Bayesian approaches to estimating the mean and standard deviation relying on the CLT in the first case, and the Bernstein-von Mises theorem in the latter.

For the Bayesian approach, I used a normal-inverse-gamma prior [13] [4] and updated it with the data to get the posterior distribution of the mean and standard deviation. I then generated a large number of samples from the estimated distributions and plotted histograms of these samples for comparison.

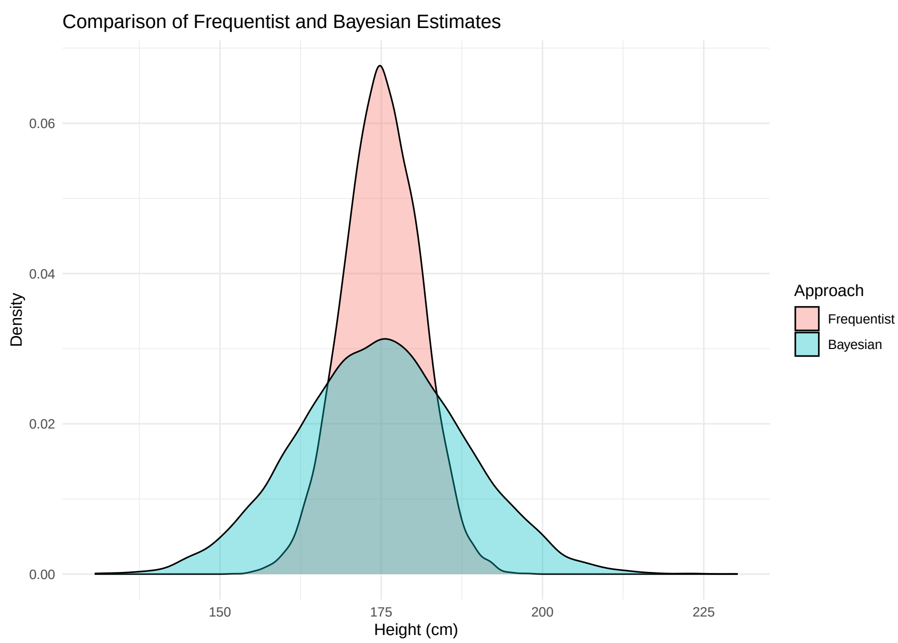

Figure 3.1: Comparison of Frequentist and Bayesian Estimates

We can note that both distributions are centered around the same value and have similar spreads, demonstrating the central limit theorem in action.

[4]Let's consider a normal distribution with unknown mean μ and unknown precision (the reciprocal of the variance) τ. In Bayesian statistics, a conjugate prior is a prior distribution that, when combined with a likelihood function through Bayes' theorem, gives a posterior distribution of the same family as the prior.

The normal-inverse-gamma distribution is defined as follows:

Given two independent random variables μ and τ, we say that (μ, τ) has a normal-inverse-gamma distribution if $\mu|\tau \sim \mathcal{N}(\mu_0, (\beta\tau)^{-1})$ and $\tau \sim \mathrm{Ga}(\alpha, \beta)$, where: $\mathcal{N}$ denotes the normal distribution, Ga denotes the gamma distribution, μ_0 is the prior mean, $\beta > 0$ is the prior precision, $\alpha > 0$ is the shape parameter for the gamma distribution

The Bayesian distribution appears narrower and taller as opposed to the frequentist one, because of the influence of the prior in Bayesian inference. The Bayesian method combines our prior beliefs (expressed in the prior distribution) with the observed data (via the likelihood function) to produce a posterior distribution.

In practice, the choice between frequentist and Bayesian approaches often comes down to the specific context and the analyst's beliefs about the underlying parameters.

3.5 Exploring Alan Turing's Contributions: A Deviation from CLT Theory

Alan Turing, one of the most notable mathematicians of the 20th century, outside of his primarily work in mathematical logic, machine computation, and artificial intelligence, made significant contributions to probability and statistics. Despite not being a widely acknowledged part of his career, Turing rediscovered a version of Lindeberg's famous formulation of CLT under the Lindeberg condition, as well as a significant part of the Feller-Lévy converse theorem to it while he was still an undergraduate student at Cambridge in 1934. [16]

For historical context, mathematical theory at Cambridge was surprisingly undeveloped at the time, as even the most respected mathematicians like G. H. Hardy lacked a basic acquaintance with the modern developments in probability theory.

