Formal Verification of Central Limit Theorem in HOL Theorem Prover

Thi Cam Tu Phan

A thesis submitted for the degree of

Master of Computing (Advanced)

The Australian National University

School of Computing

May 2025

Abstract

Contents

Draft ideas

0.0.1 Abstract

- What is the CLT? Why is formal verification important?
- What limitations or gaps does this work address?
- Briefly summarize the methods, results, and significance.

0.1 Introduction

Purpose: Set the stage for this thesis, establish its relevance, and provide an overview of the challenges and contributions.

0.1.1 Background of the Central Limit Theorem (CLT)

- Briefly describe the CLT's significance in probability theory and its foundational role in various fields.
- Highlight the diversity of proofs for the CLT, focusing on Lyapunov's direction as a generalization of the i.i.d. case.

0.1.2 Formal Verification of Mathematical Theorems

- Discuss the importance of formalizing theorems like the CLT for computational mathematics.
- Introduce HOL4 and its potential for advancing probability theory through rigorous formalization.

0.1.3 Challenges and Motivation

- Outline the unique challenges of formalizing Lyapunov's approach in HOL4:
- HOL4 lacks a robust library for complex analysis and probability.
- The proof depends on Taylor expansions, Big-O notation, and Lyapunov's inequality, which needed to be formalized from scratch.

- Describe the failed attempt to use moment generating functions (MGFs) and how it inspired the shift to Lyapunov's approach.

0.1.4 Thesis Contributions

- Summarize key contributions:
  1. Formalization of Lyapunov's inequality in HOL4.
  2. Development of supporting libraries for Taylor expansions, Big-O notation, and moment analysis.
  3. Complete formalization of the CLT based on Lyapunov's theorem.

0.2 Background and Related Work

Purpose: Position this work in the context of previous efforts and provide foundational concepts.

0.2.1 Overview of the Central Limit Theorem

- Explain the statement of the CLT and its generalizations (e.g., Lyapunov's and Lindeberg's approaches).
- Focus on Lyapunov's version as the basis of this proof:

$$\text{If } \frac{\Gamma_n}{s_n^3} \to 0 \text{ as } n \to \infty, \tag{1}$$

  then

$$\frac{S_n}{s_n} \xrightarrow{d} \Phi.$$

  Define key terms: Γ_n, s_n, S_n, and Lyapunov's condition.

0.2.2 Formal Proof Assistants in Probability Theory

- Discuss Isabelle/HOL and its use of characteristic functions to prove the CLT (mention Luke Serafin's work briefly here).
- Highlight HOL4's potential and this thesis's motivation to enrich its probability theory libraries.

0.2.3 Lyapunov's Direction and Proof Techniques

- Provide an informal overview of Lyapunov's proof steps:
  - Construction of normal approximations (Y_j).
  - Use of Taylor expansions to bound differences.
  - Application of Lyapunov's inequality to ensure convergence.
- Contrast Lyapunov's approach with other CLT proofs, emphasizing why it's more suitable for HOL4's existing framework.

0.3 The Formalization Process

Purpose: Detail the technical challenges and solutions in formalizing Lyapunov's proof in HOL4.

0.3.1 Overview of Lyapunov's Proof Structure

- Provide a roadmap of the proof components you formalized:
  1. Taylor expansions and approximation of test functions.
  2. Telescoping sums and bounding higher-order moments.
  3. Lyapunov's inequality and its application to ensure convergence.

0.3.2 Formalizing the Supporting Theorems in HOL4

- Taylor Expansion
- Big-O Notation
- Lyapunov's Inequality

0.3.3 Technical Challenges and Innovations

- Share how HOL4's lack of complex analysis tools required alternative approaches (e.g., using Lyapunov's proof instead of characteristic functions).
- ...

0.4 The Formal Proof of the CLT in HOL4

Purpose: Present the completed proof and validate its correctness.

0.4.1 Step-by-Step Proof

- Present the formalized proof of Lyapunov's theorem in HOL4.
- Break it into logical steps, mirroring the informal proof:
  1. Setup and normalization of S_n.
  2. Construction of Z_j and comparison with X_j.
  3. Application of Taylor expansions and bounding error terms.
  4. Application of Lyapunov's inequality to complete the proof.

0.4.2 Validation and Soundness

- Discuss how HOL4 ensures the proof's correctness and soundness.
- Highlight key intermediate results that confirm the integrity of this thesis's approach.

0.5 Results and Comparisons

Purpose: Showcase the impact of this work and compare it to related efforts.

0.5.1 Achievements in HOL4

- Summarize the completed formalization of Lyapunov's CLT proof.
- Highlight the additional libraries (e.g., Taylor expansions, Big-O notation) and their potential for future work.

0.5.2 Comparison with Isabelle/HOL

- Briefly compare this work with Luke Serafin's formalization of the CLT in Isabelle:
  - HOL4 uses Lyapunov's approach instead of characteristic functions.

- Discuss the trade-offs: HOL4's modularity and flexibility versus Isabelle's existing tools for complex analysis.

0.6 Conclusion

Purpose: Reflect on the thesis's contributions and look ahead.

0.6.1 Summary of Contributions

Restate achievements:

1. Formalized Lyapunov's theorem and the CLT in HOL4.
2. Extended HOL4's libraries for probability theory.
3. Overcame challenges related to HOL4's lack of complex analysis tools.

Also, discuss the significance of enriching HOL4 for future formalizations in probability and functional analysis.

0.6.2 Future Work

Suggest potential extensions, such as:

- Formalizing multivariate versions of the CLT.
- Extending HOL4's libraries for advanced probabilistic theorems.

0.6.3 Appendices and Supplementary Material

Code listings, etc.

