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

Chapter 1

Introduction

The idea of a formalised mechanist is simple but powerful: to trust mathematics in software, we must mechanise not only calculations but also proofs. A mechanist builds a logic engine where every theorem arises from a small set of primitive rules. If those rules are sound, everything derived from them inherits that soundness.

HOL4 supports this mechanised philosophy [12]. It encodes logic using a typed λ-calculus and restricts theorem creation to a protected MetaLanguage (ML) type, ensuring that every step is checked with mathematical rigour.

In this thesis, we formalise one of the most important results in probability theory: the Central Limit Theorem (CLT). When many independent effects accumulate, their combined influence often resembles a bell curve. The CLT explains this universal phenomenon observed across diverse fields—from biology and economics to physics and computation.

We formalise the Lyapunov version of the CLT from [3], using Lindeberg's replacement method. Let $\{X_j\}_{j=1}^n$ be a sequence of independent random variables with zero mean, finite variance, and third absolute moment:

$$\mathbb{E}[X_j] = 0, \quad \mathbb{E}[X_j^2] = \sigma_j^2, \quad \mathbb{E}[|X_j|^3] = \gamma_j < \infty.$$

Define the sum $S_n = \sum_{j=1}^n X_j$, total variance $s_n^2 = \sum \sigma_j^2$, and total third moment $\Gamma_n = \sum \gamma_j$. Then, under the Lyapunov condition:

$$\lim_{n\to\infty} \frac{\Gamma_n}{s_n^3} = 0,$$

we have convergence in distribution:

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

Here is the Lyapunov's form of the Central Limit Theorem in HOL4:

```
Theorem central_limit_theorem :
⊢ ∀p X N.
prob_space p ∧
ext_normal_rv N p 0 1 ∧
(∀i. real_random_variable (X i) p) ∧
(∀n. indep_vars p X (λi. Borel) (count n)) ∧
(∀i. expectation p (X i) = 0) ∧
(∀i. expectation p (λx. (abs (X i x)){3) < +∞) ∧
(∀i. variance p (X i) < PosInf) ∧
(∀i. variance p (X i) ≠ 0) ∧
(∀n. (sqrt (second_moments p X n)) ≠ 0) ∧
((λn. (third_moments p X n) / ((sqrt (second_moments p X n)) pow 3)) --> 0) sequentially
⇒
((λn x. (SIGMA (λi. X i x) (count n)) / (sqrt (second_-
moments p X n))) --> N)
(in_distribution p)
```

Although the final convergence proof is not yet complete, this thesis represents a significant step toward a full mechanisation. Over 6000 lines of HOL4 code were developed, including formal definitions, theorems, and supporting libraries. An initial attempt using moment-generating functions was abandoned due to insufficient HOL4 support. The current Taylor-based approach has allowed us to formalise all major components. The only missing piece is a bound involving the gamma function, which requires integration techniques not yet supported in HOL4. This is a matter of technical realisation, not conceptual difficulty.

Chapter 2

Background and Related Work

2.1 Background: The Central Limit Theorem

The central limit theory is one of the fundamental results of the Probability theory, a theoretical connection from individual randomness to collective predictability. It states the empirical fact that the sum of large numbers of independent random variables converges to a normal distribution, regardless of the individual distributions, as long as some general assumptions hold. This theorem explains the universality of the normal distribution in nature and is the foundation for an extremely broad range of applications in statistics, physics, and computer science.

Historically, the beginnings of the CLT trace back to Abraham de Moivre, who showed in 1733 that the binomial distribution is approximated very well by the normal distribution if the number of trials is large [5]. Pierre-Simon Laplace later generalised this result and formalised it in terms of the Laplace–de Moivre theorem. Chebyshev and his students, especially Lyapunov and Markov, pushed the theory further by relaxing the identically distributed condition and replacing it with moment conditions. Lyapunov's 1901 version introduced a now-famous condition on the third absolute moment, which is still the cornerstone for general versions of the CLT.

A full history of the development of the CLT can be found in Fischer's *A History of the Central Limit Theorem* [5], which chronicles how the theorem evolved from numerical approximations to a fundamental limit law in probability. The evolution reflects a broader trend in mathematics: from combinatorial methods to analysis, and finally, to measure-theoretic and functional analytic foundations.

In this thesis, we focus on the Lyapunov form of the CLT, as presented in Chung's textbook [3]. It assumes independence, but not identical distribution, and uses the Lindeberg replacement trick to incrementally replace variables with normal ones, bounding the error using Taylor's theorem. This analytic method avoids characteristic functions and relies only on real analysis and moment estimates. As a result, it is well-suited to formalisation in systems like HOL4, which are strong in real analysis but currently lack mature libraries for complex integration.

2.2 Related Work

Formal proofs of the CLT have previously been attempted in several proof assistants. The most notable example is the Isabelle/HOL formalisation by [11], which proves the CLT under the assumption of independent identically distributed (i.i.d.) random variables. That work follows the classical approach using characteristic functions and the Lévy Continuity Theorem.

While the Isabelle/HOL technique is elegant and mathematically sound, it relies on involved analysis and Fourier transforms, which are well-supported in Isabelle but not yet fully formalised in HOL4. HOL4 is lacking key theorems for complex-valued functions at present, so the characteristic function approach is unworkable here.

Moreover, the i.i.d. assumption limits the scope of the Isabelle formalisation. In contrast, this thesis formalises a strictly more general version of the CLT—Lyapunov's form—which requires only independence and finiteness of variances and third absolute moments. The approach avoids characteristic functions entirely and instead uses the Lindeberg replacement method with Taylor expansion and asymptotic error bounds.

This makes the result both broader in applicability and better aligned with the current capabilities of HOL4's real and measure-theoretic libraries. The infrastructure developed in this work—including handling of expectations, variances, random variable sequences, and Taylor bounds—may serve as a foundation for generalising beyond Lyapunov's condition in future work.

To the best of our knowledge, this is the first mechanised proof of the Central Limit Theorem in HOL4 that goes beyond the i.i.d. case and formalises the full structure of the Lindeberg–Lyapunov strategy.

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) [9, 12] is derived from the Logic of Computable Functions (LCF) [7, 10] created by Robin Milner and colleagues in 1972. HOL is an adaptation of Church's Simple Theory of Types (STT) [4], 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.

The type system of HOL establishes the structural framework within which all terms and expressions are guaranteed to be well-defined and logically consistent. Types in HOL denote sets within the universe U, and every term bears a certain type. The type grammar is simple and very expressive, and thus able to construct a wide variety of mathematical and logical objects.

The type grammar is defined as:

$$\sigma ::= \alpha \mid c \mid (\sigma_1, \ldots, \sigma_n)\text{op} \mid \sigma_1 \to \sigma_2$$

where:

1. **Type Variables ($\alpha, \beta, \ldots$):** Generic placeholders that allow polymorphism to provide functions and predicates over different types.
   - Example: The type variable α could indicate integers, Booleans, or functions.
2. **Atomic Types (c):** Fixed and pre-defined types within HOL. The two initial atomic types are:
   - `bool`: The set of Boolean values $\{T, F\}$.
   - `ind`: The set composed by individuals (an infinite set).
3. **Compound Types ($(\sigma_1, \ldots, \sigma_n)$op):** Formed by applying type operators to other types. Their examples include Cartesian products, which designate the tuples over the elements.
   - Example: The type $(\texttt{bool}, \texttt{ind})\times$ represents pairs of a Boolean and an individual.
4. **Function Types ($\sigma_1 \to \sigma_2$):** Represent total functions mapping elements from a domain (σ_1) to a codomain (σ_2).
   - Example: The type $\texttt{bool} \to \texttt{ind}$ indicates a function mapping both Boolean-values to individual-elements.

For example, consider the following types:

1. A function from integers to Booleans:

   $$f : \texttt{int} \to \texttt{bool}$$

   This type indicates that $f(x)$ is a function taking an integer x and returning a Boolean.
2. A tuple containing a Boolean and a function:

   $$p : (\texttt{bool}, (\texttt{int} \to \texttt{bool}))$$

   This is a pair type $p = (b, f)$, where b is a Boolean, and f is a function mapping integers to Booleans.

3. The type system guarantees the consistency by making sure all terms are properly typed. So if g : `int` → `bool`, then $g(5)$ as 5 is an integer, but, $g(T)$ would be invalid since T is a Boolean, not an integer. Such stringent typing is at the level of terms to avoid self-contradictory values and assure that proofs built up in HOL are sound.

In HOL, terms are representatives for elements of sets represented by their types. The grammar of the term defines the syntax and structure for the logical expressions that can be expressed and hence statements that could be well typed and logically valid. Terms in HOL are constructed from the following components:

$$t ::= x \mid c \mid t\ t' \mid \lambda x.t$$

where:

1. **Variables** $(x, y, \ldots)$:
   - Represent placeholders for elements of a type.
   - Example: x : `bool` stands for a Boolean variable.
2. **Constants (c):**
   - Fixed entities such as T, F, mathematical operators, or predefined functions.
   - Example: The constant + defines addition for numeric types.
3. **Function Applications ($t\ t'$):**
   - Define the application of a function to an argument. The term $f(x)$ applies the function f to the variable x.
   - Example: If f : `int` → `real` and x : `int`, then $f(x)$ is a valid term of type `real`.
4. **λ-Abstractions ($\lambda x.t$):**
   - Denote anonymous functions where x is the input variable, and t is the function body.
   - Example: $\lambda x.x + 1$ defines a function that increments its input by 1.

