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

Supervisor: Dr. Chun Tian

May 2025

© Copyright by Thi Cam Tu Phan, 2025

All Rights Reserved

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.

This thesis formalises one of the most fundamental 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 focus on Lyapunov's version of the CLT as presented in Chung [3], using Lindeberg's replacement method.

The Central Limit Theorem (Lyapunov form):

> Let $\{X_i\}$ be a sequence of independent real-valued random variables defined over a probability space p, with the following properties:
>
> - Each X_i has mean zero: $\mathbb{E}[X_i] = 0$;
> - Each X_i has finite, non-zero variance: $0 < \mathrm{Var}(X_i) < \infty$;
> - Each X_i has finite third absolute moment: $\mathbb{E}[|X_i|^3] < \infty$;
> - The sequence X_i is independent in the formal sense used in HOL4 (via `indep_vars` over Borel sets).
>
> Define the sum $S_n = \sum_{i=0}^{n-1} X_i$, and let the total variance be $s_n^2 = \sum_{i=0}^{n-1} \mathrm{Var}(X_i)$.
>
> If the Lyapunov condition holds:
>
> $$\frac{\sum_{i=0}^{n-1} \mathbb{E}[|X_i|^3]}{s_n^3} \to 0 \quad \text{as } n \to \infty,$$

then the normalised sum converges in distribution to the standard normal:

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

In HOL4, we express this formally as:

```
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 result is not yet complete, this work marks a significant advance toward a fully formalised CLT. Over 6000 lines of HOL4 code were written, including formal definitions, measure-theoretic infrastructure, Taylor approximations, moment bounds, and Big-O analysis.

An initial attempt to follow the moment-generating function route was abandoned due to limited HOL4 support for improper integrals and Laplace transforms. The current approach, based on Taylor expansions and the Lindeberg method, allowed all core analytic steps to be formalised. The only missing component involves evaluating the third absolute moment of the normal distribution, which requires gamma-function integrals not yet formalised in HOL4.

This remaining gap is technical, not conceptual. Once suitable analysis tools become available, the full convergence proof can be completed with minimal additional effort.

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 characteristic functions (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 as 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:

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

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:

   $$\frac{}{t \ \texttt{|-}\ t}$$

   - Example: From the assumption P, we conclude P.

2. **Reflexivity (REFL):**

   - States that any term is equal to itself.
   - Rule:

   $$\frac{}{\texttt{|-}\ t = t}$$

   - Example: For x : `int`, $x = x$ is always true.

3. **Beta Conversion (BETA_CONV):**

   - Applies substitution in lambda abstractions.
   - Rule:

   $$\frac{}{\texttt{|-}\ (\backslash x.t_1)t_2 = t_1[t_2/x]}$$

   - 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:

   $$\frac{\Gamma_1 \ \texttt{|-}\ t_1{=}t_1' \quad \cdots \quad \Gamma_n \ \texttt{|-}\ t_n{=}t_n' \quad \Gamma \ \texttt{|-}\ t[t_1,\ldots,t_n]}{\Gamma_1 \cup \cdots \cup \Gamma_n \cup \Gamma \ \texttt{|-}\ t[t_1',\ldots,t_n']}$$

   - Example: From $x = y$ and $P(x)$, infer $P(y)$.

5. **Abstraction (ABS):**

- Generalizes an equation by abstracting a variable.
- Rule:

$$\frac{\Gamma \ \texttt{|-}\ t_1 \ \texttt{=}\ t_2}{\Gamma \ \texttt{|-}\ (\backslash x.t_1) \ \texttt{=}\ (\backslash x.t_2)}$$

- 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:

$$\frac{\Gamma \ \texttt{|-}\ t}{\Gamma[\sigma_1,\ \dots\ ,\sigma_n/\alpha_1,\ \dots\ ,\alpha_n] \ \texttt{|-}\ t[\sigma_1,\ \dots\ ,\sigma_n/\alpha_1,\ \dots\ ,\alpha_n]}$$

- Example: Consider the polymorphic identity theorem: $\vdash \forall x.\ x = x$ This holds for any type α. Using `INST_TYPE`, we can instantiate the type variable to a specific type, such as `real`. The result is: $\vdash \forall x : \texttt{real}.\ x = x$ This instantiated version now applies to real numbers specifically, which is often necessary in the context of formalising analysis or probability theory, where many theorems apply to real-valued functions or variables.

7. **Discharging Assumptions (DISCH):**

- Converts an assumption into an implication.
- Rule:

$$\frac{\Gamma \ \texttt{|-}\ t_2}{\Gamma - \{t_1\} \ \texttt{|-}\ t_1 \ \texttt{==>}\ t_2}$$

- 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:

$$\frac{\Gamma_1 \ \texttt{|-}\ t_1 \ \texttt{==>}\ t_2 \qquad \Gamma_2 \ \texttt{|-}\ t_1}{\Gamma_1 \cup \Gamma_2 \ \texttt{|-}\ t_2}$$

- 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** ($\Longrightarrow$). 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
     ```
     T_DEF              |- T  = ((\x:bool. x) = (\x. x))
     ```
   - True is represented as the equality of two identical boolean functions.
2. **Logical Falsehood (F)**
   - Rule
     ```
     F_DEF           |- F  = !t. t
     ```
   - False is defined to satisfy any boolean implication.
3. **Negation ($\neg$)**
   - Rule
     ```
     NOT_DEF         |- ~  = (\t. t ==> F)
     ```
   - Negation is the implication of a boolean value leading to falsehood.
4. **Conjunction ($\wedge$)**
   - Rule
     ```
     AND_DEF         |- /\ = \t1 t2. !t. (t1 ==> t2 ==> t) ==> t
     ```
   - Conjunction is defined as a logical function that evaluates nested implications.
5. **Disjunction ($\vee$)**
   - Rule
     ```
     OR_DEF          |- \/ = \t1 t2. !t. (t1 ==> t) ==> (t2 ==> t) ==> t
     ```
   - Disjunction is expressed through sequential implications.
6. **Universal Quantifier ($\forall$)**
   - Rule