Turing's work on the CLT reflected a striking comprehension of several complex mathematical concepts. In his fellowship dissertation, where most of his contributions lie, he identified the significance of examining distribution functions rather than densities — a considerable shift from traditional practices of the time. This transition paralleled Lindeberg's approach, indicating that Turing was moving along similar mathematical trajectories as contemporaries in the field.

One of the most significant contributions is Turing's discussion of the quasi-necessary conditions, which are not strictly necessary but serve as requirements for the convergence of U_n to Φ.[5] The two quasi-necessary conditions are expressed as:

[5] U_n denotes the shape function of the nth distribution function F_n. It is defined as $U_n(x) = F_n(\sigma_n(x - \mu_n))$, where μ_n represents the expectation and σ_n^2 represents the mean square deviation (MSD) of the nth distribution.

Φ represents the shape function of the Gaussian error. In this context, $\Phi(x)$ denotes the shape function of a standard Gaussian distribution, which has zero expectation and unit Mean Square Deviation.

Both U_n and Φ are functions that transform the input variable x according to specific distribution parameters (μ and σ^2 for U_n, and 0 and 1 for Φ) in order to standardize the distribution.

$$\sum_{k=1}^{\infty} \sigma_k^2 = \infty \quad \text{and} \quad \frac{\sigma_n^2}{s_n^2} \to 0.$$

Turing was also able to discern the necessary conditions for convergence to the normal distribution and additionally made contributions on the sufficient conditions. He displayed an impressive understanding of a Cramer-type factorization theorem at the time, a crucial component in deriving these conditions. Furthermore, he independently uncovered the Feller subsequence phenomenon. [3]

Finally, Turing concludes his work with a counterexample that satisfies his "quasi-necessary conditions" but does not fulfill the CLT. This example provides a fitting end to his discussion, demonstrating his comprehensive understanding of limit theorems at the time.

Following his dissertation, Turing's subsequent work, particularly during World War II, showcases his determination in the consideration of the statistical aspects of problems. His work in Bletchley Park (the main centre of Allied code-breaking during the Second World War), for instance, involved the development of statistical methods for decrypting German messages by making use of his seminal contributions in the fields of sequential analysis, empirical Bayes methods, and logarithms of the Bayes factor. [16]

Turing's work exemplifies the state of transition in mathematical probability during the 1930s, preceding the impact of the famous Soviet mathematician Andrey Kolmogorov's revolutionary work on probability theory. His influence continued to be felt in the field of statistics long after his active involvement in the discipline.

Chapter 4

Unfolding Progress: The Central Limit Theorem After Mid-20th Century

Post the Second World War, extensive studies on probability theories carried on, resulting in more profound developments and broader generalizations of the Central Limit Theorem (CLT). Particularly, two notable advancements were the extension of the CLT to limit theorems concerning Brownian motion [2], and the further refinement of the CLT to incorporate random variables operating within Hilbert spaces [5]. These were areas sporadically touched upon in the 1930s but gained significant attention by the 1950s: the exploration of stochastic processes as random elements in function spaces, the widening of fundamental concepts such as expectation and convergence, and the promising influence of functional analysis in the realm of probability theory.

Several tenets, ideas, and findings from the conventional theory relating to the sums of independent, real-valued random variables were extended to random elements within more abstract spaces. Instances include infinitely divisible distributions, characteristic functions, and iterations of the Lindeberg's condition. However, this marked a substantial departure from the early contributors to the CLT in two main ways. Firstly, traditional applications of the CLT, became less relevant in a scientific context increasingly dominated by stochastic viewpoints. Secondly, the analytical nature and conventional tools of analytical probability theory were largely substituted by nuanced measure-theoretic considerations. This transition is highlighted by the trajectory of characteristic functions. Despite noticeable similarities between properties of characteristic functions defined on dual spaces and traditional characteristic functions, the key property of finite-dimensional random variables — the continuous correspondence between characteristic functions and distributions — couldn't be sufficiently expanded to scenarios involving random variables with values in infinite-dimensional spaces. The methods typically applied to understand random variables in finite settings don't work as seamlessly when those variables can exist in a universe of endless possibilities, reducing the interest and research on