0.6.4 The proof of Central Limit Theorem

0.6.4.1 Informal Proof without introducing Y ((the replacement Gaussian random variables))

$\mathbb{E}[X_j] = 0,$

$\mathrm{Var}(X_j) = \sigma_j^2$ (finite variances, $\sigma_j^2 < \infty$),

$\mathbb{E}[|X_j|^3] = \gamma_j$ (finite third moments, $\gamma_j < \infty$).

Define:

$S_n = \sum_{j=1}^n X_j,$

$s_n^2 = \sum_{j=1}^n \sigma_j^2$ (total variance),

$\Gamma_n = \sum_{j=1}^n \gamma_j$ (total third moment).

Statement:

If

$$\frac{\Gamma_n}{s_n^3} \to 0 \quad \text{as } n \to \infty.$$

then

$$\frac{S_n}{s_n} \xrightarrow{d} \Phi$$

where Φ denotes the standard normal distribution $\mathcal{N}(0, 1)$.

Let $\boldsymbol{Z_n = S_n/s_n}$, to prove $\boldsymbol{Z_n \xrightarrow{d} \mathcal{N}(0, 1)}$, it suffices to show that for any bounded, continuous test function $f \in C_b^3(\mathbb{R})$ (bounded continuous functions with three bounded derivatives):

$$\mathbb{E}[f(Z_n)] \to \mathbb{E}[f(\mathcal{N})],$$

Using Taylor's theorem, we expand $f(Z_n)$ around 0:

$$f(Z_n) = f(0) + f'(0)Z_n + \frac{f''(0)}{2} Z_n^2 + R(Z_n),$$

where $R(Z_n)$ is the remainder:

$$R(Z_n) = \frac{f^{(3)}(\xi)}{6} Z_n^3,$$

for some ξ between 0 and Z_n.

Taking expectations:

$$\mathbb{E}[f(Z_n)] = f(0) + f'(0)\mathbb{E}[Z_n] + \frac{f''(0)}{2}\mathbb{E}[Z_n^2] + \mathbb{E}[R(Z_n)].$$

Similarly, for $\mathcal{N}(0,1)$, we expand $f(\mathcal{N})$:

$$\mathbb{E}[f(\mathcal{N})] = f(0) + f'(0)\mathbb{E}[\mathcal{N}] + \frac{f''(0)}{2}\mathbb{E}[\mathcal{N}^2] + \mathbb{E}[R(\mathcal{N})].$$

Thus, the difference becomes:

$$\mathbb{E}[f(Z_n)] - \mathbb{E}[f(\mathcal{N})] = \underbrace{f'(0)(\mathbb{E}[Z_n] - \mathbb{E}[\mathcal{N}])}_{(a)} + \underbrace{\frac{f''(0)}{2}(\mathbb{E}[Z_n^2] - \mathbb{E}[\mathcal{N}^2])}_{(b)} + \underbrace{(\mathbb{E}[R(Z_n)] - \mathbb{E}[R(\mathcal{N})])}_{(c)}.$$

By analyzing each term, we have:

(a) $f'(0)(\mathbb{E}[Z_n] - \mathbb{E}[\mathcal{N}])$

- By assumption, $\mathbb{E}[X_j] = 0$, so $\mathbb{E}[S_n] = 0$, and:

$$\mathbb{E}[Z_n] = \frac{\mathbb{E}[S_n]}{s_n} = 0.$$

- For $\mathcal{N}(0,1)$, $\mathbb{E}[\mathcal{N}] = 0$.

Therefore:

$$f'(0)(\mathbb{E}[Z_n] - \mathbb{E}[\mathcal{N}]) = 0.$$

(b) $\frac{f''(0)}{2}(\mathbb{E}[Z_n^2] - \mathbb{E}[\mathcal{N}^2])$

- The variance of S_n is $s_n^2 = \sum_{j=1}^n \sigma_j^2$, so:

$$\mathrm{Var}(Z_n) = \frac{\mathrm{Var}(S_n)}{s_n^2} = \frac{s_n^2}{s_n^2} = 1.$$

- Thus:

$$\mathbb{E}[Z_n^2] = 1.$$

- For $\mathcal{N}(0,1)$, $\mathbb{E}[\mathcal{N}^2] = 1$.

Therefore:

$$\frac{f''(0)}{2}(\mathbb{E}[Z_n^2] - \mathbb{E}[\mathcal{N}^2]) = 0.$$

(c) $\mathbb{E}[\boldsymbol{R(Z_n)}] - \mathbb{E}[\boldsymbol{R(\mathcal{N})}]$

By Taylor's theorem, We have:

$$\mathbb{E}[R(Z_n)] \leq \frac{M}{6}\mathbb{E}[Z_n^3],$$

and

$$\mathbb{E}[R(\mathcal{N})] \leq \frac{M}{6}\mathbb{E}[\mathcal{N}^3]$$

where $M = \sup_{x\in\mathbb{R}} |f^{(3)}(x)|$ is the maximum bound on the third derivative of f.

Thus,

$$\left|\mathbb{E}[R(Z_n)] - \mathbb{E}[R(\mathcal{N})]\right| \leq \frac{M}{6}\left|\mathbb{E}[Z_n^3] - \mathbb{E}[\mathcal{N}^3]\right| \tag{2}$$

For $\boldsymbol{Z_n}$:

The normalized third moment of $Z_n = S_n/s_n$ is:

$$\mathbb{E}[Z_n^3] = \frac{\mathbb{E}[|S_n|^3]}{s_n^3}. \tag{3}$$

Using Triangle inequality for sums, which states:

$$\mathbb{E}[|S_n|^3] = \mathbb{E}\left[\left|\sum_{j=1}^{n} X_j\right|^3\right] \leq \sum_{j=1}^{n} \mathbb{E}[|X_j|^3], \tag{4}$$

Combining (**??**_cube and (**??**_Z we get:

$$\mathbb{E}[|Z_n|^3] \leq \frac{\sum_{j=1}^{n} \gamma_j}{s_n^3},$$

where $\gamma_j = \mathbb{E}[|X_j|^3]$.