To ensure consistency, the terms of HOL should be well typed. Given a term t_σ of type σ, its grammar can be generalized with type annotations:

$$t_\sigma ::= x_\sigma \mid c_\sigma \mid (t_{\sigma_1 \to \sigma_2}\ t'_{\sigma_1})_{\sigma_2} \mid (\lambda x_{\sigma_1}.t_{\sigma_2})_{\sigma_1 \to \sigma_2}$$

HOL's deductive system is considered the logical foundation for forming and checking a proof. HOL's deductive system may consist of eight primitive rules of inference, the definition of new theorems by existing theorems. These rules are the basic components and are required for all logical reasoning within HOL, ensuring that proofs are consistent, logically valid, and traced. The following are the eight main primitive inference rules in HOL:

1. **Assumption Introduction (ASSUME):**
   - Introduces a formula as an assumption.
   - Rule:
   - Example: From the assumption P, we conclude P.
2. **Reflexivity (REFL):**
   - States that any term is equal to itself.
   - Rule:
   - Example: For $x : \texttt{int}$, $x = x$ is always true.
3. **Beta Conversion (BETA_CONV):**
   - Applies substitution in lambda abstractions.
   - Rule:
   - Example: $(\lambda x.x + 1)(5) \vdash 5 + 1$.
4. **Substitution (SUBST):**
   - Replaces a term in a formula with another term proven to be equal.
   - Rule:
   - Example: From $x = y$ and $P(x)$, infer $P(y)$.
5. **Abstraction (ABS):**
   - Generalizes an equation by abstracting a variable.
   - Rule:
   - Example: From $5 + 1 = 6$, infer $\lambda x.x + 1 = \lambda x.6$.
6. **Type Instantiation (INST_TYPE):**
   - Specializes polymorphic functions or predicates to specific types.
   - Rule:

- Example:

7. **Discharging Assumptions (DISCH):**
   - Converts an assumption into an implication.
   - Rule:
   - Example: From $P \wedge Q$, infer $P \Rightarrow (Q \wedge P)$.

8. **Modus Ponens (MP):**
   - Combines an implication and its premise to infer the conclusion.
   - Rule:
   - Example: From $x > 0 \Rightarrow x^2 > 0$ and $x > 0$, infer $x^2 > 0$.

These inference rules ensure that all logical derivations are traceable to basic axioms and established theorems. Additionally, the deductive system forms the backbone of HOL4, ensuring that proofs are both rigorous and reliable.

All proofs in HOL are fundamentally derived from a set of primitive inference rules and a core logical foundation. These rules define the semantics of two fundamental logical connectives: **equality** (=) and **implication** ($\implies$). Other logical connectives and firstorder quantifiers, such as logical truth (T), falsehood (F), conjunction ($\wedge$), disjunction ($\vee$), and existential quantification ($\exists$), are defined as lambda (λ) functions for consistency within the HOL framework:

1. **Logical Truth (T)**
   - Rule
   - True is represented as the equality of two identical boolean functions.

2. **Logical Falsehood (F)**
   - Rule
   - False is defined to satisfy any boolean implication.

3. **Negation ($\neg$)**
   - Rule
   - Negation is the implication of a boolean value leading to falsehood.

4. **Conjunction ($\wedge$)**
   - Rule

- Conjunction is defined as a logical function that evaluates nested implications.

5. **Conjunction (∧)**
   - Rule
   - Conjunction is defined as a logical function that evaluates nested implications.
6. **Disjunction** (∨)
   - Rule
   - Disjunction is expressed through sequential implications.
7. **Universal Quantifier (∀)**
   - Rule
   - Universality asserts that a predicate holds for all elements of a type.
8. **Existential Quantifier (∃)**
   - Rule
   - Existence is defined using Hilbert's choice operator (ε).

   HOL also defines constructs for mathematical operations, such as **one-to-one functions** (One_One) and **onto functions** ($Onto$), to extend logical capabilities:
9. **One-to-One (**One_One**)**
   - Rule
10. **Onto (**$Onto$**)**
    - Rule

    HOL includes the constant $Type\_Definition$, which defines new types as bijections of subsets of existing types:
11. **Type Definition** $Type\_Definition$
    - Rule
    - This process is automated by the HOL Datatype package, simplifying the creation of new types.

HOL's standard theory is built upon four foundational axioms:

1. **Boolean Cases (BOOL_CASES_AX)**

- Rule
- This axiom ensures that any boolean value is either true or false.

2. **Eta Conversion (ETA_AX)**
   - Rule
   - Eta conversion describes the extensionality of functions.
3. **Hilbert's Choice (SELECT_AX)**
   - Rule
   - This axiom relates the choice operator to existential quantification.
4. **Infinity (INFINITY_AX)**
   - Rule
   - The Axiom of Infinity ensures the existence of an infinite set.

These axioms are generally sufficient for conventional formalization projects in HOL4. Adding new axioms is strongly discouraged, as it can compromise logical consistency.

3.2 Measure Theory

Measure theory provides the mathematical foundation for probability and integration. In this project, we rely on the HOL4 formalisation of measure spaces, σ-algebras, and measurable functions, as defined in the 'measureTheory' library [9].

3.2.1 Measure Spaces

A *measure space* is a triple (X, Σ, μ), where:

- X is the underlying space,
- Σ is a σ-algebra of subsets of X,
- $\mu : \Sigma \to \overline{\mathbb{R}}$ is a measure.

In HOL4, these components are represented as:

```
Definition m_space_def :
⊢ ∀sp sts μ.
m_space (sp, sts, μ) = sp
```

```
Definition measurable_sets_def :
⊢ ∀sp sts μ.
measurable_sets (sp, sts, μ) = sts
```

```
Definition measure_def :
⊢ ∀sp sts μ.
measure (sp, sts, μ) = μ
```

A measure μ satisfies:

1. **Non-negativity:** $\mu(A) \geq 0$ for all $A \in \Sigma$,
2. **Null empty set:** $\mu(\emptyset) = 0$,
3. **Countable additivity:** For disjoint sets $\{A_i\} \subseteq \Sigma$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

HOL4 uses the extended real number type `extreal`, which includes $+\infty$, $-\infty$, and undefined values, to represent measure values.

3.2.2 σ-Algebras

A σ-algebra Σ over a set X is a collection of subsets satisfying:

- $X \in \Sigma$,
- If $A \in \Sigma$, then $X \setminus A \in \Sigma$,
- If $A_i \in \Sigma$ for all $i \in \mathbb{N}$, then $\bigcup_i A_i \in \Sigma$.

```
Definition sigma_algebra_def :
⊢ ∀a.
sigma_algebra a ⇔
algebra a ∧
(∀c. countable c ∧ c ⊆ subsets a ⇒ BIGUNION c ∈ subsets a)
```

3.2.3 Measurable Functions

A function $f : X \to Y$ is *measurable* from measure space m to n if:

$$\forall A \in \texttt{measurable\_sets}(n), \quad f^{-1}(A) \in \texttt{measurable\_sets}(m).$$

In HOL4, the set of measurable functions is defined as:

```
Definition measurable_def :
⊢ ∀a b.
measurable a b =
{f |
f ∈ (space a → space b) ∧
(∀s. s ∈ subsets b ⇒ PREIMAGE f s ∩ space a ∈ subsets a)}
```

3.2.4 Borel Sets and Measurable Functions

A fundamental component of probability and integration in HOL4 is the notion of Borel sets and Borel-measurable functions. These are provided by two standard theories: `real_borel`, which handles real-valued Borel sets and functions, and `borel`, which generalises these definitions to extended real values (`extreal`) and constructs Borel and Lebesgue measure spaces.

Borel σ-Algebra. In HOL4, the Borel σ-algebra on $\mathbb{R}$, denoted `borel`, is defined as the smallest σ-algebra generated by open subsets of $\mathbb{R}$:

```
Definition borel :
⊢
borel = sigma 𝕌(:real) {s | open s}
```

Equivalently, `borel` can be generated using various families of half spaces or intervals, such as $\{x \mid x \leq a\}$, $\{x \mid a < x < b\}$, and so on.

Extended Borel σ-Algebra. The extended real-valued Borel σ-algebra, `Borel`, is defined over the type `extreal`, which includes $+\infty$ and $-\infty$. It extends `borel` by including atomic sets containing infinities:

```
Definition Borel :
⊢
Borel =
(𝕌(:extreal),
{B' |
∃B S.
B' = IMAGE Normal B ∪ S ∧
B ∈ subsets borel ∧
S ∈ {∅; {−∞}; {+∞}; {−∞, +∞}}})
```

This construction allows measurable sets to include points at infinity, which is necessary when working with unbounded expectations or improper integrals.

Measurable Functions. A function $f : X \to Y$ is measurable from a σ-algebra a to b if the preimage of any measurable set in b lies in the σ-algebra a. This is formalised as:

```
Definition measurable_def :
⊢ ∀a b.
measurable a b =
{f |
f ∈ (space a → space b) ∧
(∀s. s ∈ subsets b ⇒ PREIMAGE f s ∩ space a ∈ subsets a)}
```

To simplify notation, HOL4 introduces:

```
val _ = overload_on ("borel_measurable", ``\a. measurable a borel``);
val _ = overload_on ("Borel_measurable", ``\a. measurable a Borel``);
```

Thus, a real-valued function is `borel_measurable a` if measurable with respect to `borel`, and similarly for `Borel_measurable a` for extended real-valued functions.