```
FORALL_DEF       |- !  = \P:'a->bool. P = (\x. T)
```

- Universality asserts that a predicate holds for all elements of a type.

7. **Existential Quantifier (∃)**

   - Rule

   ```
   EXISTS_DEF       |- ?  = \P:'a->bool. P($@ P)
   ```

   - 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:

8. **One-to-One (*One_One*)**

   - Rule

   ```
   ⊢ ONE_ONE = (λf. ∀x1 x2. f x1 = f x2 ⇒ x1 = x2)
   ```

9. **Onto (*Onto*)**

   - Rule

   ```
   ⊢ ONTO = (λf. ∀y. ∃x. y = f x)
   ```

   HOL includes the constant $Type\_Definition$, which defines new types as bijections of subsets of existing types:

10. **Type Definition (Type_Definition)**

    - Rule

    ```
    |- TYPE_DEFINITION (P:'a->bool) (rep:'b->'a) =
    (!x' x''. (rep x' = rep x'') ==> (x' = x'')) /\
    (!x. P x = (?x'. x = rep x'))
    ```

    - 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

```
⊢ ∀t. (t ⇔ T) ∨ (t ⇔ F)
```

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

2. **Eta Conversion (ETA_AX)**
   - Rule

     ```
     ⊢ ∀t. (λx. t x) = t
     ```

   - Eta conversion describes the extensionality of functions.

3. **Hilbert's Choice (SELECT_AX)**
   - Rule

     ```
     ⊢ ∀P x. P x ⇒ P (@ P)
     ```

   - This axiom relates the choice operator to existential quantification.

4. **Infinity (INFINITY_AX)**
   - Rule

     ```
     ⊢ ∃f. ONE_ONE f ∧ ¬ONTO f
     ```

   - 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 further fundamental building block for probability and integration in HOL4 is the theory of Borel sets and Borel-measurable functions. These are provided by two standard theories: `real_borel`, which considers real-valued Borel sets and functions, and `borel`, which extends these definitions to extended real values (`extreal`) and builds up 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, e.g $\{x \mid x \leq a\}$, $\{x \mid a < x < b\}$, 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 ∈ {∅; {−∞}; {+∞}; {−∞, +∞}}})
```

The construction allows for measurable sets to have points at infinity, as required 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 every set in b that is measurable is in the σ-algebra a.

Or, formally:

```
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

HOL4 formalisation of Lebesgue integration provides a strong foundation for reasoning about expectations, variances, and convergence in probability. This section briefly introduces the key constructs and theorems that we use throughout the Central Limit Theorem formalisation.

3.3.1 Positive Simple Functions and Integration

Nonnegative measurable functions in HOL4 are approximated by positive simple functions, piecewise constant functions taking finitely many values on measurable disjoint sets. A function f is represented as:

$$f(x) = \sum_{j=1}^{n} y_j \cdot \mathbf{1}_{A_j}(x), \quad \text{where } A_j \text{ are disjoint measurable sets.}$$

This is formalised in HOL4 using:

```
Theorem pos_simple_fn_def :
⊢ ∀m f s a x.
pos_simple_fn m f s a x ⇔
(∀t. t ∈ m_space m ⇒ 0 ≤ f t) ∧
(∀t. t ∈ m_space m ⇒
        f t = ∑ (λi. Normal (x i) * 𝟙 (a i) t) s) ∧
(∀i. i ∈ s ⇒ a i ∈ measurable_sets m) ∧
FINITE s ∧
(∀i. i ∈ s ⇒ 0 ≤ x i) ∧
(∀i j. i ∈ s ∧ j ∈ s ∧ i ≠ j ⇒
        DISJOINT (a i) (a j)) ∧
BIGUNION (IMAGE a s) = m_space m
```

The integral of such a function is defined as:

```
Definition pos_simple_fn_integral_def :
⊢ ∀m s a x.
pos_simple_fn_integral m s a x =
∑ (λi. Normal (x i) * measure m (a i)) s
```

3.3.2 Positive and General Lebesgue Integral

The positive integral of a nonnegative function f is defined as the supremum of the integrals of all positive simple functions below f:

```
Definition pos_fn_integral_def :
⊢ ∀m f.
∫⁺ m f =
sup {r | (∃g. r ∈ psfis m g ∧
                ∀x. x ∈ m_space m ⇒ g x ≤ f x)}
```

The general Lebesgue integral of a function f is defined as:

$$\int f = \int^{+} f^{+} - \int^{+} f^{-}$$

where $f^{+} = \max(f, 0)$ and $f^{-} = \max(-f, 0)$.

```
Definition integral_def :
⊢ ∀m f.
∫ m f = ∫⁺ m f⁺ − ∫⁺ m f⁻
```

```
Definition fn_plus_def :
⊢ ∀f.
f⁺ = (λx. if 0 < f x then f x else 0)
```

```
Definition fn_minus_def :
⊢ ∀f.
f⁻ = (λx. if f x < 0 then -f x else 0)
```

3.3.3 Integrability

A function is integrable if it is Borel-measurable and the integrals of both its positive and negative parts are finite:

```
Definition integrable_def :
⊢ ∀m f.
integrable m f ⇔
f ∈ Borel_measurable (measurable_space m) ∧
∫⁺ m f⁺ ≠ +∞ ∧
∫⁺ m f⁻ ≠ +∞
```

3.3.4 Almost Everywhere and Null Sets

A property $P(x)$ holds almost everywhere (AE) if it holds outside of a null set:

```
Definition AE_DEF :
⊢ ∀m P.
(AE x::m. P x) ⇔
∃N. null_set m N ∧
∀x. x ∈ m_space m DIFF N ⇒ P x
```

3.3.5 Monotone Convergence Theorem

In approximating general nonnegative functions by simple functions, the following theorem allows taking limits under the integral:

```
Theorem lebesgue_monotone_convergence :
⊢ ∀m f fi.
measure_space m ∧
(∀i. fi i ∈ Borel_measurable (measurable_space m)) ∧
(∀i x. x ∈ m_space m ⇒ 0 ≤ fi i x) ∧
(∀x. x ∈ m_space m ⇒
   mono_increasing (λi. fi i x)) ∧
(∀x. x ∈ m_space m ⇒
   sup (IMAGE (λi. fi i x) 𝕌(:num)) = f x) ⇒
∫⁺ m f = sup (IMAGE (λi. ∫⁺ m (fi i)) 𝕌(:num))
```

3.3.6 Fubini and Tonelli Theorems

In formalising the Lindeberg replacement method, we often integrate over product spaces involving independent sequences of random variables. To justify iterated integrals in these settings, we rely on HOL4's formalisation of Tonelli's and Fubini's Theorems, which ensure that integrals over product measure spaces can be computed as nested one-dimensional integrals.

Tonelli's Theorem. Tonelli's theorem applies to nonnegative functions and guarantees that the double integral equals the iterated integral regardless of whether the value is finite or infinite:

```
Theorem TONELLI :
⊢ ∀X Y A B u v f.
sigma_finite_measure_space (X,A,u) ∧
sigma_finite_measure_space (Y,B,v) ∧
f ∈ Borel_measurable ((X,A) × (Y,B)) ∧
(∀s. s ∈ X × Y ⇒ 0 ≤ f s) ⇒
(∀y. y ∈ Y ⇒ (λx. f (x,y)) ∈ Borel_measurable (X,A)) ∧
(∀x. x ∈ X ⇒ (λy. f (x,y)) ∈ Borel_measurable (Y,B)) ∧
(λx. ∫⁺ (Y,B,v) (λy. f (x,y))) ∈ Borel_measurable (X,A) ∧
(λy. ∫⁺ (X,A,u) (λx. f (x,y))) ∈ Borel_measurable (Y,B) ∧
∫⁺ ((X,A,u) × (Y,B,v)) f =
∫⁺ (Y,B,v) (λy. ∫⁺ (X,A,u) (λx. f (x,y))) ∧
∫⁺ ((X,A,u) × (Y,B,v)) f =
∫⁺ (X,A,u) (λx. ∫⁺ (Y,B,v) (λy. f (x,y)))
```

Fubini's Theorem. Fubini's theorem generalises Tonelli's result to integrable (not necessarily nonnegative) functions. It justifies interchanging the order of integration under integrability assumptions:

```
Theorem FUBINI :
⊢ ∀X Y A B u v f.
sigma_finite_measure_space (X,A,u) ∧
sigma_finite_measure_space (Y,B,v) ∧
f ∈ Borel_measurable ((X,A) × (Y,B)) ∧
(∫⁺ ((X,A,u) × (Y,B,v)) (abs ∘ f) ≠ +∞ ∨
 ∫⁺ (Y,B,v) (λy. ∫⁺ (X,A,u) (λx. (abs ∘ f) (x,y))) ≠ +∞ ∨
 ∫⁺ (X,A,u) (λx. ∫⁺ (Y,B,v) (λy. (abs ∘ f) (x,y))) ≠ +∞) ⇒
∫⁺ ((X,A,u) × (Y,B,v)) (abs ∘ f) ≠ +∞ ∧
∫⁺ (Y,B,v) (λy. ∫⁺ (X,A,u) (λx. (abs ∘ f) (x,y))) ≠ +∞ ∧
∫⁺ (X,A,u) (λx. ∫⁺ (Y,B,v) (λy. (abs ∘ f) (x,y))) ≠ +∞ ∧
integrable ((X,A,u) × (Y,B,v)) f ∧
(AE y::(Y,B,v). integrable (X,A,u) (λx. f (x,y))) ∧
(AE x::(X,A,u). integrable (Y,B,v) (λy. f (x,y))) ∧
integrable (X,A,u) (λx. ∫ (Y,B,v) (λy. f (x,y))) ∧
integrable (Y,B,v) (λy. ∫ (X,A,u) (λx. f (x,y))) ∧
∫ ((X,A,u) × (Y,B,v)) f =
∫ (Y,B,v) (λy. ∫ (X,A,u) (λx. f (x,y))) ∧
∫ ((X,A,u) × (Y,B,v)) f =
∫ (X,A,u) (λx. ∫ (Y,B,v) (λy. f (x,y)))
```

In this thesis, these theorems are applied in constructing auxiliary sequences of independent random variables from product spaces and evaluating their expectations. Their use guarantees that swapping integration order is mathematically and formally valid in our HOL4 development.

3.3.7 $\mathcal{L}^p$ Spaces and Fundamental Inequalities

In HOL4, the $\mathcal{L}^p$ space over a measure space m is defined as the set of (extended) real-valued Borel-measurable functions whose p-th absolute power is integrable:

```
Theorem lp_space_alt_finite :
⊢ ∀p m f.
0 < p ∧ p ≠ +∞ ⇒
(f ∈ lp_space p m ⇔
 f ∈ Borel_measurable (measurable_space m) ∧
 ∫⁺ m (λx. abs (f x) powr p) ≠ +∞)
```

The associated seminorm is defined as:

$$\|f\|_p := \left(\int |f(x)|^p \, d\mu(x) \right)^{1/p}$$