Thus:

$$\mathbb{E}[R(Z_n)] \leq \frac{M}{6}\frac{\sum_{j=1}^{n} \gamma_j}{s_n^3}.$$

For $\boldsymbol{\mathcal{N}(0,1)}$: The third moment of $\mathcal{N}$ is constant:

$$\mathbb{E}[\mathcal{N}^3] = 0, \tag{5}$$

To prove (**??**_moment, we use moment generating function (MGF). The moment generating function (MGF) of a random variable X is defined as:

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

For $X \sim N(0,1)$, the MGF is known to be:

$$M_X(t) = e^{t^2/2}.$$

This follows from integrating the product of e^{tx} and the PDF of $N(0,1)$:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = e^{t^2/2}.$$

Then, third derivative:

$$M_X^{(3)}(t) = \frac{d}{dt}\left(e^{t^2/2} + t^2 e^{t^2/2}\right).$$

Differentiate term by term:

$$M_X^{(3)}(t) = t^3 e^{t^2/2} + 3te^{t^2/2}.$$

Now evaluate $M_X^{(3)}(t)$ at $t = 0$:

$$M_X^{(3)}(0) = (0^3 + 3 \cdot 0)e^0 = 0.$$

Thus:

$$\mathbb{E}[X^3] = M_X^{(3)}(0) = 0.$$

Hence, the difference simplifies to:

$$\left|\mathbb{E}[Z_n^3] - \mathbb{E}[N^3]\right| = \left|\mathbb{E}[Z_n^3]\right| \leq \frac{\sum_{j=1}^n \gamma_j}{s_n^3} = \frac{\Gamma_n}{s_n^3}.$$

Since $\mathbb{E}[\mathcal{N}^3] = 0$, the (**??**_eq simplifies to:

$$\left|\mathbb{E}[R(Z_n)] - \mathbb{E}[R(\mathcal{N})]\right| \leq \frac{M}{6}\mathbb{E}[Z_n^3].$$

Substituting $\mathbb{E}[Z_n^3] \leq \frac{\Gamma_n}{s_n^3}$, we get:

$$\mathbb{E}[R(Z_n)] \leq \frac{M}{6}\frac{\Gamma_n}{s_n^3}.$$

Finally, by the assumption:

$$\frac{\Gamma_n}{s_n^3} \to 0 \quad \text{as } n \to \infty,$$

it follows that:

$$\mathbb{E}[f(Z_n)] \to \mathbb{E}[f(\mathcal{N})].$$

0.6.4.2 Informal Proof

Informal proof of Lyapounov's theorem for a single sequence by the method of Lindeberg Chung (2000).

- **Notation and Statement**

  1. $\{X_j\}_{j=1}^n$ be a sequence of independent random variables with:
     $\mathbb{E}[X_j] = 0$,
     $\mathrm{Var}(X_j) = \sigma_j^2$ (finite variances, $\sigma_j^2 < \infty$),
     $\mathbb{E}[|X_j|^3] = \gamma_j$ (finite third moments, $\gamma_j < \infty$).
  2. Define:
     $S_n = \sum_{j=1}^n X_j$,
     $s_n^2 = \sum_{j=1}^n \sigma_j^2$ (total variance),
     $\Gamma_n = \sum_{j=1}^n \gamma_j$ (total third moment).
  3. Statement: If
     $$\frac{\Gamma_n}{s_n^3} \to 0 \quad \text{as } n \to \infty.$$
     then
     $$\frac{S_n}{s_n} \xrightarrow{d} \Phi$$
     where Φ denotes the standard normal distribution $\mathcal{N}(0,1)$.

- **Step-by-Step Proof**

  1. **Normalize the Random Variables**
     Define the normalized variables:
     $$X_{n,j} = \frac{X_j}{s_n}, \quad \text{for } j = 1, \ldots, n.$$
     The sum S_n is now expressed as:
     $$\frac{S_n}{s_n} = \sum_{j=1}^n X_{n,j}.$$
  2. **Use the Lindeberg Replacement to approximate**
     Replace each X_j with a corresponding normal random variable $Y_j \sim \mathcal{N}(0, \sigma_j^2)$, where $\{Y_j\}_{j=1}^n$ are independent and have the same mean and variance as X_j, then satisfy:
     $$\mathbb{E}[Y_j] = 0 \quad \text{and} \quad \mathrm{Var}(Y_j) = \sigma_j^2.$$

Let all the X's and Y's be totally independent.
Since $Y_j \sim \mathcal{N}(0, \sigma_j^2)$, the normalized random variable:

$$\frac{Y_j}{s_n} \sim \mathcal{N}\left(0, \frac{\sigma_j^2}{s_n^2}\right).$$

Adding these normalized variables gives:

$$\frac{1}{s_n}\sum_{j=1}^{n} Y_j \sim \mathcal{N}\left(0, \sum_{j=1}^{n} \frac{\sigma_j^2}{s_n^2}\right).$$

Since $s_n^2 = \sum_{j=1}^n \sigma_j^2$, this simplifies to:

$$\frac{1}{s_n}\sum_{j=1}^{n} Y_j \sim \mathcal{N}(0, 1).$$

Now, construct the sequence Z:

$$Z_j = Y_1 + \cdots + Y_{j-1} + X_{j+1} + \cdots + X_n, \quad 1 \le j \le n.$$

Thus:

$$\begin{aligned}
Z_1 &= X_2 + X_3 + \cdots + X_n,\\
Z_2 &= Y_1 + X_3 + \cdots + X_n,\\
Z_n &= Y_1 + Y_2 + \cdots + Y_{n-1}.
\end{aligned}$$

In general, each Z_j represents a sum where:

- All variables before X_j are replaced by $Y_1, \dots, Y_{j-1}$,
- X_j is excluded from the summation,
- All variables after X_j remain as $X_{j+1}, X_{j+2}, \dots, X_n$.