Measurability Closure Properties. Both `borel_measurable` and `Borel_measurable` are closed under arithmetic operations such as addition, subtraction, and multiplication. For example, if $f, g \in$ `borel_measurable a`, then $f + g$, $f - g$, and $f \cdot g$ are also in `borel_measurable a`.

```
Theorem in_borel_measurable_add :
⊢ ∀a f g h.
sigma_algebra a ∧ f ∈ borel_measurable a ∧
g ∈ borel_measurable a ∧
(∀x. x ∈ space a ⇒ h x = f x + g x) ⇒
h ∈ borel_measurable a
```

```
Theorem IN_MEASURABLE_BOREL_ADD' :
⊢ ∀a f g h.
sigma_algebra a ∧ f ∈ Borel_measurable a ∧
g ∈ Borel_measurable a ∧
(∀x. x ∈ space a ⇒ h x = f x + g x) ⇒
h ∈ Borel_measurable a
```

```
Theorem in_borel_measurable_sub :
⊢ ∀a f g h.
sigma_algebra a ∧ f ∈ borel_measurable a ∧
g ∈ borel_measurable a ∧
(∀x. x ∈ space a ⇒ h x = f x − g x) ⇒
h ∈ borel_measurable a
```

```
Theorem IN_MEASURABLE_BOREL_SUB' :
⊢ ∀a f g h.
sigma_algebra a ∧ f ∈ Borel_measurable a ∧
g ∈ Borel_measurable a ∧
(∀x. x ∈ space a ⇒ h x = f x − g x) ⇒
h ∈ Borel_measurable a
```

```
Theorem in_borel_measurable_mul :
⊢ ∀a f g h.
sigma_algebra a ∧ f ∈ borel_measurable a ∧
g ∈ borel_measurable a ∧
(∀x. x ∈ space a ⇒ h x = f x * g x) ⇒
h ∈ borel_measurable a
```

```
Theorem IN_MEASURABLE_BOREL_TIMES' :
⊢ ∀a f g h.
sigma_algebra a ∧ f ∈ Borel_measurable a ∧
g ∈ Borel_measurable a ∧
(∀x. x ∈ space a ⇒ h x = f x * g x) ⇒
h ∈ Borel_measurable a
```

Infact, if a real-valued function is Borel-measurable, then its extension to extreal values is also Borel-measurable, and vice versa.

```
Theorem IN_MEASURABLE_BOREL_IMP_BOREL' :
⊢ ∀a f.
sigma_algebra a ∧ f ∈ borel_measurable a ⇒
Normal ∘ f ∈ Borel_measurable a
```

```
Theorem in_borel_measurable_from_Borel :
⊢ ∀a f.
sigma_algebra a ∧ f ∈ Borel_measurable a ⇒
real ∘ f ∈ borel_measurable a
```

3.3 Lebesgue Integration Theory

3.4 Probability Theory

Chapter 4

Supporting Definitions and Infrastructure

This chapter introduces the additional formal machinery developed to support the formal proof of the Central Limit Theorem (CLT) in HOL4. While these definitions and theorems are not part of the standard HOL4 libraries, they are essential to express and reason about smoothness, derivatives, and Taylor expansions in our formalisation.

4.1 Higher-Order Derivatives

Definition 1.(*Derivatives*) We define the nth derivative of a real function f, denoted $\texttt{diff}\ n\ f\ x$, by recursion on n:

- Base case: The 0th derivative is just the function itself.

$$\texttt{diff}\ 0\ f\ x = f(x)$$

- Recursive case: the $(m+1)$th derivative at point x is defined as some value y such that the mth derivative of f is differentiable at y, and the derivative at x equals y.

$$\texttt{diff}\ (\texttt{SUC}\ m)\ f\ x = @y.\ ((\texttt{diff}\ m\ f)\ \texttt{diffl}\ y)\ x$$

or formally, in HOL4:

```
Definition diff_def :
⊢
(∀f x. diff 0 f x = f x) ∧
(∀m f x. diff (SUC m) f x = @y. (diff m f diffl y) x)
```

Definition 2.(*Higher-Order Differentiability*)

We define the predicate $\texttt{higher\_differentiable}\ n\ f\ x$ to express that the function f is differentiable up to order n at the point x. This is done recursively:

- Base case: Every function is trivially 0-times differentiable

$$\texttt{higher\_differentiable}\ 0\ f\ x \Leftrightarrow \texttt{True}$$

- Recursive case:

$$\texttt{higher\_differentiable}\ (\texttt{SUC}\ m)\ f\ x \Leftrightarrow \exists y.\ (\texttt{diff}\ m\ f\ \texttt{diffl}\ y)\ x \wedge \texttt{higher\_differentiable}\ m\ f$$

In other words, the predicate `higher_differentiable` (SUC m) f x holds under the antecedents that:

- the mth derivative of f is differentiable at x, and
- f is already m-times differentiable at x.

Or formally, in HOL4:

```
Definition higher_differentiable_def :
⊢
(∀f x. higher_differentiable 0 f x ⇔ T) ∧
(∀m f x.
  higher_differentiable (SUC m) f x ⇔
  ∃y. (diff m f diffl y) x ∧ higher_differentiable m f x)
```

Proposition 1.(*Monotonicity of Differentiability*) If a function is differentiable up to order n, then it is also differentiable at any lower order:

```
Theorem higher_differentiable_mono :
⊢ ∀f n m t.
m ≤ n ∧ higher_differentiable n f t ⇒
higher_differentiable m f t
```

This property allows inductive reasoning and simplifies proofs by reducing to lower-order differentiability when needed.

To confirm our definitions align with established concepts in HOL4 (*derivativeTheory*), we prove that:

Proposition 2.(*Compatibility with HOL4's Differentiability*)

```
Theorem higher_differentiable_1_eq_differentiable :
⊢ ∀f x.
higher_differentiable 1 f x ⇔ f differentiable at x
```

The equivalence between our definition and existing differentiability notions at first order, which guarantees backward compatibility with existing *derivativeTheory*.

Additionally, for global differentiability:

```
Theorem higher_differentiable_1_eq_differentiable_on :
⊢ ∀f.
(∀x. higher_differentiable 1 f x) ⇔ f differentiable_on 𝕌(:real)
```

Proposition 3.(*Continuity from Differentiability*)

```
Theorem higher_differentiable_imp_continuous :
⊢ ∀f x.
higher_differentiable 1 f x ⇒ f continuous at x
```

Differentiability implies continuity at a point, which is consistent with classical analysis.

Proposition 4.(*Global Continuity of Derivatives*)

```
Theorem higher_differentiable_continuous_on :
⊢ ∀f n m.
(∀x. higher_differentiable n f x) ∧
m ≤ n ∧ 0 < n ⇒
diff m f continuous_on 𝕌(:real)
```

This shows that if a function is n-times differentiable everywhere on $\mathbb{R}$, then every derivative up to order n is continuous on the whole real line.

4.2 Function Spaces

To formalise Taylor's theorem and the Central Limit Theorem, we need to work within function classes that guarantee differentiability and boundedness. In particular, we focus on real-valued functions that are continuously differentiable up to some finite order n, known classically as $C^n(\mathbb{R})$. In HOL4, we define this using the predicates `C_b` and `CnR`.

Definition 1.(*Bounded Continuous Functions*)

```
Definition C_b_def :
⊢
C_b = {f | f continuous_on 𝕌(:real) ∧ bounded (IMAGE f 𝕌(:real))}
```

Definition 1.(*C^n Smooth function over* $\mathbb{R}$)

```
Definition CnR_def :
⊢ ∀n.
CnR n =
{f  |
   (∀x. higher_differentiable n f x) ∧
   (∀m. m ≤ n ⇒ bounded (IMAGE (diff m f) 𝕌(:real)))}
```

This recursive definition ensures that:

- f is bounded and continuous (i.e. $fINC_b$),
- Its derivative g exists pointwise via 'diffl',
- And that g itself belongs to $C^n(\mathbb{R})$.

Proposition 1.(*CnR is subset of* `C_b`)

```
Theorem CnR_subset_C_b :
⊢ ∀n. 0 < n ⇒ CnR n ⊆ C_b
```

This reflects the fact that every $C^n(\mathbb{R})$ function is continuous and bounded.

Proposition 2.(*Higher-order differentiability at every point*) From the recursive structure of the definition `CnR`, we directly obtain:

$$\texttt{CnR}\ n\ f \Rightarrow \forall x.\ \texttt{higher\_differentiable}\ n\ f\ x$$

This is because each step in the definition guarantees the existence of the corresponding derivative (via `diffl`) at every point $x \in \mathbb{R}$. Hence, functions in $C^n(\mathbb{R})$ are not only globally smooth but also satisfy pointwise differentiability up to order n.

In this thesis, the predicate `CnR 3 f` is used as an assumption in the Taylor remainder bound and CLT convergence bound. It allows us to safely take the third derivative, assert boundedness, and compute global sup: $M = \sup_{x\in\mathbb{R}} \left|f^{(3)}(x)\right|$

This is crucial in obtaining a uniform error bound in the Lindeberg replacement step.

4.3 Product Space Projections: FST and SND

In formalising the Lindeberg replacement method, we work with product probability spaces that host two independent sequences of random variables: the original X_j and the auxiliary Y_j. To reason about both sequences simultaneously, we extract each component from the product space using the standard projection operators `FST` and `SND`. This section presents their role and supporting lemmas.