CLT as it was known. [16]

However, among notable mathematicians who continued their work on the theorem, the work of William Feller stands out. Feller, a revered figure in probability theory, made seminal contributions in the field of stochastic processes with his two-volume book, "An Introduction to Probability Theory and Its Applications," which remains a classic source on the subject, providing invaluable insights into the CLT and its applications. [6]

Another critical figure is Andrey Kolmogorov, who laid the groundwork for modern probability theory with his axiomatic approach. His significant contributions extended to the theory of stochastic processes, where the CLT plays a vital role. Mark Kac is also worthy of mention, renowned for his work in probability theory and statistical physics. His research into stochastic processes and random matrices added to the understanding of the CLT.

Within the Bayesian framwerok, Bruno de Finetti, while not directly working on the CLT, profoundly influenced its application through his work on probability theory. De Finetti introduced the concept of "exchangeability", which underpins Bayesian statistical inference. His work on infinitely exchangeable sequences of random variables enhanced our understanding of probability models' structure and provided a new context for interpreting various limit theorems, including the CLT. [6]

The nature of research in this period had significantly changed, as highlighted in this chapter, opening pathways for new formulations and developments in entirely novel mathematical frameworks.

Chapter 5

Conclusion

In this thesis, I have embarked on a comprehensive study of the Central Limit Theorem (CLT), an indispensable concept in the realms of probability and statistics. I traced the theorem's historical origins, examines the rigorous mathematical framework that underpins it, investigates its expansion, and illuminates its impactful role across various disciplines. This narrative underscores the remarkable evolution and enduring relevance of the CLT over the centuries.
I began the discussion with the early pioneers of the field — Jacob Bernoulli, Abraham de Moivre, and Pierre-Simon Laplace. Their work in the 17th to 19th centuries served as the basies for the formulation of what later became known as "The Central Limit Theorem". However, as is the nature of scientific progress, their initial theoretical models, while substantial, were not without room for refinement and improvement.

The upcoming research was subsequently taken up by a succession of mathematicians who worked diligently to add mathematical rigour and extend the theorem's applicability. Notably, Pafnuty Chebyshev and his students Aleksandr Lyapunov and Andrey Markov were instrumental in developing more mathematically sound proofs, distinguishing the CLT as a proper limit theorem and broadening its scope beyond classical probability. Their contributions significantly solidified the theorem's standing in mathematics.

In the 20th century, several mathematicians extended the CLT further, adding depth to its mathematical reach and breadth to its range of applications. Lévy, whose seminal work transcended the Gaussian distribution's boundaries and previous assumptions. They introduced new variations of the theorem that laid the foundation for modern approaches to probability theory and statistical analysis.

The dual examination of the Central Limit Theorem from both the Bayesian and frequentist perspectives underscored its adaptability and utility across various methodological approaches, illustrating its versatility. Indeed, the Central Limit Theorem has proven invaluable not only within the realm of probability theory and statistics but also in various other disciplines such as finance, physics, computer science, data analysis, business, and economics.

The post mid-20th century era, as explored in Chapter 4, opened up a multitude of developments, illuminating the works of mathematicians such as Bruno de Finetti, Mark Kac, William Feller, and others who contributed to the theorem's evolution and refolmulated and adapted it to the rapid changes in mathematical theory.

The Central Limit Theorem, with its various forms and wide-ranging applications, stands as a testament to the dynamic, expansive nature of probability theory and mathematical inquiry at large. It illustrates the interplay of intellectual curiosity, rigorous mathematical reasoning, and practical utility, making it a quintessential element of mathematical history and a powerful tool for current and future research.

In conclusion, the journey of the Central Limit Theorem, from its early origins to modern extensions, mirrors the evolution of probability theory itself, from a mathematically less formal discipline to an axiomatic and rigorous branch of mathematics. The Central Limit Theorem, while firmly rooted in its historical heritage, continues to inspire, influence, and inform research across various fields. Its narrative serves as a reminder of the continuous quest for knowledge that characterizes the mathematical discipline. Despite centuries of intensive study and the classical period coming to an end, the Central Limit Theorem remains a ground for exploration, with possible novel advancements in the years to come.