Thus, by telescoping property, we have:

$$Y_j + Z_j = X_{j+1} + Z_{j+1}.$$

3. **Compare Distributions**

To show that $\frac{S_n}{s_n} \xrightarrow{d} \mathcal{N}(0, 1)$, we compare their expectations.

Using Test Funtions

Consider a test function f from C_B^3, the class of bounded continuous functions with three bounded derivatives.

Consider:

$$\mathbb{E}\left[f\left(\frac{S_n}{s_n}\right)\right] \quad \text{and} \quad \mathbb{E}\left[f\left(\mathcal{N}\right)\right],$$

By introducing the replacement sequence Z_j, we rewrite:

$$\mathbb{E}\left[f\left(\frac{S_n}{s_n}\right)\right] - \mathbb{E}\left[f\left(\frac{Z_n}{s_n}\right)\right].$$

By telescoping:

$$\mathbb{E}\left[f\left(\frac{S_n}{s_n}\right)\right] - \mathbb{E}\left[f\left(\frac{Z_n}{s_n}\right)\right].$$

$$= \mathbb{E}\left[f\left(\frac{X_1 + \cdots + X_n}{s_n}\right)\right] - \mathbb{E}\left[f\left(\frac{Y_1 + \cdots + Y_n}{s_n}\right)\right].$$

$$= \sum_{j=1}^{n}\left[\mathbb{E}\left[f\left(\frac{X_j + Z_j}{s_n}\right)\right] - \mathbb{E}\left[f\left(\frac{Y_j + Z_j}{s_n}\right)\right]\right].$$

4. **Expand expectations with Taylor's theorem**

By Taylor's theorem for $f \in C_B^3$:

$$\left|f(x+y) - \left[f(x) + f'(x)y + \frac{f''(x)}{2}y^2\right]\right| \leq \frac{M}{6}|y|^3,$$

where $M = \sup_{x\in\mathbb{R}} |f^{(3)}(x)|$ is the maximum bound on the third derivative of f.

Apply this to $f(\xi + \eta)$ and take Expectations, we write:

$$\left|\mathbb{E}[f(\xi+\eta)] - \mathbb{E}[f(\xi)] - \mathbb{E}[f'(\xi)]\mathbb{E}[\eta] - \frac{1}{2}\mathbb{E}[f''(\xi)]\mathbb{E}[\eta^2]\right| \leq \frac{M}{6}\mathbb{E}[|\eta|^3],$$

where ξ and η are independent random variables such that $\mathbb{E}[|\eta|^3] < \infty$.

Then, let ζ be another independent random variable, having the same mean and variance as η. Now, consider:

$$\mathbb{E}[f(\xi+\eta)] - \mathbb{E}[f(\xi+\zeta)].$$

Using Taylor's theorem again, this difference can be bounded as:

$$|\mathbb{E}[f(\xi+\eta)] - \mathbb{E}[f(\xi+\zeta)]| \leq \frac{M}{6}(\mathbb{E}[|\eta|^3] + \mathbb{E}[|\zeta|^3]),$$

where the sum $\mathbb{E}[|\eta|^3] + \mathbb{E}[|\zeta|^3]$ accounts for the third absolute moments of both random variables.

Then, substitute with:

- $\xi = Z_j/s_n$,
- $\eta = X_j/s_n$,

– $\zeta = Y_j/s_n$.

We have:

$$\left|\mathbb{E}\left[f\left(\frac{Z_j}{s_n}+\frac{X_j}{s_n}\right)\right]-\mathbb{E}\left[f\left(\frac{Z_j}{s_n}+\frac{Y_j}{s_n}\right)\right]\right| \leq \frac{M}{6}\left(\mathbb{E}\left[\left|\frac{X_j}{s_n}\right|^3\right]+\mathbb{E}\left[\left|\frac{Y_j}{s_n}\right|^3\right]\right). \tag{6}$$

5. **Bound error terms using Big-O notation**

From the definition of X_j/s_n and Y_j/s_n, we know:

$$\mathbb{E}\left[\left|\frac{X_j}{s_n}\right|^3\right] = \frac{\mathbb{E}[|X_j|^3]}{s_n^3}, \quad \mathbb{E}\left[\left|\frac{Y_j}{s_n}\right|^3\right] = \frac{\mathbb{E}[|Y_j|^3]}{s_n^3}.$$

For $Y_j \sim \mathcal{N}(0, \sigma_j^2)$, the third absolute moment is:

$$\mathbb{E}[|Y_j|^3] = c\sigma_j^3,$$

where $c = \sqrt{8/\pi}$ (a constant depending on the normal distribution).