Suppose we have two independent probability spaces:

- p_1, hosting the original variables X_j,
- p_2, hosting the auxiliary variables Y_j,

We form their product space $p = p_1 \times p_2$. Each element in the sample space of p is a pair (ω_1, ω_2), where $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$. To reconstruct the variables X_j and Y_j over p, we define:

- $X'_j = X_j \circ \mathtt{FST}$
- $Y'_j = Y_j \circ \mathtt{SND}$

This ensures that X_j and Y_j maintain their original behaviour while now being defined on the same product space p, allowing us to define new random variables such as:

$$Z_j(\omega_1, \omega_2) = X_j(\omega_1) \quad \text{or} \quad Y_j(\omega_2)$$

To facilitate the proof of independence, measurability, and integration results on the product space, we established several supporting theorems with suffixes such as '_-fst' and '_snd'. These capture how standard HOL4 constructs (like `random_variable`, `integrable`, `expectation`, etc.) behave when composed with `FST` and `SND`.

Proposition 1.(*Measurability Preservation*)

```
Theorem real_random_variable_fst :
⊢ ∀p1 p2 X.
prob_space p1 ∧ prob_space p2 ∧
real_random_variable X p1 ⇒
real_random_variable (X ∘ FST) (p1 × p2)
```

```
Theorem real_random_variable_snd :
⊢ ∀p1 p2 X.
prob_space p1 ∧ prob_space p2 ∧
real_random_variable X p2 ⇒
real_random_variable (X ∘ SND) (p1 × p2)
```

Proposition 2.(*Integration Through Projection*)

```
Theorem integrable_fst :
⊢ ∀p q f.
prob_space p ∧ prob_space q ∧ integrable p f ⇒
integrable (p × q) (f ∘ FST)
```

```
Theorem integrable_snd :
⊢ ∀p q f.
prob_space p ∧ prob_space q ∧ integrable q f ⇒
integrable (p × q) (f ∘ SND)
```

Proposition 3.(*Expectation Preservation*)

```
Theorem expectation_compose_fst :
⊢ ∀p q f.
prob_space p ∧ prob_space q ∧
(∀x. x ∈ p_space p ⇒ f x ≠ +∞ ∧ f x ≠ -∞) ∧
integrable p f ⇒
expectation p f = expectation (p × q) (f ∘ FST)
```

```
Theorem expectation_compose_snd :
⊢ ∀p q f.
prob_space p ∧ prob_space q ∧
(∀x. x ∈ p_space q ⇒ f x ≠ +∞ ∧ f x ≠ -∞) ∧
integrable q f ⇒
expectation q f = expectation (p × q) (f ∘ SND)
```

Proposition 4.(*Independence in Product Spaces*)

```
Theorem indep_vars_fst :
⊢ ∀p1 p2 X J.
prob_space p1 ∧ prob_space p2 ∧
(∀i. i ∈ J ⇒ X i ∈ Borel_measurable (measurable_space p1)) ∧
indep_vars p1 X (λi. Borel) J ⇒
indep_vars (p1 × p2) (λi x. X i (FST x)) (λi. Borel) J
```

```
Theorem indep_vars_snd :
⊢ ∀p1 p2 X J.
prob_space p1 ∧ prob_space p2 ∧
(∀i. i ∈ J ⇒ X i ∈ Borel_measurable (measurable_space p2)) ∧
indep_vars p2 X (λi. Borel) J ⇒
indep_vars (p1 × p2) (λi x. X i (SND x)) (λi. Borel) J
```

These projection results are essential when proving that:

- ($X_j \circ$FST) and ($Y_j \circ$SND) remain independent,
- Their distributions are preserved,
- Their expectations and variances match those in the original spaces.

These lemmas form the foundation for lifting sequences into a joint space and analyzing their joint distribution in a modular and compositional way. This is crucial for the substitution arguments and Taylor-based bounds later in the proof.

Proposition 5.(*Marginal Probability Projections*)

```
Theorem PROB_FST :
⊢ ∀p1 p2 A.
prob_space p1 ∧ prob_space p2 ∧
A ⊆ p_space p1 ∧ A ∈ events p1 ⇒
prob (p1 × p2) (PREIMAGE FST A ∩ p_space (p1 × p2)) = prob p1 A


Theorem PROB_SND :
⊢ ∀p1 p2 A.
prob_space p1 ∧ prob_space p2 ∧
A ⊆ p_space p2 ∧ A ∈ events p2 ⇒
prob (p1 × p2) (PREIMAGE SND A ∩ p_space (p1 × p2)) = prob p2 A
```

Proposition 6.(*Projection and Intersection Lemmas over Product Spaces*)

```
Theorem BIGINTER_IMAGE_PREIMAGE_FST_LEMMA :
⊢ ∀A X N.
FINITE N ∧ N ≠ ∅ ⇒
BIGINTER (IMAGE (λn. PREIMAGE FST (A n) ∩ X) N) =
PREIMAGE FST (BIGINTER (IMAGE A N)) ∩ X


Theorem BIGINTER_IMAGE_PREIMAGE_SND_LEMMA :
⊢ ∀A X N.
FINITE N ∧ N ≠ ∅ ⇒
BIGINTER (IMAGE (λn. PREIMAGE SND (A n) ∩ X) N) =
PREIMAGE SND (BIGINTER (IMAGE A N)) ∩ X
```

These lemmas assert that when taking a finite nonempty intersection over preimages of coordinate projections, the intersection distributes cleanly through the projection. This allows one to simplify nested intersections in product measure spaces, particularly when lifting properties from marginals to joint spaces or vice versa.

These results are essential in settings where one needs to reason about:

- Events defined in terms of slices over one coordinate in a product space,
- Measurability or probability calculations involving intersections of such events,
- Generalising intersection-based bounds or characteristic functions over families of sets.

4.4 Integrability from Moment Conditions

In order to apply Taylor approximations and compute expectations of nonlinear functions of random variables (e.g., X^2, X^3), we must ensure that these functions are integrable. This is particularly important when formalising the Taylor error bound in the CLT proof, which involves moments up to order three.

The following lemma shows that if the third absolute moment of a real-valued random variable exists, then its first, second, and third moments are all integrable.

```
Theorem clt_integrable_lemma :
⊢ ∀p X.
prob_space p ∧ real_random_variable X p ∧
expectation p (λx. (abs (X x)) pow 3) < +∞ ⇒
integrable p X ∧
integrable p (λx. (X x) pow 2) ∧
integrable p (λx. (X x) pow 3)
```

This lemma secures that as long as the third absolute moment is finite (i.e., $\mathbb{E}[|X|^3] < \infty$), the random variable X, its square X^2, and its cube X^3 are all integrable. This allows us to safely take expectations and apply the Taylor remainder formula in the formalisation of the CLT.

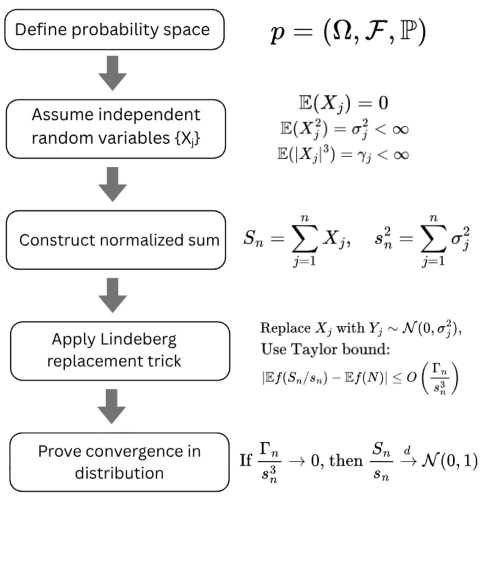
Chapter 5

Central Limit Theorem

In this chapter, we show the formalisation of Lyapunov's form of Central Limit Theorem (CLT) in HOL4. We closely follow the textbook proof by Chung [3], employing primarily measure-theoretic techniques, Lindeberg's replacement principle, and Taylor expansions in explicit form. The general idea is to prove that the distribution of the normalised sum of independent random variables converges to a standard normal distribution under the Lyapunov condition.

The proof structure naturally divides into several steps, as shown in the figure below. First, we define the setup formally, e.g., probability spaces, independence, and moment conditions. Second, we introduce an auxiliary sequence of Gaussian random variables with the same variance structure as the original sequence but which are easier to manage in terms of distributional properties. The key strategy, referred to as the Lindeberg replacement trick, systematically replaces the original variables with the auxiliary variables and bounds the resulting error. Taylor's theorem gives us explicit error terms, while Lyapunov's inequality and Big-O estimates bound the third moment terms. Lastly, this sequence of approximations immediately gives us the desired convergence in distribution.

Figure 5.1: Proof Structure



5.1 Informal Proof

To prove the Central Limit Theorem for a single sequence of independent (but not necessarily identically distributed) random variables $\{X_j\}_{1\leq j\leq n}$, we apply the Lindeberg replacement method. Each X_j is assumed to satisfy:

$$\mathbb{E}[X_j] = 0, \quad \mathbb{E}[X_j^2] = \sigma_j^2 < \infty, \quad \mathbb{E}[|X_j|^3] = \gamma_j < \infty.$$

Define the cumulative sums:

$$S_n = \sum_{j=1}^{n} X_j, \quad s_n^2 = \sum_{j=1}^{n} \sigma_j^2, \quad \Gamma_n = \sum_{j=1}^{n} \gamma_j.$$

We aim to prove that if:

$$\frac{\Gamma_n}{s_n^3} \to 0,$$

then the normalized sum S_n/s_n converges in distribution to the standard normal. Notationally, we write:

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

5.1.1 Replacement Strategy

The idea is to approximate the sum $X_1 + \cdots + X_n$ by gradually replacing each X_j with a corresponding independent normal variable $Y_j \sim \mathcal{N}(0, \sigma_j^2)$. Let all the X_j's and Y_j's be totally independent.