```
Theorem seminorm_normal :
⊢ ∀p m f.
0 < p ∧ p ≠ +∞ ⇒
seminorm p m f = ∫⁺ m (λx. abs (f x) powr p) powr p⁻¹
```

This seminorm forms the basis for all moment bounds and continuity results used in the CLT proof. In particular, the finiteness of third moments used in Lyapunov's condition corresponds to membership in $\mathcal{L}^3$.

Hölder's Inequality. For functions $u \in \mathcal{L}^p(m)$ and $v \in \mathcal{L}^q(m)$ where $1/p + 1/q = 1$, Hölder's inequality bounds their product:

```
Theorem Hoelder_inequality :
⊢ ∀m u v p q.
measure_space m ∧ 0 < p ∧ 0 < q ∧
p⁻¹ + q⁻¹ = 1 ∧
u ∈ lp_space p m ∧ v ∈ lp_space q m ⇒
integrable m (λx. u x * v x) ∧
∫ m (λx. abs (u x * v x)) ≤ seminorm p m u * seminorm q m v
```

Minkowski's Inequality. Also known as the triangle inequality for $\mathcal{L}^p$ norms:

```
Theorem Minkowski_inequality :
⊢ ∀p m u v.
measure_space m ∧ 1 ≤ p ∧
u ∈ lp_space p m ∧ v ∈ lp_space p m ⇒
(λx. u x + v x) ∈ lp_space p m ∧
seminorm p m (λx. u x + v x) ≤
seminorm p m u + seminorm p m v
```

Cauchy–Schwarz Inequality. This is Hölder's inequality for the particular case $p = q = 2$, which assures us that the inner products of square-integrable functions are bounded:

```
Theorem Cauchy_Schwarz_inequality :
⊢ ∀m u v.
measure_space m ∧ u ∈ L2_space m ∧ v ∈ L2_space m ⇒
integrable m (λx. u x * v x) ∧
∫ m (λx. abs (u x * v x)) ≤ seminorm 2 m u * seminorm 2 m v
```

These inequalities are foundational tools in functional analysis. In this thesis, only Hölder's inequality is explicitly used—namely, in the proof of the Lyapunov inequality ('liapounov_ineq_lemma') to relate L^p norms. While Minkowski's and Cauchy–Schwarz inequalities are stated here for completeness, they were not directly invoked in the formalisation. The Taylor approximation error in the CLT proof is instead bounded via a custom lemma using Lyapunov's inequality and third moment control.

3.4 Probability Theory

This section reviews the essential formal components of probability theory used throughout the formalisation of the Central Limit Theorem (CLT) in HOL4. We focus on probability spaces, random variables, independence, expectation, variance, and convergence of random sequences—only including results directly used or closely related to our proof.

3.4.1 Probability Spaces

A probability space is a triple $(\Omega, \mathscr{F}, \mathscr{P})$, where Ω is the sample space, $\mathscr{F}$ is a σ-algebra of measurable events, and $\mathscr{P}$ is a probability measure satisfying $\mathscr{P}(\Omega) = 1$. In HOL4, this is encoded as:

```
Definition prob_space_def :
⊢ ∀p.
prob_space p ⇔
measure_space p ∧ measure p (m_space p) = 1
```

All the fundamental components of a probability space—event sets, probability values, and sample spaces—are accessed via:

$$\texttt{events p} = \texttt{measurable\_sets p}, \quad \texttt{prob p} = \texttt{measure p}, \quad \texttt{p\_space p} = \texttt{m\_space p}$$

3.4.2 Random Variables and Distributions

A random variable is a measurable function $X : \Omega \to \mathbb{R}$ from a probability space to the Borel σ-algebra. HOL4 defines real-valued random variables as:

```
Definition real_random_variable_def :
⊢ ∀X p.
real_random_variable X p ⇔
random_variable X p Borel ∧
∀x. x ∈ p_space p ⇒ X x ≠ -∞ ∧ X x ≠ +∞
```

The distribution of a random variable is the pushforward measure of X, and the corresponding cumulative distribution function (cdf) is:

$$\texttt{distribution\_function p X t} = \texttt{prob p}\{x \in p\_space p | X x \leq t\}$$

3.4.3 Mathematical Expectation and Moments

Expectation is defined as Lebesgue integration over a probability space:

```
Definition expectation_def :
⊢ expectation = ∫
```

A variable is integrable if its absolute value has finite integral. Variance is defined as the second central moment:

```
Theorem variance_eq :
⊢ ∀p X.
prob_space p ∧ real_random_variable X p ∧
integrable p (λx. (X x){2) ⇒
variance p X =
expectation p (λx. (X x){2) - (expectation p X){2
```

This thesis frequently verifies integrability and bounds on moments, e.g., when applying Lyapunov's condition or Taylor approximation.

3.4.4 Independence

Random variables X_i are independent if for all measurable sets A_i, their inverse images under X_i form independent events. HOL4 formalises total independence of a sequence via:

```
Definition indep_vars_def :
⊢ ∀p X A J.
indep_vars p X A J ⇔
∀E N.
  N ⊆ J ∧ N ≠ ∅ ∧ FINITE N ∧
  E ∈ N ⇒ subsets ∘ A ⇒
  prob p (BIGINTER (IMAGE (λn. PREIMAGE (X n) (E n) ∩ p_space p) N)) =
  ∏ (prob p ∘ (λn. PREIMAGE (X n) (E n) ∩ p_space p)) N
```

```
Definition indep_rv_def :
⊢ ∀p X Y s t.
indep_vars p X Y s t ⇔
∀a b.
  a ∈ subsets s ∧ b ∈ subsets t ⇒
  indep p (PREIMAGE X a ∩ p_space p) (PREIMAGE Y b ∩ p_space p)
```

We use this extensively when constructing auxiliary sequences for the Lindeberg method.

3.4.5 Convergence of Random Sequences

In our formalisation of the Central Limit Theorem, we primarily rely on *convergence in distribution*, which captures the idea that the sequence of random variables X_n becomes increasingly close to a limiting variable Y in terms of their probability distributions.

Formally, convergence in distribution is defined in HOL4 as:

```
Theorem converge_in_dist_alt :
⊢ ∀p X Y.
prob_space p ∧ (∀n. real_random_variable (X n) p) ∧
real_random_variable Y p ⇒
((X ⟶ Y) (in_distribution p) ⇔
  (λn. distribution p (X n)) ⟶ distribution p Y)
```

This equivalence allows us to reduce convergence in distribution to pointwise convergence of the associated distributions.

A more analytic characterisation—useful in proofs using Taylor approximations and moment estimates—is given by:

```
Theorem converge_in_dist_alt' :
⊢ ∀p X Y.
prob_space p ∧ (∀n. real_random_variable (X n) p) ∧
real_random_variable Y p ⇒
((X ⟶ Y) (in_distribution p) ⇔
 ∀f. bounded (IMAGE f 𝕌(:real)) ∧
        f continuous_on 𝕌(:real) ⇒
   ((λn. expectation p (Normal ∘ f ∘ real ∘ X n)) ⟶
    expectation p (Normal ∘ f ∘ real ∘ Y)) sequentially)
```

This version states that convergence in distribution is equivalent to the convergence of expectations for all bounded continuous test functions f, which aligns closely with the Portmanteau theorem.

In our CLT proof, this form enables us to focus on expectations of composed functions and verify the convergence using Taylor expansion and moment bounds, without needing pointwise or L^p convergence of random variables.

3.4.6 Summary

This section outlines only the essential building blocks from HOL4's formalisation of probability theory that are directly used in the formalisation of the CLT using the Lindeberg method. In particular, it sets up the notion of independence, variance control, and convergence that underpin the proof strategy.

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 4.1 (Higher-order Derivative Function)**.** We define the *n*th derivative of a real function f, denoted $\mathtt{diff}\ n\ f\ x$, by recursion on n:

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

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

- 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.

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

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 4.2 (Higher-order Differentiability)**.**

We define the predicate $\mathtt{higher\_differentiable}\ n\ f\ x$ 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 $\texttt{higher\_differentiable}\ (\texttt{SUC}\ m)\ f\ x$ 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 4.1.1 (Monotonicity of Differentiability). If a function is differentiable up to order n, then it is also differentiable at any lower order.

In HOL4, it formalised as:

```
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 4.1.2 (First-Order Compatibility).

```
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 4.1.3 (Differentiability Implies Continuity).

```
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 4.1.4 (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 4.3 (Bounded Continuous Functions `C_b`).

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

Definition 4.4 (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. $f INC_b$),
- Its derivative g exists pointwise via ‘diffl‘,
- And that g itself belongs to $C^n(\mathbb{R})$.

Proposition 4.2.1 (CnR ⊆ 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:

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

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 4.3.1 (Measurability Under Projections).

```
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 4.3.2 (Integrability Under Projections).

```
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 4.3.3 (Expectation Under Projection).

```
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 4.3.4 (Independence Under Projection).

```
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∘FST) and (Y_j∘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 4.3.5 (Marginal Probability Projection).

```
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 4.3.6 (Projection and Intersection).

```
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.

Lemma 4.4.1 (Integrability from Moment Condition).

```
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.

4.5 Real Convergence of `extreal` Sequences

In this section, we provide lemmas that connect convergence of extreal sequences to standard real-valued convergence. These results are necessary since many existing limit theorems in HOL4 are only available for real sequences, whereas probabilistic constructs (e.g. expectations, variances) often use extreal values.

Lemma 4.5.1 (Convergence of `extreal` Sequences via `real`). *Let $f : \mathbb{N} \to extreal$ be a sequence and $l \in extreal$. If there exists N such that $\forall n \geq N . f(n) \in \mathbb{R}$, and $l \in \mathbb{R}$, then*

$$(f \to l) \; in \; \mathtt{extreal} \iff (real(f(n)) \to real(l)) \; in \; \mathbb{R}.$$

Formally:

```
Theorem lim_null_equiv_extreal_real :
⊢ ∀f l.
(∃N. ∀n. N ≤ n ⇒ f n ≠ +∞ ∧ f n ≠ -∞) ∧
l ≠ +∞ ∧ l ≠ -∞ ⇒
((λx. f x - l) ⟶ 0) sequentially ⇔
((λn. real (f n) - real l) ⟶ 0) sequentially
```

4.6 Finite Sums and Products over `extreal` Functions

This section collects auxiliary lemmas about finite sums and products of extreal-valued functions. These results enrich the HOL4 library by providing convenient algebraic identities and inequalities over extreal, which are not available in the standard distribution. They are frequently used in our CLT formalisation to manipulate bounds, normalisations, and absolute values in sum expressions.

Lemma 4.6.1 (`Product with Default One Outside Subset`). *If f is 1 outside a finite subset $s \subseteq t$, then*

$$\prod_{x\in t} f(x) = \prod_{x\in s} f(x).$$

```
Theorem EXTREAL_PROD_IMAGE_SUPPORT :
⊢ ∀s t f.
FINITE s ∧ FINITE t ∧ s ⊆ t ∧
(∀x. x ∈ t DIFF s ⇒ f x = 1) ⇒
∏ f t = ∏ f s
```

```
Theorem EXTREAL_PROD_IMAGE_SUPPORT' :
⊢ ∀s t f.
FINITE t ∧ FINITE s ∧ s ⊆ t ⇒
∏ (λx. if x ∈ s then f x else 1) t = ∏ f s
```

Lemma 4.6.2 (`Bounding Sum of Absolutes by Dominating Function`). *If each $|f(x)| \leq g(x)$, then*

$$\left|\sum_{x\in s} f(x)\right| \leq \sum_{x\in s} g(x).$$