Since $\mathbb{E}[|X_j|^3] = \gamma_j$ and $\mathbb{E}[|Y_j|^3] = c\sigma_j^3$, we wrire:

$$\mathbb{E}\left[\left|\frac{X_j}{s_n}\right|^3\right]+\mathbb{E}\left[\left|\frac{Y_j}{s_n}\right|^3\right] = \frac{\gamma_j}{s_n^3}+\frac{c\sigma_j^3}{s_n^3}.$$

By summing over $j = 1, \ldots, n$, (2) becomes:

$$\sum_{j=1}^{n}\left|\mathbb{E}\left[f\left(\frac{Z_j}{s_n}+\frac{X_j}{s_n}\right)\right]-\mathbb{E}\left[f\left(\frac{Z_j}{s_n}+\frac{Y_j}{s_n}\right)\right]\right| \leq \frac{M}{6}\sum_{j=1}^{n}\left(\frac{\gamma_j}{s_n^3}+\frac{c\sigma_j^3}{s_n^3}\right).$$

Lyapunov's inequality states that:

$$\sigma_j^3 \leq \gamma_j.$$

Thus:

$$\frac{\gamma_j}{s_n^3}+\frac{c\cdot\sigma_j^3}{s_n^3} \leq \frac{\gamma_j}{s_n^3}+\frac{c\cdot\gamma_j}{s_n^3} = \frac{(1+c)\cdot\gamma_j}{s_n^3}.$$

Simplify the Sum, we have:

$$\frac{M}{6}\sum_{j=1}^{n}\left(\frac{\gamma_j}{s_n^3}+\frac{c\cdot\sigma_j^3}{s_n^3}\right) \leq \frac{M}{6}\cdot\frac{1}{s_n^3}\sum_{j=1}^{n}(1+c)\gamma_j. \tag{7}$$

Let $K = \frac{M}{6}(1+c)$, and the total third absolute moment of X_j:

$$\Gamma_n = \sum_{j=1}^{n}\gamma_j(asdefineabove).$$

We substitute these into the inequality (3):

$$\frac{M}{6}\sum_{j=1}^{n}\left(\frac{\gamma_j}{s_n^3}+\frac{c\cdot\sigma_j^3}{s_n^3}\right)\leq K\cdot\frac{\Gamma_n}{s_n^3}. \tag{8}$$

By the definition of Big-O Notation, we have

> For a function $f(n)$, we say that $f(n) = O(g(n))$ if there exist positive constants $C > 0$ and $n_0 \geq 1$ such that:
>
> $$|f(n)| \leq C|g(n)|, \quad \text{for all } n \geq n_0.$$

Here:

- $f(n) = \frac{\Gamma_n}{s_n^3}$,
- $g(n) = \frac{\Gamma_n}{s_n^3}$.

Clearly:

$$|f(n)| = \left|\frac{\Gamma_n}{s_n^3}\right| \quad \text{and} \quad |g(n)| = \left|\frac{\Gamma_n}{s_n^3}\right|.$$

Since $f(n)$ and $g(n)$ are identical, we have:

$$|f(n)| \leq 1\cdot|g(n)| \quad \text{for all } n \geq 1.$$

Here, the constant $C = 1$. Then, the inequality $|f(n)| \leq C|g(n)|$ holds for all $n \geq 1$, so we can choose $n_0 = 1$.
Since we have found constants $C = 1$ and $n_0 = 1$ such that:

$$|f(n)| \leq C|g(n)| \quad \text{for all } n \geq n_0,$$

it follows by the definition of Big-O notation that:

$$\frac{\Gamma_n}{s_n^3} = O\left(\frac{\Gamma_n}{s_n^3}\right). \tag{9}$$

The Big O Notation property called **Multiplication by a constant** states that: Let k be a nonzero constant. Then:

$$O(|k|\cdot g) = O(g)$$

Apply this to (5), we have:

$$K\cdot\frac{\Gamma_n}{s_n^3} = O\left(\frac{\Gamma_n}{s_n^3}\right).$$

Then (4) becomes:

$$\frac{M}{6}\sum_{j=1}^{n}\left(\frac{\gamma_j}{s_n^3}+\frac{c\cdot\sigma_j^3}{s_n^3}\right)\leq O\left(\frac{\Gamma_n}{s_n^3}\right).$$

We have thus obtained the following estimate:

$$\forall f\in C^3:\left|\mathbb{E}\left[f\left(\frac{S_n}{s_n}\right)\right]-\mathbb{E}\left[f\left(N\right)\right]\right|\leq O\left(\frac{\Gamma_n}{s_n^3}\right).$$

6. **Convergence to Normal**

By Lyapunov's condition:

$$\frac{\Gamma_n}{s_n^3}\to 0\quad\text{as } n\to\infty.$$

This directly implies that:

$$\left|\mathbb{E}\left[f\left(\frac{S_n}{s_n}\right)\right]-\mathbb{E}\left[f(N)\right]\right|\to 0.$$

Since $\mathbb{E}[f(S_n/s_n)]\to\mathbb{E}[f(N)]$ for all bounded and continuous test functions f, by the general criterion for vague convergence (in Theorem 6.1.6 in Chung (2000)), we conclude that:

$$\frac{S_n}{s_n}\xrightarrow{d}\mathcal{N}(0,1).$$

Chapter 1

Introduction

Formal verification has continued revolutionizing the understanding and application of foundational theorems in mathematics: the surety of accompanying rigor and its implementation in computational systems. Thus have built chances for innovative mathematical services. Probability theory is a domain with both important theoretical results and practical applications, and it has increasingly become a focus of formalization efforts. The most important from such results is the Central Limit Theorem (CLT), which carries very strong implications in theory and practice.

The CLT is perhaps the most prominent result in probability theory describing how the sum of independent random variables, if suitably normalized, approximates a standard normal distribution when the number of terms increases indefinitely. It closes the gap between random single events and predictable behavior in the aggregate, forms the basis of inferential statistics, and underpins models in a vast variety of fields such as physics, finance, and artificial intelligence. Given this importance, formal verification of the CLT using proof assistants will be an important step toward proving its correctness in automated reasoning systems.