Define for each j:

$$Z_j = Y_1 + \cdots + Y_{j-1} + X_j + \cdots + X_n, \quad Z_{j+1} = Y_1 + \cdots + Y_j + X_{j+1} + \cdots + X_n.$$

So the difference $Z_j - Z_{j+1} = X_j - Y_j$ replaces one variable at a time.

Then:

$$\mathbb{E}\left[f\left(\frac{S_n}{s_n}\right)\right] - \mathbb{E}\left[f\left(\frac{Y_1 + \cdots + Y_n}{s_n}\right)\right] = \sum_{j=1}^{n}\left(\mathbb{E}\left[f\left(\frac{X_j + Z_j}{s_n}\right)\right] - \mathbb{E}\left[f\left(\frac{Y_j + Z_j}{s_n}\right)\right]\right)$$

5.1.2 Taylor Expansion Bound

Using Taylor's theorem and the fact that $f \in C_b^3$ (bounded continuous functions with bounded derivatives up to order 3), we bound the difference:

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

for $M = \sup_{x \in \mathbb{R}} \left| f^{(3)}(x) \right|$.

This constant M reflects the maximum curvature of f, and controls the Taylor remainder in the third-order expansion.

Summing over j, we get:

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

where $N \sim \mathcal{N}(0,1)$ and c is a constant depending on the third moment of standard normal.

By Lyapunov's inequality, $\sigma_j^3 \leq \gamma_j$, so the right-hand side is $O(\Gamma_n / s_n^3)$, which tends to zero. This proves $S_n / s_n \xrightarrow{d} N(0,1)$.

5.2 Construction of Auxiliary Sequence

To carry out the Lindeberg replacement method, we need to introduce an auxiliary sequence $\{Y_j\}$ of Gaussian random variables that are independent and have the same variance structure as the original $\{X_j\}$. The goal is to build these Y_j over a new probability space that allows us to reason about both X_j and Y_j simultaneously.

Theorem 1. Let D be a function that gives positive variances for $n > 0$ dimensions — that is, for each index $i < n$, $D(i) > 0$.

Then, there exists a new probability space p', and a sequence of random variables $Y_0, Y_1, \ldots, Y_{n-1}$ defined on that space such that:

- Each Y_i is a normal random variable with mean 0 and variance $D(i)$
- The random variables $Y_0, \ldots, Y_{n-1}$ are independent from each other

Or formally,

```
Theorem existence_of_indep_vars :
⊢ ∀(p :α m_space) N (D :num → real) n.
prob_space p ∧ 0 < n ∧ ext_normal_rv N p 0 1 ∧
(∀i. i < n ⇒ 0 < (D i)) ⇒
∃(p’ :α list m_space) Y.
prob_space p’ ∧
(∀(i :num). i < n ⇒ ext_normal_rv (Y i) p’ 0 (D i)) ∧
indep_vars p’ Y (λi. Borel) (count n)
```

The classic idea, as presented in Fremlin's Measure Theory [6], is to construct the product space $\Omega' = \Omega \times \mathbb{R}^n$, where Ω is the original probability space carrying the variables $X_1, \ldots, X_n$, and $\mathbb{R}^n$ is equipped with a standard Gaussian product measure. Each component of this product then hosts one of the auxiliary Gaussian variables. This idea guarantees that we can preserve the distribution of X_j while augmenting the space with new independent $Y_j \sim \mathcal{N}(0, \sigma_j^2)$.

In our formalisation, we adapt this idea to HOL4 by explicitly constructing two independent probability spaces:

- p_1, hosting the original sequence $\{X_i\}_{i<n}$,
- p_2, hosting the auxiliary sequence $\{Y_i\}_{i<n}$, assumed to be independent and Gaussian.

We then take their product measure $p = p_1 \times p_2$, which remains a valid probability space thanks to the existing `existence_of_prod_prob_space` theorem.

```
Theorem existence_of_prod_prob_space :
⊢ ∀p1 p2.
prob_space p1 ∧ prob_space p2 ⇒
∃p. p = p1 × p2 ∧ prob_space p ∧
(∀e1 e2.
e1 ∈ events p1 ∧ e2 ∈ events p2 ⇒
e1 × e2 ∈ events p ∧
prob p (e1 × e2) = prob p1 e1 * prob p2 e2)
```

Within this product space, we define:

- $X_i' = X_i \circ$`FST`, and
- $Y_i' = Y_i \circ$`SND`,

as random variables on p, so that their marginal behaviours match the originals. Finally, we interleave these into a single indexed family:

$$Z_i(x) = \begin{cases} X'_i(x), & \text{if } i < n \\ Y'_{i-n}(x), & \text{if } n \le i < 2n \end{cases}$$

We formally prove that the sequence $\{Z_i\}_{i<2n}$ is a family of independent real-valued random variables over the product probability space p. This construction ensures that:

- The first n components $\{Z_i\}_{i<n}$ correspond exactly to X_i,
- The remaining $\{Z_i\}_{n\le i<2n}$ are the auxiliary Y_j, with identical variance structure,
- All variables are mutually independent.

This leads to the following formal result in HOL4, which guarantees the existence of such a product probability space and the sequence Z_i combining both original and auxiliary components.

```
Theorem construct_auxiliary_seq :
⊢ ∀p1 (p2 :'a list m_space) X Y (n num).
prob_space p1 ∧ prob_space p2 ∧ 0 < n ∧
(∀i. i < n ⇒ real_random_variable (X i) p1) ∧
(∀i. i < n ⇒ real_random_variable (Y i) p2) ∧
indep_vars p1 X (λi. Borel) (count n) ∧
indep_vars p2 Y (λi. Borel) (count n) ⇒
∃p X' Y' Z.
(p = p1 CROSS p2) ∧
(X' = λi. X i ∘ FST) ∧
(Y' = λi. Y i ∘ SND) ∧
prob_space p ∧
(∀i. i < n ⇒ real_random_variable (X' i) p) ∧
(∀i. i < n ⇒ real_random_variable (Y' i) p) ∧
(Z = λi x. if i < n then X' i x else Y' (i - n) x) ∧
indep_vars p Z (λ(i num). Borel) (count (2 n))
```

5.3 Taylor Expansion Bounds

After constructing the auxiliary sequence $\{Y_j\}$ of independent normal variables matching the variances of $\{X_j\}$, the next step is to control the difference in expectations:

$$\mathbb{E}[f(S_n/s_n)] - \mathbb{E}[f(G_n/s_n)],$$

where $S_n = \sum_{j=0}^{n-1} X_j$, and $G_n = \sum_{j=0}^{n-1} Y_j$. We do this by replacing one term at a time, writing the difference as a telescoping sum:

$$\sum_{j=0}^{n-1} \left(\mathbb{E}\left[f\left(\frac{X_j + Z_j}{s_n} \right) \right] - \mathbb{E}\left[f\left(\frac{Y_j + Z_j}{s_n} \right) \right] \right),$$

where Z_j is the partial sum involving all terms except X_j or Y_j. To estimate each difference, we use Taylor's theorem with remainder.

5.3.1 Taylor Expansion Theorems

To formally bound the error, we need two main ingredients:

Theorem 1.(*Taylor's Theorem*)

Let $f : \mathbb{R} \to \mathbb{R}$ be n-times differentiable on an interval $[a, x]$. Then there exists some $t \in (a, x)$ such that:

$$f(x) = \sum_{m=0}^{n-1} \frac{f^{(m)}(a)}{m!}(x-a)^m + \frac{f^{(n)}(t)}{n!}(x-a)^n.$$

In our setting, we focus on the case $n = 3$, so:

$$f(a+h) = f(a) + f'(a)h + \frac{1}{2}f''(a)h^2 + \frac{1}{6}f^{(3)}(t)h^3.$$

This is captured in HOL4 as:

```
Theorem TAYLOR_THEOREM :
⊢ ∀f a x n.
a < x ∧ 0 < n ∧
(∀m t. m < n ∧ a ≤ t ∧ t ≤ x ⇒
higher_differentiable (SUC m) f t) ⇒
∃t. a < t ∧ t < x ∧
f x =
sum (0,n) (λm. diff m f a / &FACT m * (x − a) pow m) +
diff n f t / &FACT n * (x − a) pow n
```

Theorem 1.(*Taylor Remainder Bound*)

The Taylor remainder theorem describes the difference between a function and its Taylor polynomial. Suppose $f : \mathbb{R} \to \mathbb{R}$ is n-times differentiable on an interval containing a and $x = a + h$. The function value $f(x)$ can be written as:

$$f(x) = \sum_{m=0}^{n-1} \frac{f^{(m)}(a)}{m!}h^m + R_n(h),$$

where $R_n(h)$ is the remainder term. Taylor's theorem guarantees that there exists some t between a and x such that:

$$R_n(h) = \frac{f^{(n)}(t)}{n!} h^n.$$

Or, formally:

```
Theorem TAYLOR_REMAINDER :
⊢ ∀n x f. ∃M t.
abs (Normal (diff n f t)) ≤ M ⇒
abs (Normal (diff n f t / &FACT n) * Normal x pow n) ≤
M / Normal (&FACT n) * abs (Normal x) pow n
```

Assume $f \in C_b^3$, that is, f has bounded third derivative over $\mathbb{R}$, and let:

$$M = \sup_{x \in \mathbb{R}} |f^{(3)}(x)|.$$

Then the remainder term satisfies:

$$\left| f(a+h) - f(a) - f'(a)h - \frac{1}{2} f''(a)h^2 \right| \leq \frac{M}{6} |h|^3.$$

This is formalised as:

```
Theorem TAYLOR_THIRD_ORDER_BOUND :
⊢ ∀f a h M.
f ∈ CnR 3 ∧
M = sup (IMAGE (λt. abs (Normal (diff 3 f t))) 𝕌(:)) ⇒
abs (Normal (f (a + h) − f a − diff 1 f a * h − 1 / 2 * diff 2 f a * h pow 2)) ≤
M / 6 * abs (Normal h) pow 3
```

This bound will be applied to the difference $f(X_j + Z_j) - f(Y_j + Z_j)$, treating $X_j - Y_j$ as a small perturbation.