```
Theorem EXTREAL_SUM_IMAGE_ABS_LE :
⊢ ∀f g s.
FINITE s ∧ (∀x. x ∈ s ⇒ abs (f x) ≤ g x) ⇒
abs (∑ f s) ≤ ∑ g s
```

Lemma 4.6.3 (`Triangle Inequality for Finite extreal Sums`). *Triangle inequality for finite `extreal` sums:*

$$\left|\sum_{x\in s} f(x)\right| \leq \sum_{x\in s} |f(x)|.$$

```
Theorem EXTREAL_SUM_IMAGE_ABS_TRIANGLE :
⊢ ∀f s.
FINITE s ⇒
abs (∑ f s) ≤ ∑ (λx. abs (f x)) s
```

4.7 Additional Analytical and Measure-Theoretic Lemmas

This section collects technical results developed throughout the project to support formalisation of the CLT. These lemmas concern measurability, composition, powers, and integrability properties of real-valued functions.

Lemma 4.7.1 (Big Intersection of Measurable Sets). *Let m be a measure space, and suppose we take the intersection of a finite non-empty collection of measurable sets. Then the intersection remains measurable:*

If $\mathtt{FINITE}(s) \wedge s \neq \emptyset \wedge \forall x \in s.\, x \in \mathtt{measurable\_sets}(m) \Rightarrow \mathtt{BIGINTER}(s) \in \mathtt{measurable\_sets}(m)$.

```
Theorem MEASURABLE_BIGINTER :
⊢ ∀m s.
measure_space m ∧ FINITE s ∧ s ≠ ∅ ∧
(∀x. x ∈ s ⇒ x ∈ measurable_sets m) ⇒
BIGINTER s ∈ measurable_sets m
```

Lemma 4.7.2 (Preimage of Borel Set is Measurable). *Let f be a Borel-measurable function over some measurable space. Then the preimage of any Borel set (restricted to the measure space) is also measurable:*

$$\begin{aligned} &f \in \mathtt{Borel\_measurable}(\mathtt{measurable\_space}(m)) \Rightarrow \\ &\forall s \in \mathtt{subsets\ Borel},\ \mathtt{PREIMAGE}(f, s) \cap \mathtt{m\_space}(m) \in \mathtt{measurable\_sets}(m). \end{aligned} \tag{4.5}$$

```
Theorem MEASURABLE_PREIMAGE_BOREL :
⊢ ∀f m.
f ∈ Borel_measurable (measurable_space m) ⇒
∀s. s ∈ subsets Borel ⇒
PREIMAGE f s ∩ m_space m ∈ measurable_sets m
```

```
Theorem measurable_preimage_borel :
⊢ ∀f m.
f ∈ borel_measurable (measurable_space m) ⇒
∀s. s ∈ subsets borel ⇒
PREIMAGE f s ∩ m_space m ∈ measurable_sets m
```

Theorem 4.7.3 (Measurability of Power Functions). *If f is a Borel measurable function on a, and f^n is defined pointwise by $g(x) = f(x)^n$, then g is also measurable.*

```
Theorem IN_BOREL_MEASURABLE_POW :
⊢ ∀a n f g.
sigma_algebra a ∧ f ∈ Borel_measurable a ∧
(∀x. x ∈ space a ⇒ g x = f x pow n) ∧
(∀x. x ∈ space a ⇒ f x ≠ -∞ ∧ f x ≠ +∞) ⇒
g ∈ Borel_measurable a
```

Theorem 4.7.4 (Measurability of Function Composition). *If $f : \mathbb{R} \to \mathbb{R}$ and g are measurable, and $h = f \circ g$, then h is also measurable.*

```
Theorem in_measurable_borel_comp_borel :
⊢ ∀a f g h.
f ∈ borel_measurable borel ∧
g ∈ borel_measurable a ∧
(∀x. x ∈ space a ⇒ h x = f (g x)) ⇒
h ∈ borel_measurable a
```

Lemma 4.7.5 (Measurability of Differentiable Functions). *If $f \in C^n(\mathbb{R})$, then all of its derivatives up to order $m \leq n$ are Borel measurable.*

```
Theorem in_borel_measurable_diff :
⊢ ∀f n m.
f ∈ CnR n ∧ m ≤ n ∧ 0 < n ⇒
diff m f ∈ borel_measurable borel
```

Lemma 4.7.6 (Measurability under Pointwise Equality). *If two functions f and g agree pointwise on the space, and g is measurable, then f is also measurable.*

```
Theorem in_measurable_borel_eq :
⊢ ∀a f g.
(∀x. x ∈ space a ⇒ f x = g x) ∧
g ∈ borel_measurable a ⇒
f ∈ borel_measurable a
```

Lemma 4.7.7 (Equivalence of Real-Valued Random Variable Forms). *If p is a probability space and X is a function into $\mathbb{R}$, then the following are equivalent:*

- *X is a real-valued random variable in the sense of ‘random_variable X p borel‘,*
- *the composed function $\mathtt{Normal} \circ X$ is a ‘real_random_variable‘ over p.*

```
Theorem real_random_variable_equiv :
⊢ ∀p X.
prob_space p ⇒
(real_random_variable (Normal ∘ X) p ⇔
random_variable X p borel)
```

4.8 Normal Random Variables over `extreal`

This section develops additional formal support for extended normal random variables in the `extreal` type. While HOL4 provides basic infrastructure for real-valued normal random variables via `normal_rv`, our formalisation of the Central Limit Theorem (CLT) requires reasoning about normal random variables with values in the extended real line. This is necessary to work with expectations and variances that naturally arise as `extreal`-valued quantities in the HOL4 probability library. The following lemmas extend and generalise standard normality properties to the `extreal` setting, enabling operations such as affine transformations, summation, and independence to be handled uniformly.

Lemma 4.8.1 (Equality Preserves Normality)**.** *If two random variables are equal on the probability space, then normality is preserved:*

$$\begin{aligned} &\mathit{prob\_space}(p) \wedge \forall x \in \mathit{p\_space}(p).\ X(x) = Y(x) \Rightarrow \\ &\mathit{normal\_rv}(X, p, \mu, \sigma) \Leftrightarrow \mathit{normal\_rv}(Y, p, \mu, \sigma). \end{aligned} \tag{4.6}$$