However, formalizing CLT is far from straightforward. There exist numerous variants of this theorem, together with their numerous proofs, ranging from classical independent and identically distributed (i.i.d.) to some more general versions. Notably, Lyapunov's theorem extends the CLT to independent, non-identically distributed (i.n.i.d.) random variables, offering a broader applicability than the classical i.i.d. case **?**. Each of those proofs uses advanced mathematical tools, including characteristic functions, moment generating functions, or even higher-order moments, according to the situation. These proofs require advanced mathematical infrastructure, which is not available with all proof assistants. In particular, the case of HOL4 Slind and Norrish (2008) has found it impossible to use some of the approaches, particularly the one based on characteristic functions, because the strong complex analysis library was lacking.

Instead, this thesis follows another path: leveraging Lyapunov's theorem and the Lindeberg approach, which avoids characteristic functions in favor of real-valued, moment-based techniques. Additionally, it pays attention to moment-based convergence criteria, trying to provide its conditions for convergence using Taylor expansions, Big-O notation of error bounds, and Lyapunov's inequality. This approach aligns well with the existing mathematical infrastructure of HOL4 while simultaneously expanding its capabilities.

1.1 Motivation

The formalization of the Central Limit Theorem in HOL4 involved interesting challenges because of the highly mathematical nature of the theorem and the specific limitations of existing libraries in HOL4. Another difference between the proof assistants, Isabelle/HOL, which has a very mature library for complex analysis and HOL4, is that there are no built-in key concepts for working with characteristic functions and others associated with the common classical proofs of CLT.

Lack of a well-developed library has been one of the major setbacks of HOL4. Proving the central limit theorem using the more traditional kinds of proofs, such as the one formalized in Isabelle/HOL, involves the use of characteristic functions for convergence proof. Characteristic functions are essentially Fourier transforms and thereby hinge a lot on complex analysis. It would have been very difficult to follow this approach in HOL4, since the proof assistant currently lacks support for these advanced tools.

An alternative approach was explored using moment generating functions (MGFs). MGFs are much more popular for proving the CLT, wherein their specialized property is utilized to show convergence to the standard normal distribution. However, it soon became evident that formally establishing the uniqueness property of MGFs would become a serious blocker. While this property is a triviality in classical analysis, the formalization work involved was highly intricate and beyond the scope of what is currently practical within HOL4's libraries. As a result, this approach was abandoned, leading to the need for a new direction.

The breakthrough came with the adoption of Lyapunov's theorem as the basis for formalizing the CLT. Lyapunov's argument has a very direct and elementary approach to the theorem without using complex analysis or moment-generating functions but rather basing on real-valued, moment-based convergence criteria; hence, it being particularly fitted into the existing framework of HOL4. Lyapunov's proof is a milestone of abstraction in the probability theory, but it is still beautifully simple and thus met the practical and strict requirements of formal verification.

However, Linderberg's method introduced its own set of challenges for formalization in HOL4:

- **Taylor Expansions:** The proof relies heavy on Taylor expansions to approximate test functions and bound errors. These had to be rigorously formalized in HOL4 from first principles.

- **Big-O Notation:** Reasoning about the asymptotic behavior of error terms involved the formalization of Big-O notation, which did not exist in HOL4 until now.

- **Lyapunov's Inequality:** An important part of the proof was the formalization of Lyapunov's inequality, which gives sufficient conditions for convergence. It was carefully adapted to fit into HOL4's logical framework.

Following the Linderberg's approach, with its valuation on real techniques and elementary methods, proved most important: it served to encapsulate the CLT under the constraints of HOL4 and was entirely consistent toward building reusable mathematics libraries. Taylor series developments, Big-O notion, and Lyapunov's inequality are indispensable not only for this proof but also as a basis of further formalizations in the realm of probability theory and elsewhere.

This shift from characteristic functions and MGFs to the more concrete, moment-based approach of Lyapunov underlines flexibility and problem-solving ingenuity necessary in formal verification projects. It thus shows how advanced mathematical results can be formalized in HOL4, even when serious deficiencies in its infrastructure must first be overcome. Moreover, it emphasizes the role of Lyapunov's approach as a turning point in the history of probability theory and as one of the bases for modern formalization.

1.2 Thesis Contributions

This thesis formalizes Lyapunov's inequality in HOL4, hence establishing a rigorous framework for convergence in probability proofs. In particular, the necessary mathematical theories and lemmas for the proof of CLT: Taylor expansions to bound errors, Big-O notation for asymptotic behavior analysis, and telescoping techniques for incremental replacements. These lemmas are used to formalize the Lyapunov's theorem and Lindeberg's approach, with a subsequent complete proof of CLT within HOL4. Unlike other approaches using characterization or moment-generating functions, this work does not use complex analysis, but works with the methods compatible with the real-valued nature of HOL4.

This thesis, beyond formalizing CLT, enriches libraries for probabilistic analysis that can be reused beyond, such as statistical inference and machine learning. This addresses some foundational gaps in HOL4 and demonstrates the use of HOL4 in advancing formalized mathematics.

1.3 Structure of the Thesis

The rest of the thesis is organized as follows. Chapter 2 provides background on the Central Limit Theorem, its history, and the setting of Lyapunov's proof. Related work, including other formalizations, using proof assistants like Isabelle/HOL is also covered. Chapter 3 then details the preliminaries, including HOL4 as proof engineering and related libraries. The key components of the proof, its completion, and its validation of the CLT are given in Chapter 4. Chapter 5 presents a discussion of results, and positioning of this work with respect to related efforts on formal verification. Finally, the thesis is concluded in Chapter 6 by summarizing the contributions made by this work and presenting the vision on future directions that probability theory in HOL4 may take.

This thesis not only shows that a formalization of the CLT in HOL4 is possible but also opens up further opportunities for work on the formal probability theory. It fills in some foundational gaps and provides reusable tools, which illustrates more general potential of HOL4 to make serious contributions to the fully verified mathematics corpus.