Bibliography

[1] William J. Adams. *The life and times of the central limit theorem*. 2nd ed. History of mathematics v. 35. OCLC: ocn403853728. Providence, R.I. : [London]: American Mathematical Society ; London Mathematical Society, 2009. ISBN: 978-0-8218-4899-9.

[2] Amir Bar. *Brownian motion and the Central Limit Theorem*. Course Notes. Based on Shang-Keng Ma, Statistical Mechanics, sections 12.2,12.7 and the course's notes section 6.2. Jan. 2013.

[3] L. Le Cam. "The Central Limit Theorem Around 1935". In: *Statistical Science* 1.1 (1986), pp. 78–91. ISSN: 08834237. URL: `http://www.jstor.org/stable/2245503`.

[4] David Draper and Erdong Guo. *The Practical Scope of the Central Limit Theorem*. en. Nov. 2021. URL: `https://ar5iv.labs.arxiv.org/html/2111.12267`.

[5] Jérôme Dedecker and Florence Merlevède. "The conditional central limit theorem in Hilbert spaces". In: *Stochastic Processes and their Applications* 108.2 (2003), pp. 229–262. ISSN: 0304-4149. DOI: `https://doi.org/10.1016/j.spa.2003.07.004`. URL: `https://www.sciencedirect.com/science/article/pii/S0304414903001157`.

[6] Hans Fischer. *A history of the central limit theorem: from classical to modern probability theory*. Sources and studies in the history of mathematics and physical sciences. OCLC: ocn567148466. New York ; London: Springer, 2011. 402 pp. ISBN: 978-0-387-87856-0.

[7] Jiqiang Guo et al. *rstan: R Interface to Stan*. Jan. 2023. URL: `https://cran.r-project.org/web/packages/rstan/index.html` (visited on 06/25/2023).

[8] Anders Hald. *A history of parametric statistical inference from Bernoulli to Fisher, 1713-1935*. eng. Sources and studies in the history of mathematics and physical sciences. New York: Springer, 2007. ISBN: 978-0-387-46409-1.

[9] Sang Gyu Kwak and Jong Hae Kim. "Central limit theorem: the cornerstone of modern statistics". en. In: *Korean Journal of Anesthesiology* 70.2 (2017), p. 144. ISSN: 2005-6419, 2005-7563. DOI: `10.4097/kjae.2017.70.2.144`. URL: `http://ekja.org/journal/view.php?doi=10.4097/kjae.2017.70.2.144`.

[10] *Normal Approximation to Binomial*. en. URL: `https://online.stat.psu.edu/stat414/lesson/28/28.1`.

[11] O. B. Sheynin. “The Appearance of Dirac’s delta functions in the works of P. S. Laplace (Russian)”. ru. In: *Istor.-Mat. Issled.* Vyp. 20 (1975), pp. 303–308, 381. URL: http://pyrkov-professor.ru/default.aspx?tabid=195&ArticleId=59.

[12] Ryan Giordano. *A stroll through the Bayesian central limit theorem. Part 1: Uniform laws of large numbers and maximum likelihood estimators.* Ryan Giordano, statistician. Dec. 3, 2020. URL: https://rgiordan.github.io/bayes/2020/12/03/bvm.html.

[13] A. F. M. Smith. “A General Bayesian Linear Model”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 35.1 (1973), pp. 67–75. DOI: https://doi.org/10.1111/j.2517-6161.1973.tb00937.x. eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1973.tb00937.x. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1973.tb00937.x.

[14] *St Petersburg Mathematical Society*. en. URL: https://mathshistory.st-andrews.ac.uk/Societies/St_Petersburg/.

[15] A. W. Van Der Vaart. *Asymptotic Statistics*. 1st ed. Cambridge University Press, Oct. 1998. DOI: 10.1017/CBO9780511802256. URL: https://www.cambridge.org/core/product/identifier/9780511802256/type/book.

[16] S. L. Zabell. “Alan Turing and the Central Limit Theorem”. en. In: *Taylor & Francis, Ltd. on behalf of the Mathematical Association of America* Vol. 102.The American Mathematical Monthly (July 1995), pp. 483–494.