```
Theorem normal_rv_cong :
⊢ ∀p X Y μ σ.
prob_space p ∧ (∀x. x ∈ p_space p ⇒ X x = Y x) ⇒
(normal_rv X p μ σ ⇔ normal_rv Y p μ σ)
```

Lemma 4.8.2 (Affine Transformation of Normal Variables)**.** *Let X be an extended normal random variable over p. Then an affine transformation $Y(x) = b + a \cdot X(x)$ is also an extended normal random variable, with appropriately shifted mean and scaled standard deviation:*

$$\mathit{ext\_normal\_rv}(X, p, \mu, \sigma) \Rightarrow \mathit{ext\_normal\_rv}(Y, p, b + a\mu, |a|\sigma).$$

```
Theorem ext_normal_rv_affine :
⊢ ∀X p μ σ Y a b.
prob_space p ∧ a ≠ 0 ∧ 0 < σ ∧
ext_normal_rv X p μ σ ∧
(∀x. Y x = Normal b + Normal a * X x) ⇒
ext_normal_rv Y p (b + a * μ) (abs a * σ)
```

Lemma 4.8.3 (Sum of Two Independent Normal Variables)**.** *The sum of two independent extended normal random variables is also an extended normal variable, with mean equal to the sum of means and variance equal to the sum of variances:*

$$\begin{aligned} &\mathit{indep\_vars}(p,\ X,\ Y,\ \mathit{Borel},\ \mathit{Borel}) \\ &\wedge\ \mathit{ext\_normal\_rv}(X,\ p,\ \mu_1,\ \sigma_1) \wedge \mathit{ext\_normal\_rv}(Y,\ p,\ \mu_2,\ \sigma_2) \\ &\Rightarrow\ \mathit{ext\_normal\_rv}((X+Y),\ p,\ \mu_1+\mu_2,\ \sqrt{\sigma_1^2+\sigma_2^2}). \end{aligned}$$

```
Theorem sum_indep_ext_normal :
⊢ ∀p X Y μ1 μ2 σ1 σ2.
prob_space p ∧ 0 < σ1 ∧ 0 < σ2 ∧
indep_vars p X Y Borel Borel ∧
ext_normal_rv X p μ1 σ1 ∧
ext_normal_rv Y p μ2 σ2 ⇒
ext_normal_rv (λx. X x + Y x) p (μ1 + μ2) (sqrt (σ1{2 + σ2{2))
```

Lemma 4.8.4 (Sum of Finite Independent Normal Variables)**.** *Given a finite sequence of independent extended normal random variables, the sum is also extended normal, with additive mean and variance:*

$$\forall i < n,\ \mathit{ext\_normal\_rv}(X_i,\ p,\ \mu_i,\ \sigma_i),\ \mathit{indep\_vars}(p,\ X,\ (\lambda i.\ \mathit{Borel}),\ \mathit{count}(n))$$

$$\Rightarrow\ \mathit{ext\_normal\_rv}\left(\lambda x. \sum_{i<n} X_i(x),\ p,\ \sum_{i<n} \mu_i,\ \sqrt{\sum_{i<n} \sigma_i^2}\right)$$

```
Theorem sum_indep_ext_normal' :
⊢ ∀p X μ σ n.
prob_space p ∧ 0 < n ∧
(∀i. i < n ⇒ ext_normal_rv (X i) p (μ i) (σ i)) ∧
indep_vars p X (λi. Borel) (count n) ∧
(∀i. i < n ⇒ 0 < σ i) ⇒
ext_normal_rv
(λx. ∑ (λi. X i x) (count n)) p
(∑ (λi. μ i) (count n))
(sqrt (∑ (λi. (σ i){2) (count n)))
```

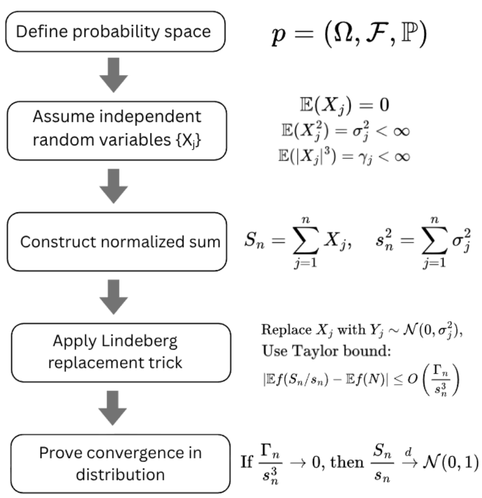
Chapter 5

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:

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

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:

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

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 5.2.1 (Existence of independent variables). *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:

- $X_i' = X_i \circ \mathtt{FST}$, and
- $Y_i' = Y_i \circ \mathtt{SND}$,

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 5.2.2 (Construction of auxiliary sequence). *There exists a product probability space $p = p_1 \times p_2$ supporting two independent sequences of random variables $X'_i = X_i \circ \mathit{FST}$ and $Y'_i = Y_i \circ \mathit{SND}$ such that the interleaved sequence Z_i combines both and preserves mutual independence and variance structure.*

Or, formally:

```
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 5.3.1 (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 5.3.2 (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.

Lemma 5.4.1 (Measurability of Partial Sums). *Given real-valued integrable random variables $X_0, \dots, X_{n-1}$ and $Y_0, \dots, Y_{n-1}$, define:*

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

Then each Z_j is also a real random variable and integrable.

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.