Chapter 2

Background and Related Work

The Central Limit Theory is one of the main theoretical results from the Probability theory, a bridge from individual randomness to aggregate predictability. It marks the gradual approach of mathematical rigor and upgrading of the theorem used by addressing more and more difficult problems. Over centuries, different forms of CLTs have been derived, each of which has added to its understanding convergence and applications. This chapter will present the historical development of the CLT that describes its milestones and contributions. It will also examine the modern theorem proving tools and ways of formalization, focusing on Lyapunov's approach.

2.1 Historical Background

Let us just trace briefly the historical evolution of the CLT, critical milestones, key contributors, and the gradual refinement of its conditions and scope. The roots of the CLT go back to the dispute of the eighteenth century, specifically to the time Abraham de Moivre laid the foundation for this theorem. In 1733, de Moivre showed that the sum of a large number of independent and identically distributed random variables converges to the normal distribution-an elegant approximation to problems such as the rolling of dice or repeated coin flips De Moivre (1733). His work initiated the interplay between discrete distributions-such as the Binomial-and their continuous approximations, and laid the ground for developments to come.

In 1810, Pierre-Simon Laplace independently extended de Moivre's insights by proving a much more general form of this theorem. Under broader conditions, Laplace demonstrated convergence of the sum of independent random variables to a normal distribution, thereby establishing CLT as a universal principle. His work was used in applying the CLT to population statistics and measurement errors Laplace (1835) and cemented its usefulness for aggregate phenomena analysis.

The 19th century would see the formalization of the central limit theorem (CLT) by Pafnuty Chebyshev, who also gave the theorem a rigorous foundation by the introduction of variance and expectation conditions. With that work, Chebyshev rendered the CLT mathematically precise, linking it to the developing area of probability theory Chebyshev

(1890). Building on this work was further advancing the generalization of the CLT by Alexander Lyapunov to sequences of independent random variables that are not identically distributed. His version of the theorem brought forth Lyapunov's condition, which contained definite criteria for convergence Lyapunov (1895).

In the early years of the 20th century, contributions by Lyapunov represented such a change in the evolution of the CLT. Unlike his predecessors, Lyapunov sought a "direct" and "elementary" proof. It would try to make the theorem internal relations simpler and relax the theorem's assumptions. His papers published in 1900 and 1901 reconstructed previous methods with a rigor of Weierstrass, and introduced new precision to the theorem by establishing explicit bounds on the normalized sum and the limiting normal distribution. This specificity answers Chebyshev's earlier demand for sharper probabilistic results although it should have simplified the proof through some clever but much deeper refinements of Poisson's arguments Fischer (2011).

Lyapunov's goals extended beyond mere generalization; he aimed to illuminate the theorem's internal structure and make it more accessible. Avoiding the complexities of characteristic functions and complex analysis, he focused on moment-based criteria, relying on real-valued tools like Taylor expansions, bounding techniques, and Lyapunov's inequality. Thus the balance here is theoretical rigor and practical applicability, two characteristics that still resonate well with modern efforts at formal verification as that undertaken in this thesis.

In contemporary times, further refinements of the CLT were provided by William Feller bringing it to discrete random variables Feller (1945), and all subsequent developments that are probabilistic have been geared towards this common extension to stable distributions and higher dimensional data. These extensions demonstrate the theorem's adaptability to various contexts, from stochastic processes to machine learning and data analysis.

Lyapunov's version of the CLT represents a key milestone in this journey. It generalizes the theorem for independent random variables with finite moments so that it becomes an observed but pliant frame of reference for convergence. It is this balancing between abstractness and concreteness that is a characteristic of the whole development of probability during the modern period. Contemporaries of Lyapunov such as von Mises, Lévy, Cramér, Khinchin, and Kolmogorov likewise pursued abstract relations in probability theory while keeping an eye on practical relevance. This tension is paradigmatically present in Lyapunov's work, which illustrates the way in which results which are purely formal mathematics may nonetheless satisfy external criteria of applicability and utility. For more information about the history and the proofs of the Central Limit Theorem, see Fischer (2011); **?**.

This thesis is based on the generalization of CLT as expressed in Lyapunov's version.

Because of its focus on real-valued methods and moment-based criteria, it avoids any complex tools like characteristic functions or complex analysis, hence being particularly suitable for formalization in HOL4. In this respect, Lyapunov's approach not only advances the mathematical generalization but also provides a practical framework to meet the rigor and accessibility requirements of automated reasoning systems by directly addressing convergence under relaxed conditions.

2.1.1 Related Work

Formalization of mathematical theorems as proofs has been a singular research focus in modern computational mathematics. The systems such as Isabelle/HOL Team (2015), Coq Bertot and Castéran (2013), and HOL4 Slind and Norrish (2008) have been heavily relied on over the years to encode and prove basic properties but offer several features and challenges.

One of the most significant achievements in this area was the formalization of CLT within the Isabelle/HOL proof assistant. This was made possible mainly by the rich library related to complex analysis in Isabelle/HOL, which provided the means to make use of characteristic functions within the proof. Characteristic functions are Fourier transforms of probability distributions that make convergence easier to analyze by drawing on their special mathematical properties. Central to the approach are two key steps: first, pointwise convergence of the characteristic function for the normalized sum of independent random variables to that of the standard normal distribution should be demonstrated; and second, by the application of the Lévy Continuity Theorem, one obtains the desired convergence in distribution. These two steps are actually the core of the proof that eventually leads to the verification of CLT **?**.

The characteristic function approach is beautifully complemented by the infrastructure provided by Isabelle/HOL, which supports careful manipulation of complex-valued functions, differentiation, and limits. This thereby enables Isabelle/HOL to achieve the rigorous performance that is demanded by Fourier analysis and probability theory. Among the striking personalities behind this effort is Luke Serafin, who has produced a giant leap forward in the formalization of the CLT **?**. Serafin's systematic encoding of the characteristic function method has ensured that it was entirely generic and modular. His contributions involved the formalization of key intermediate results such as the for sums of independent random variables and the fact that components of his work are reusable, enabling extension into applications that relate to probabilistic theorems.