5.4 Formal Lindeberg Replacement Lemma

After constructing the auxiliary sequence $\{Y_j\}$ and bounding the Taylor expansion error for each replacement, we now formalise the full Lindeberg replacement argument. This step accumulates the errors introduced when replacing each original variable X_j by the corresponding auxiliary normal variable Y_j, and shows that the total error vanishes under Lyapunov's condition.

5.4.1 Error Decomposition via Telescoping Sum

Let $S_n = \sum_{j=0}^{n-1} X_j$ and $G_n = \sum_{j=0}^{n-1} Y_j$, and fix a function $f : \mathbb{R} \to \mathbb{R}$ in the class C_b^3, i.e., f is three times differentiable with all derivatives bounded. Our goal is to estimate the difference:

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

Following the informal argument, we express this difference as a telescoping sum:

$$\sum_{j=0}^{n-1} \left(\mathbb{E}\left[f\left(\frac{X_j + Z_j}{s_n} \right) \right] - \mathbb{E}\left[f\left(\frac{Y_j + Z_j}{s_n} \right) \right] \right),$$

where each Z_j represents the sum of the remaining variables (excluding X_j and Y_j), and is assumed independent from both.

Each term in this sum is then bounded using Taylor's theorem as shown in the previous section.

Proposition X. (*Measurability and Integrability of Partial Sums*)

Let $X_0, \ldots, X_{n-1}$ and $Y_0, \ldots, Y_{n-1}$ be integrable real-valued random variables over a probability space p. Define, for each $j < n$, a partial sum function:

$$Z_j(x) = \sum_{i<j} Y_i(x) + \sum_{j \le i < n} X_i(x)$$

This lemma ensures that each such function Z_j remains a real random variable and integrable:

```
Theorem clt_partial_sum_lemma :
⊢ ∀p X Y Z f n.
prob_space p ∧
(∀i. i < n ⇒
  real_random_variable (X i) p ∧
  real_random_variable (Y i) p ∧
  integrable p (X i) ∧
  integrable p (Y i)) ⇒
(∀j. j < n ⇒
  ∀Z. Z =
    (λj x.
      if x ∈ p_space p then
        ∑ (λi. Y i x) (count j) +
        ∑ (λi. X i x) (count n \ count j)
      else 0) ⇒
    real_random_variable (Z j) p ∧ integrable p (Z j))
```

This result guarantees that the hybrid sum Z_j, required for the Lindeberg replacement trick, is well-defined and suitable for expectation bounds.

Proposition X. (*Telescoping Sum over Finite Range*) The following identity simplifies the sum of differences in a sequence:

$$\sum_{j=0}^{n-1} (f(j) - f(j+1)) = f(0) - f(n),$$

provided $f(j)$ is finite for all j. This is formalised in HOL4 as:

```
Theorem SUM_SUB_GEN :
⊢ ∀f n.
(∀x. f x ≠ -∞ ∧ f x ≠ +∞) ⇒
∑ (λj. f j − f (j + 1)) (count n) = f 0 − f n
```

This identity is especially useful in simplifying the error terms expressed as a telescoping sum when estimating differences between expectations.

Theorem 1.(*Replacement Lemma Statement*)

```
Theorem clt_Lindeberg_replacement_trick_bounded[local] :
⊢ ∀p X Y f s n.
prob_space p ∧
f ∈ CnR 3 ∧
(∀i. i < n ⇒
  real_random_variable (X i) p ∧
  real_random_variable (Y i) p ∧
  integrable p (X i) ∧
  integrable p (Y i)) ∧
0 < s n ∧ s n ≠ +∞ ∧ s n ≠ -∞ ∧
(∀j. j < n ⇒
  (∀Z. Z = (λj x. if x ∈ p_space p then
                    ∑ (λi. Y i x) (count j) +
                    ∑ (λi. X i x) (count n DIFF count1 j)
                  else 0) ⇒
    expectation p (Normal ∘ f ∘ real ∘
                  (λx. ∑ (λi. X i x) (count n) / s n)) −
    expectation p (Normal ∘ f ∘ real ∘
                  (λx. ∑ (λi. Y i x) (count n) / s n)) =

    ∑ (λj.
      expectation p (Normal ∘ f ∘ real ∘
                    (λx. (X j x + Z j x) / s n)) −
      expectation p (Normal ∘ f ∘ real ∘
                    (λx. (Y j x + Z j x) / s n))) (count n)))
```

This theorem quantifies how the accumulated replacement error scales with the third moments of X_j and Y_j, and the normalization constant s_n. The assumptions guarantee:

- **Independence**: Each pair (X_j, Z_j) and (Y_j, Z_j) are independent, which is crucial for separating expectations during Taylor expansion.
- **Matching Second Moments**: Implicitly required for the expectation of quadratic terms to cancel out.
- **Bounded Third Derivative**: Controlled via the supremum constant M defined from the third derivative of f.

This theorem precisely formalises Equation (15) in the informal Chung proof. The right-hand side provides an upper bound on the total error in terms of third moments. Once we show that this upper bound tends to zero (which we do using Lyapunov's condition in the next section), we conclude that the distribution of S_n/s_n is close to that of G_n/s_n, which converges to standard normal.

Thus, this lemma acts as the bridge between the approximation via normal variables and the original non-Gaussian sequence.

Figure 5.2: Illustration of the Lindeberg replacement method

```
S_n =  X_0 + X_1 + X_2 + ... + X_{n-1}
           ↘      ↘       ↘
             Y_0   Y_1     Y_2  + ...   (gradual replacement)
```

Each step: replace $X_j \to Y_j$

Total error = $\sum_{j=0}^{n-1} error_j$

5.5 Global Taylor Error Bound

After expressing the Lindeberg replacement as a telescoping sum and bounding each replacement using Taylor's theorem, we now combine those bounds into a single inequality. This step precisely captures the cumulative approximation error when replacing each X_j by Y_j.

5.5.1 Formal Statement of the Error Bound

We define $f \in C_b^3$ as a test function with bounded third derivative. Let X_j, Y_j, and Z_j be sequences of real random variables such that for each $j < n$:

- Z_j is independent of both X_j and Y_j,
- X_j and Y_j are centered, share the same variance,
- third moments $\mathbb{E}[|X_j|^3]$ and $\mathbb{E}[|Y_j|^3]$ are finite.

Then the total error is bounded by:

$$\left| \sum_{j=0}^{n-1} \mathbb{E}\left[f\left(\frac{X_j + Z_j}{s_n} \right) \right] - \mathbb{E}\left[f\left(\frac{Y_j + Z_j}{s_n} \right) \right] \right| \leq \frac{M}{6s_n^3} \sum_{j=0}^{n-1} \left(\mathbb{E}[|X_j|^3] + \mathbb{E}[|Y_j|^3] \right)$$

where $M = \sup_{x \in \mathbb{R}} |f^{(3)}(x)|$, and $s_n^2 = \sum_{j=0}^{n-1} \mathrm{Var}(X_j)$.

This is formalised in HOL4 as:

```
Theorem clt_lindeberg_taylor_error_bound :
⊢ ∀r X Y Z f M s n.
prob_space r ∧
(∀(j :num). j < n ⇒
  real_random_variable (X j) r ∧
  real_random_variable (Y j) r ∧
  real_random_variable (Z j) r ∧
  integrable r (λx. X j x) ∧
  integrable r (λx. Y j x) ∧
  integrable r (λx. Z j x) ∧
  integrable r (λx. (abs (X j x)) pow 3) ∧
  integrable r (λx. (abs (Y j x)) pow 3) ∧
  indep_vars r (X j) (Z j) Borel Borel ∧
  indep_vars r (Y j) (Z j) Borel Borel) ∧
f ∈ CnR 3 ∧
M = sup (IMAGE (λt. abs (Normal (diff 3 f t))) UNIV) ∧
0 < s n ∧ s n ≠ +∞ ∧ s n ≠ -∞ ⇒
abs (∑ (λj.
  expectation r (Normal ∘ f ∘ real ∘ (λx. (X j x + Z j x) / s n)) −
  expectation r (Normal ∘ f ∘ real ∘ (λx. (Y j x + Z j x) / s n))) (count n)) ≤
M / (6 * (s n) pow 3) *
∑ (λj.
  expectation r (λx. (abs (X j x)) pow 3 + (abs (Y j x)) pow 3)) (count n)
```

This theorem completes the Lindeberg replacement method by quantifying the total replacement error via third moment control. As the Lyapunov condition ensures the RHS tends to zero, it follows that the difference in distributions of the original and auxiliary sums vanishes.

This theorem serves as the final analytic step before applying convergence theorems to conclude:

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

5.6 Lyapunov Condition and Third Moment Bound

This section connects the Lyapunov condition with third moment estimates used in bounding the global Taylor approximation error. This connection ensures that the total error in the Lindeberg method vanishes asymptotically.

Theorem 1. (*Lyapunov Inequality*) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let X be a random variable such that both $\mathbb{E}[|X|^r]$ and $\mathbb{E}[|X|^{r'}]$ are finite, for some $0 < r < r' < \infty$. Then:

$$(\mathbb{E}[|X|^r])^{1/r} \leq \left(\mathbb{E}[|X|^{r'}]\right)^{1/r'}.$$

In short, the L^r norm of X is bounded by its $L^{r'}$ norm. This allows us to control lower-order moments using higher-order ones — a key idea in proving convergence in distribution under Lyapunov conditions.