Lemma 5.4.2 (Telescoping Sum Identity). *For any real-valued function f where each term is finite, we have:*

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

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 5.4.3 (Lindeberg Replacement Lemma). *Let X_j, Y_j be independent sequences of real random variables with bounded third moments, and let $f \in C_b^3$. Then the total replacement error between X_j and Y_j when replacing them one by one is bounded as follows...*

In HOL4, it states as:

```
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:

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

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

Theorem 5.5.1 (Global Taylor Error Bound)**.** *Given independent sequences X_j, Y_j, Z_j with bounded third moments and a test function $f \in C_b^3$, the cumulative difference in expectations is bounded by the total third moments scaled by s_n^{-3}.*

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 5.6.1 (Lyapunov Inequality)**.** *For a random variable X with finite r-th and r'-th absolute moments where $0 < r < r'$, we have:*

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

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:

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

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

By Lyapunov's inequality, we know $\sigma_j^3 \leq y_j$. 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 5.3 (Big-O Notation). $f(n) = \mathcal{O}(g(n))$ means that there exists a constant $c > 0$ and threshold n_0 such that for all $n \geq n_0$, we have $|f(n)| \leq c \cdot |g(n)|$.

Or, formally:

```
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 5.6.2 (Big-O algebra). (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.

Theorem 5.6.3 (Central Limit Theorem: Lyapunov Form)**.** *Let $\{X_i\}$ be a sequence of independent, mean-zero, real random variables with variances $Var(X_i)$ and third absolute moments. Under Lyapunov's condition:*

$$\frac{\sum \mathbb{E}[|X_i|^3]}{\left(\sum Var(X_i)\right)^{3/2}} \to 0,$$

we have:

$$\frac{\sum X_i}{\sqrt{\sum Var(X_i)}} \xrightarrow{d} \mathcal{N}(0,1).$$

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

This thesis has formalised a general version of the Central Limit Theorem (CLT) in HOL4 under Lyapunov's condition, using the Lindeberg replacement method. The version presented here is far more general than the classical i.i.d. case: it handles independent (but not identically distributed) summands with varying variances and third absolute moments. Our development closely follows Chung's analytical presentation [3], and surpasses earlier formal work such as the CLT in Isabelle/HOL [11] in both generality and technical depth.

To our knowledge, this is the first mechanised proof of Lyapunov's CLT in HOL4 for general independent sequences, employing the Lindeberg replacement trick, Taylor expansion, and Big-O reasoning. Unlike earlier approaches that assume identical distributions, our formalisation accommodates inhomogeneous sequences, which introduces substantial technical overhead—particularly in bounding variable-specific replacement errors. However, this effort yields a more widely applicable theorem under Lyapunov's condition.

Over 6000 lines of HOL4 code were developed to support this formalisation, including infrastructure for moment inequalities, summability, Taylor approximation, and asymptotic analysis. Much of this was written from scratch or adapted from existing theories to suit the strict requirements of probabilistic reasoning in HOL4. The resulting tools significantly enhance HOL4's capacity to reason about convergence in distribution and limit theorems.

Initial attempts to formalise the CLT via moment-generating functions proved impractical due to the lack of support for improper integrals and MGFs in HOL4's libraries. The switch to the Lindeberg method, though more technically involved, allowed for a successful proof strategy grounded entirely in measure theory and real analysis.

What has been completed:

- Construction of auxiliary sequences of Gaussian random variables;
- Moment and variance control via Lyapunov-type inequalities;
- Taylor expansion bounds on individual replacement steps;

- Big-O asymptotic estimation of the total 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.

Beyond the lack of special-function support, a deeper gap remains between classical Riemann-based probabilistic analysis and the rigorous Lebesgue-style formalism adopted in HOL4. Many standard results rely on transitions between these views, which must be made fully explicit in a mechanised setting. Bridging this divide—by relating improper Riemann and Lebesgue integrals—is essential for validating textbook-level probabilistic computations, and forms an important direction for future work.

At the time of writing, all analytic components have been formalised, and the final error bound has been cleanly isolated. The only missing piece is the formalisation of the third moment of the Gaussian, a purely technical step requiring extended integration support in HOL4.

This work lays a solid foundation for future research on formalising limit theorems in probability theory. As support for complex integration and special functions in HOL4 grows, the full proof of the CLT will become attainable. Beyond that, the techniques developed here could be extended to CLTs for dependent sequences, martingales, or even random vectors. This thesis not only demonstrates the feasibility of formalising one of probability theory's central results but also provides reusable infrastructure for reasoning about expectations, variances, Taylor approximations, and convergence in distribution within a theorem prover.

Chapter 7

Future Work

The formalisation of Lyapunov's Central Limit Theorem (CLT) in this thesis is nearly complete. All analytic components—including moment bounds, Taylor expansion, auxiliary sequences, and asymptotic estimates—have been mechanised. The final convergence theorem has been stated and is fully supported by the developed infrastructure. The only remaining step is to formalise the third absolute moment of the Gaussian distribution:

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

This requires formalising improper integrals and special functions, particularly the gamma function. Currently, HOL4 lacks sufficient support for such techniques. Bridging the gap between classical Riemann-style and formal Lebesgue integration will be critical for validating these results. The author intends to address this in future work.

Beyond completing this proof, several directions arise naturally:

1. Generalising the Central Limit Theorem

The current version handles independent, non-identically distributed variables. Future formal developments could include:

- **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 Mathematical Libraries

The formalisation also highlights areas where HOL4 could be extended to support more advanced probability theory:

- Definitions and properties of the gamma, beta, and error functions.
- Support for improper Lebesgue integration on infinite domains.
- Formalisation of complex-valued functions and Fourier/characteristic methods.
- Improved automation for Big-O and asymptotic reasoning in formal proofs.

These enhancements would enable broader mechanisation of probability theory, from classical results to modern topics in stochastic processes and statistics. The infrastructure developed in this thesis demonstrates the feasibility of such mechanisation and provides a reusable foundation for further research.

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.