The strength of Isabelle/HOL is to borrow from its excellent complex analysis library. The characteristic function-based proofs are formalized smoothly, and the modular and extensible framework allows researchers to always build on previous work, gradually increasing the scope of formalized mathematics. The CLT formalization within Isabelle/HOL is exemplary in showing how even the most intricate probabilistic results can be encoded rigorously with the help of proof assistants. But this also points to some of the challenges with this approach: it is very much based on Fourier transforms and complex-valued methods, an infrastructure which may not always be as easily available when moving to other proof systems. Also, characteristic functions are powerful but often complicate proofs in situations where simpler real-valued methods, such as those based on moments, would suffice.

The formalization of CLT in Isabelle/HOL serves as an inspiration and comparison for the research presented in this thesis. The Example of Isabelle/HOL shows that the CLT can be formalized with the help of characteristic functions but also brings into relief the limitations of such a strategy-if proof assistants like HOL4 lack libraries comparable for complex analysis with it. This thesis takes a different site following the approach with the theorem of Lyapunov, which involves real-valued moment conditions rather than characteristic functions. Following this path, the work in HOL4 demonstrates how easier and less convoluted tools enable formulating the CLT while considering the inherent challenges of formal verification.

Chapter 3

Preliminaries

This chapter describes an overview of the theoretical and formal theories required for the formalization of the Central Limit Theorem . This includes HOL Formalization, Measure Theory, Lebesgue Integration, and Probability Theory.

3.1 HOL Formalization

Higher Order Logic (HOL) Slind and Norrish (2008); **?** is derived from the Logic of Computable Functions (LCF) **??** created by Robin Milner and colleagues in 1972. HOL is an adaptation of Church's Simple Theory of Types (STT) **?**, where a higher-order version of Hilbert's choice operator ε, Axiom of Infinity, and Rank-1 polymorphism have been added. HOL4 implements the original HOL framework, while other theorem provers in the HOL family, such as Isabelle/HOL, include important extensions. Such a simple logical basis makes HOL more accessible than those systems founded on much more advanced dependent type theories, such such the Calculus of Inductive and Co-Inductive Constructions constructed by Coq. Therefore, theories and proofs founded on HOL are easier for a layman to comprehend rather than being lost in a complicated type theory.

HOL refers both to the logical system and the software implementing it. HOL4 is the latest version of this software and written in Standard ML (SML), a general-purpose functional programming language. SML has played the most vital role in the HOL4 for implementing its core engine, enabled automation due to which proof tactics have been written in that and also for interaction, whether it is through a proof script or in direct correspondence with the user. Integrated SML gives a way in which HOL4 is versatile and can easily be extended such that complex verification tools are provided to develop the management of proofs by a user efficiently.

HOL terms are representatives of such things and the grammar includes constants, variables, applications (function calls) and lambda abstractions. Quantifiers, for instance universal (`!x. P(x)`) and existential (`?x. P(x)`), are also provided in HOL - they are defined as specific lambda functions.

3.2 Measure Theory

3.3 Lebesgue Integration Theory

3.4 Probability Theory

Chapter 4

Central Limit Theorem

asldfjkldasjflkdfjldsa;f

Chapter 5

Future Work

Future Work

Chapter 6

Conclusion

asdfsdfasdf

asdasasfdfdf

References

Yves Bertot and Pierre Castéran. *Interactive theorem proving and program development: Coq'Art: the calculus of inductive constructions*. Springer Science & Business Media, 2013.

Manfred Broy, Bengt Jonsson, J-P Katoen, Martin Leucker, and Alexander Pretschner. Model-based testing of reactive systems. In *Volume 3472 of Springer LNCS*. Springer, 2005.

Jerry R Burch, Edmund M Clarke, Kenneth L McMillan, David L Dill, and Lain-Jinn Hwang. Symbolic model checking: 1020 states and beyond. *Information and computation*, 98 (2):142–170, 1992.

Pafnutii Lvovich Chebyshev. Sur deux théorèmes relatifs aux probabilités. *Acta math*, 14 (1):305–315, 1890.

Kai Lai Chung. *A course in probability theory*. Elsevier, 2000.

Edmund M Clarke, Thomas A Henzinger, Helmut Veith, Roderick Bloem, et al. *Handbook of model checking*, volume 10. Springer, 2018.

Abraham De Moivre. *Approximatio ad summam terminorum binomii (a+ b) n in seriem expansi*. 1733.

Willy Feller. The fundamental limit theorems in probability. *Bulletin of the American Mathematical Society*, 51:800–832, 1945. Reprinted in Adams 2009, pp. 80–113.

Hans Fischer. *A history of the central limit theorem: from classical to modern probability theory*, volume 4. Springer, 2011.

John Harrison. Hol light, 2015. URL `http://www.cl.cam.ac.uk/~jrh13/hol-light/`. Accessed: 2024-12-01.

Pierre Simon Laplace. *Oeuvres complètes de Laplace*. Gautier-Villars, 1835.

Aleksandr Mikhailovich Lyapunov. Pafnutii lvovich chebyshev, 1895.

Sheldon Ross. *First Course in Probability, A*. Pearson Higher Ed, 2019.

Konrad Slind and Michael Norrish. A brief overview of hol4. In *International Conference on Theorem Proving in Higher Order Logics*, pages 28–32. Springer, 2008.

Isabelle Team. Isabelle/hol, 2015. URL `https://isabelle.in.tum.de/`. Accessed: 2024-12-01.

Index