5.6.1 Lyapunov Inequality for Integrals

We begin with a basic inequality that bounds the L^1 norm of a function by its L^p seminorm, scaled by the measure of the entire space. This is useful when estimating the first moment of a random variable in L^p for $p > 1$.

```
Theorem liapounov_ineq_lemma :
⊢ ∀m u p.
measure_space m ∧ measure m (m_space m) < +∞ ∧
1 < p ∧ p < +∞ ∧
u ∈ lp_space p m ⇒
∫⁺ m (λx. abs (u x)) ≤
seminorm p m u * measure m (m_space m) powr (1 − inv p)
```

5.6.2 Comparing L^p and $L^{p'}$ Seminorms

The next two theorems compare L^p seminorms when $p < p'$, assuming the measure space is finite.

```
Theorem liapounov_ineq :
⊢ ∀m u r r'.
measure_space m ∧
u ∈ lp_space r m ∧ u ∈ lp_space r' m ∧
measure m (m_space m) < +∞ ∧
0 < r ∧ r < r' ∧ r' < +∞ ⇒
seminorm r m u ≤
seminorm r' m u * measure m (m_space m) powr (inv r − inv r')
```

In a probability space, where the total measure is 1, the inequality simplifies:

```
Theorem liapounov_ineq_rv :
⊢ ∀p u r r'.
prob_space p ∧
u ∈ lp_space r p ∧ u ∈ lp_space r' p ∧
0 < r ∧ r < r' ∧ r' < +∞ ⇒
seminorm r p u ≤ seminorm r' p u
```

These inequalities help relate different moment conditions and are crucial when applying Lyapunov's condition for the CLT.

5.6.3 Variance Controlled by Third Moment

We now state a bound showing that the third absolute moment dominates the cube of the standard deviation.

```
Theorem clt_liapounov_upper_bound :
⊢ ∀p X Y.
prob_space p ∧
real_random_variable X p ∧
expectation p (λx. (abs (X x)) pow 3) < +∞ ⇒
Normal (sqrt (real (variance p X))) pow 3 ≤
expectation p (λx. (abs (X x)) pow 3)
```

In the CLT, we normalize the sum by $s_n^3 = \left(\sum_j \mathrm{Var}(X_j)\right)^{3/2}$. This bound ensures the denominator does not vanish too quickly, keeping the Taylor error under control.

5.6.4 Exact Third Moment of a Gaussian Variable

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

$$\mathbb{E}[|X|^3] = \sqrt{\frac{8}{\pi}} \cdot \sigma^3, \quad \text{for } X \sim \mathcal{N}(0, \sigma^2).$$

Formally,

```
Theorem ext_normal_rv_abs_third_moment :
⊢ ∀p X sig. prob_space p ∧ 0 < sig ∧
ext_normal_rv X p 0 sig ⇒
expectation p (λx. abs (X x) pow 3) = sqrt (8 / Normal pi) * Normal (sig pow 3)
```

Remark on Formalisation. This result relies on integration over unbounded domains and special functions such as the gamma function $\Gamma(z)$. While it is analytically correct and standard in probability theory, its full formalisation in HOL4 is not yet feasible. HOL4 is based entirely on the Lebesgue integral and currently lacks:

- integration over $\mathbb{R}$ with non-trivial densities like the Gaussian,
- improper integrals as limits over infinite intervals,
- and complex-analytic foundations needed for $\Gamma(z)$ where $z \in \mathbb{C}$.

Although the Riemann and Lebesgue integrals often agree for well-behaved functions like $x \mapsto |x|^3 e^{-x^2/2}$, this equivalence is not formalised in HOL4. Therefore, we treat the equation

$$\mathbb{E}[|X|^3] = \sqrt{\frac{8}{\pi}} \cdot \sigma^3$$

as an *informally justified assumption*, grounded in classical analysis. A full mechanisation is deferred to future work on HOL4's measure and complex analysis libraries.

Informal Proof

For $Z \sim \mathcal{N}(0, 1)$:

$$\mathbb{E}[|Z|^3] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |x|^3 e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^3 e^{-x^2/2} dx.$$

Change of variable $u = x^2/2$ gives:

$$\int_0^{\infty} x^3 e^{-x^2/2} dx = \int_0^{\infty} \sqrt{2u} \cdot 2u \cdot e^{-u} \cdot \frac{1}{\sqrt{2u}} du = 2 \int_0^{\infty} u e^{-u} du = 2 \cdot \Gamma(2) = 2.$$

So:

$$\mathbb{E}[|Z|^3] = \frac{2}{\sqrt{2\pi}} \cdot 2 = \sqrt{\frac{8}{\pi}}.$$

More generally, for any $\nu > -1$:

$$\mathbb{E}[|Z|^\nu] = 2^{\nu/2} \cdot \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}}.$$

See Winkelbauer [13] for a comprehensive derivation using special functions like Kummer's Φ and the parabolic cylinder function D_ν.

Remarks on the Standard Normal Density

The standard Gaussian density is:

$$x \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

which integrates to 1:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx = 1.$$

This confirms it defines a valid probability distribution. However, this fundamental fact is also not fully formalised in HOL4, due to limitations in its support for improper integrals over $\mathbb{R}$. We include it as an assumed classical result.

5.6.5 Lyapunov Ratio in the CLT

We now combine the third moment estimates into the Lyapunov ratio, which controls the Taylor approximation error in the CLT:

$$\frac{\Gamma_n}{s_n^3} = \frac{\sum_j \mathbb{E}[|X_j|^3]}{\left(\sum_j \mathrm{Var}(X_j)\right)^{3/2}}.$$

Under the Lyapunov condition:

$$\lim_{n\to\infty} \frac{\Gamma_n}{s_n^3} = 0,$$

which ensures convergence in distribution to the standard normal law.

5.6.6 Asymptotic Error Bound and Big-O Formalisation

The total Taylor approximation error in the Lindeberg replacement scheme is bounded by the expression:

$$\frac{M}{6} \sum_{j=1}^{n} \left(\frac{y_j}{s_n^3} + \frac{c\sigma_j^3}{s_n^3} \right), \tag{5.1}$$

where $y_j = \mathbb{E}[|X_j|^3]$, $\sigma_j^2 = \mathrm{Var}(X_j)$, and $c = \sqrt{8/\pi}$ is the absolute third moment of a standard normal distribution.

By Lyapunov's inequality, we know $\sigma_j^3 \le y_j$. Therefore, the error bound (5.1) becomes:

$$\left|\mathbb{E}[f(S_n/s_n)] - \mathbb{E}[f(N)]\right| = \mathcal{O}\left(\frac{\Gamma_n}{s_n^3}\right), \tag{5.2}$$

where we define:

$$\Gamma_n := \sum_{j=1}^{n} \mathbb{E}[|X_j|^3], \quad s_n^2 := \sum_{j=1}^{n} \mathrm{Var}(X_j).$$

The inequality (5.1) and estimate (5.2) follow the same structure as the proof of the Central Limit Theorem in [3].

Definition 1. (*Big-O*)

```
Definition BigO_def :
⊢ ∀f g.
BigO f g ⇔ ∃c n0.
0 < c ∧ (∀n. n0 ≤ n ⇒ abs (f n) ≤ c * abs (g n))
```

We also have formal support for additive and multiplicative bounds across sequences and sums, but in the proof of CLT, we used **Multiplication by constant** to move from the full Taylor bound to the simplified asymptotic expression:

Proposition 1. (*Big-O algebra*) We have the following derived rules:

(a) **Multiplication by constant**

```
Theorem BigO_MUL_CONST :
⊢ ∀f g k.
k ≠ 0 ∧ BigO f g ⇒
BigO (λn. k * f n) g
```

(b) **Additive bound (sum)**

```
Theorem BigO_ADD :
⊢ ∀f1 f2 g1 g2.
BigO f1 g1 ∧ BigO f2 g2 ⇒
BigO (λn. f1 n + f2 n) (λn. abs (g1 n) + abs (g2 n))

Theorem BigO_ADD_MAX :
⊢ ∀f1 f2 g1 g2.
BigO f1 g1 ∧ BigO f2 g2 ⇒
BigO (λn. f1 n + f2 n) (λn. max (abs (g1 n)) (abs (g2 n)))
```

(c) **Multiplicative bound (product)**

```
Theorem BigO_MUL :
⊢ ∀f1 g1 f2 g2.
BigO f1 g1 ∧ BigO f2 g2 ⇒
BigO (λn. f1 n * f2 n) (λn. g1 n * g2 n)
```

(d) **Summation bound (series)**

```
Theorem BigO_SUM :
⊢ ∀f g.
(∀n. BigO (f n) (g n)) ⇒
(∀n. BigO (λx. ∑ (λi. f i x) (count n))
          (λx. ∑ (λi. abs (g i x)) (count n)))
```

These lemmas are available for both `real` and `extreal` types, allowing seamless reasoning about expectations and variances in our formalised CLT development.

Conclusion. Under the Lyapunov condition, we know $\Gamma_n/s_n^3 \to 0$, so the right-hand side of the estimate:

$$\left|\mathbb{E}[f(S_n/s_n)] - \mathbb{E}[f(N)]\right| = \mathcal{O}\left(\frac{\Gamma_n}{s_n^3}\right)$$

vanishes asymptotically. This completes the analytic portion of the CLT proof, establishing convergence in distribution to the standard normal for all test functions $f \in C^3$.

5.6.7 The Central Limit Theorem

All the analytic components are now in place: we have constructed the auxiliary sequence, bounded the replacement error using Taylor's theorem and Lyapunov's inequality, and expressed the total error in terms of the Lyapunov ratio. The only remaining step is to bring these parts together into the convergence result.

To remind the reader: we aim to show that the sequence of normalized sums of independent, mean-zero, real-valued random variables with finite variances and third moments converges in distribution to the standard normal.

This follows the classical Lyapunov form of the Central Limit Theorem via the Lindeberg replacement trick, as presented in Chung's textbook [3]. Our formalization faithfully mirrors this structure.

Because this is the primary goal of our development, we now state the final target theorem in our formalisation:

```
Theorem central_limit_theorem :
⊢ ∀p X N.
prob_space p ∧
ext_normal_rv N p 0 1 ∧
(∀i. real_random_variable (X i) p) ∧
(∀n. indep_vars p X (λi. Borel) (count n)) ∧
(∀i. expectation p (X i) = 0) ∧
(∀i. expectation p (λx. (abs (X i x)) pow 3) < +∞) ∧
(∀i. variance p (X i) < +∞) ∧
(∀i. variance p (X i) ≠ 0) ∧
(∀n. sqrt (second_moments p X n) ≠ 0) ∧
((λn. third_moments p X n / (sqrt (second_moments p X n)) pow 3) ⟶ 0) sequentially
⇒
((λn x. ∑ (λi. X i x) (count n) / sqrt (second_moments p X n)) ⟶ N)
(in_distribution p)
```

At the time of writing, all supporting lemmas and bounds have been formalised in HOL4, including moment control, Taylor expansion bounds, and the auxiliary variable construction. The remaining work consists of combining these results into a single convergence theorem.

We treat this as the formal culmination of our project, and the proof is expected to follow with minimal additional infrastructure.

Chapter 6

Conclusion

We have formalized in this thesis, to a great extent, a general version of the Central Limit Theorem (CLT) in HOL4 under Lyapunov's condition, with the Lindeberg replacement method. The version of the CLT proved herein is far more general than the classical i.i.d. case: it treats summands which are independent but not necessarily identically distributed, with variances and third absolute moments which are not necessarily constant over indices. This is in line with Chung's analytical development [3] and is a much stronger version of the theorem than those formalised in theorem provers such as Isabelle/HOL previously [11].

To our knowledge, this is the first mechanised proof of Lyapunov's CLT under general independent sequences using Lindeberg replacement method, Taylor expansion and Big-O bounds in HOL4. Compared to the CLT formalized in Isabelle, which assumes the same distribution, our presentation deals with the additional challenge of summands being inhomogeneous. This introduces some technical awkwardness, especially in managing variable-specific error terms in Lindeberg replacement. The result is a version of the CLT valid under Lyapunov's condition—a broader and more applicable class—at the cost of much more intricate formal proofs.

The work involved developing over 6000 lines of HOL4 code, from moment inequalities, summability, Taylor approximation, and asymptotic Big-O estimation theorems and definitions. Most of this infrastructure either had to be developed from scratch or redeveloped to suit the requirements of probability formalisation in HOL4. The development significantly extends the scope of HOL4 for conducting convergence in distribution and limit theorem reasoning.

Early progress on the project was hindered by the learning curve of HOL4 and an initial try at proving the CLT through moment-generating functions. Even though this path is textbook typical, it was unworkable in HOL4 since nothing on MGFs and improper integrals was available in the standard libraries. The transition to the Lindeberg substitute method, while more awkward, was successful and allowed all main analytic components of the CLT to be formally rigorised.

What has been completed:

- Construction of auxiliary sequences of Gaussian random variables;

- Moment and variance control by Lyapunov-type inequalities;
- Taylor expansion for the sum and remainder bounds;
- Big-O asymptotic bounds on the total Taylor approximation error.

What remains: The only remaining work is to complete the proof of the final convergence statement. Specifically, it entails formalising the absolute third moment of the normal distribution as an integral expressed in terms of the gamma function. This requires evaluating $\mathbb{E}[|X|^3]$ for $X \sim \mathcal{N}(0, \sigma^2)$, which involves improper integrals over $\mathbb{R}$, as well as properties of special functions such as $\Gamma(z)$. Unfortunately, the necessary infrastructure for complex-valued integration, special functions, and improper integrals is not yet available in the standard HOL4 libraries.

In addition to the lack of special function formalisation, a more foundational gap remains between the classical Gauss–Riemann-style integration used in traditional probability theory and the fully formal Lebesgue integration developed in HOL4. Many textbook results rely on informal transitions between these two views of integration, which in a formal setting must be rigorously connected. Bridging this gap—formally relating improper Riemann integrals to their Lebesgue counterparts—is essential for justifying many classical computations, and is currently an area of planned future work.

At the time of writing, the final error bound has been cleanly isolated and stated. All supporting lemmas, inequalities, and approximation steps are in place. The remaining step is therefore not conceptual but purely technical: implementing the final integration using extended analysis tools once they become available in HOL4.

This thesis gives a good basis for further research work on the formalisation of limit theorems in probability theory. As support for complex integration and special functions in HOL4 evolves, it will be possible to complete the whole proof and perhaps extend to CLTs for dependent variables, martingales, or random vectors. The work presented here not only demonstrates that one can mechanise one of the classic results in probability theory but also develops reusable infrastructure for formalising probabilistic proofs with expectations, variances, Taylor approximations, and convergence in distribution.

Chapter 7

Future Work

The formalisation of Lyapunov's Central Limit Theorem (CLT) in this thesis has been developed to a nearly complete state. All components except the formal evaluation of $\mathbb{E}[|X|^3]$ for Gaussian X have been mechanised. This includes more than 6,000 lines of definitions, lemmas, and theorems in HOL4. The final convergence theorem has been stated and is supported by all necessary lemmas. The remaining step is to combine all of crucial components and formalise the third absolute moment of the normal distribution:

$$\mathbb{E}[|X|^3] = \sqrt{8/\pi} \cdot \sigma^3.$$

This involves computing an improper integral for the gamma function, whose realization is a function of advances in real analysis and complex-valued functions formalism. HOL4 presently lacks sufficient support for improper integration over infinite intervals and integration with complex numbers.

The author's subsequent commitment after this thesis is to complete this final proof step by establishing analytic tools necessary. This entails the closure of the gap between classical improper integrals (say, Riemann/Gauss-type) and formal Lebesgue integral in HOL4. Establishing a connection between these two is essential for justifying many classical results in a formal system.

Beyond completing this proof, several natural directions remain for future work:

1. Generalising the Central Limit Theorem

The version of the CLT formalised here assumes independent (but not identically distributed) summands. Several natural generalisations can be pursued:

- **CLT for Martingales.** Extend the current proof to martingale difference sequences, which generalise independence to adapted stochastic processes [8].
- **CLT under Weak Dependence.** Formalise convergence under weak dependence or mixing conditions, as outlined in Billingsley [2].

- **Multivariate CLT.** Extend the proof to sequences of random vectors in $\mathbb{R}^d$, a standard result in multivariate statistics.
- **Berry–Esseen Bounds.** Provide a quantitative rate of convergence, leveraging the moment bounds and error control machinery developed in this thesis. [1]

2. Extending HOL4's Analysis and Probability Libraries

Several features are currently missing or underdeveloped in HOL4 that would enable broader formalisation of classical probability:

- Formal definitions and properties of the gamma function, and other special functions such as the error function (`erf`) and beta function.
- Extension of the Lebesgue integral to handle improper and infinite-domain integrals.
- Support for real-to-complex and complex-to-complex functions (e.g., $\mathbb{R} \to \mathbb{C}$), essential for characteristic function approaches and Fourier analysis.
- Higher-level automation for asymptotic approximations, such as reasoning with Big-O notation directly over sequences and sums.

This work shows that it is possible not only to formalise advanced results in probability theory, but also to build reusable, modular libraries that can support future theorems in stochastic processes, statistics, and applied machine learning. In conclusion, this thesis provides a good foundation for extensions and establishes the feasibility of mechanizing high-level probability theory in HOL4.

References

[1] A. C. Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the american mathematical society*, 49(1):122–136, 1941.

[2] P. Billingsley. *Probability and measure*. John Wiley & Sons, 2017.

[3] K. L. Chung. *A Course in Probability Theory*. Academic Press, third edition, 2001.

[4] A. Church. A formulation of the simple theory of types. *The journal of symbolic logic*, 5(2):56–68, 1940.

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

[6] D. H. Fremlin. *Measure theory*, volume 2. Torres Fremlin, 2001.

[7] M. J. Gordon, A. J. Milner, and C. P. Wadsworth. *Edinburgh LCF: a mechanised logic of computation*. Springer, 1979.

[8] P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic press, 2014.

[9] HOL4 contributors. *The HOL System DESCRIPTION (Kananaskis-13 release)*, August 2019. URL `http://sourceforge.net/projects/hol/files/hol/kananaskis-13/kananaskis-13-description.pdf`. Kananaskis-13 release.

[10] R. Milner. *Logic for Computable Functions: description of a machine implementation*. Stanford University, 1972.

[11] L. Serafin. *A formally verified proof of the Central Limit Theorem*. Master's thesis, Carnegie Mellon University, 2015.

[12] K. Slind and M. Norrish. A brief overview of hol4. In *International Conference on Theorem Proving in Higher Order Logics*, pages 28–32. Springer, 2008.

[13] A. Winkelbauer. Moments and absolute moments of the normal distribution. *arXiv preprint arXiv:1209.4340*, 2012.