On the Formalization of the Lebesgue Integration Theory in HOL

Tarek Mhamdi, Osman Hasan, and Sofiène Tahar

ECE Department, Concordia University, Montreal, QC, Canada
{mhamdi,o_hasan,tahar}@ece.concordia.ca

Technical Report

January, 2010

Abstract

Lebesgue integration is a fundamental concept in many mathematical theories, such as real analysis, probability and information theories. Reported higher-order-logic formalizations of the Lebesgue integral either do not include, or have a limited support for the Borel algebra, which is the canonical sigma algebra used on any metric space over which the Lebesgue integral is defined. In this report, we overcome this limitation by presenting a formalization of the Borel sigma algebra that can be used on any metric space, such as the complex numbers or the n-dimensional Euclidean space. Building on top of this framework, we have been able to prove some key Lebesgue integral properties, like its linearity and monotone convergence. Furthermore, we present the formalization of the "almost everywhere" relation and prove that the Lebesgue integral does not distinguish between functions which differ on a null set as well as other important results based on this concept. As applications, we present the verification of Markov and Chebyshev inequalities and the Weak Law of Large Numbers theorem.

Table of Contents

1 Introduction

Formal modeling of physical systems or devices is not a very straightforward task due to the presence of many continuous and unpredictable components. For example, embedded systems are operating in a concrete physical environment with continuous dynamics; cryptography heavily relies upon information theoretic concepts; a broad area of chemistry and biology (and biophysics) worries about stochastic effects and phenomena, etc. Formal models of computation have in the past mostly been considered independent of the continuous or unpredictable world. In classical formal verification efforts, hardware and software are viewed as discrete models of computation. But due to the dire need of accurate analysis in safety-critical domains, there is a growing trend towards incorporating continuous and unpredictable physical realities in the formal models of physical systems.

Lebesgue integration [1] is a fundamental concept in many mathematical theories, such as real analysis [6], probability [7] and information, which are widely used to model and reason about the continuous and unpredictable components of physical systems. The reasons for its extensive usage, compared to the commonly known Riemann integral, include the ability to handle a broader class of functions, which are defined over more general types than the real line, and its better behavior when it comes to interchanging limits and integrals. In order to facilitate the formal analysis of physical systems, two higher-order-logic formalizations of the Lebesgue integral have been recently reported [3, 17]. However, they either do not include, or have a very limited support for the Borel algebra [2], which is a sigma algebra generated by the open sets. These deficiencies restrict the formal reasoning about some very useful Lebesgue integral properties, which in turn limits the scope of formally analyzing physical systems.

In this report, we present a generalized formalization of the Lebesgue integral in order to exploit its full potential for the formal analysis of other systems. We first formalize the Borel algebra that provides a unified framework to prove the Lebesgue integral properties and measurability theorems on any metric space, such as the real numbers, the complex numbers or the n-dimensional Euclidean space. Building on top of this formalization, we prove some of the key Lebesgue integral properties as well as its convergence theorems. Similarly, we formalize the notion of "almost everywhere" [1] and prove that the Lebesgue integral does not distinguish between functions which differ on a null set along with some other useful results based on the "almost everywhere" relation. In order to illustrate the practical effectiveness of our work, we utilize it to verify the Chebyshev and Markov inequalities and the Weak Law of Large Numbers (WLLN) [16], which are widely used properties in probability and information theories.

We used the HOL theorem prover for the above mentioned formalization and verification tasks. The main motivation behind this choice was to build upon existing formalizations of measure [11] and Lebesgue integration [3] theories.

The rest of the report is organized as follows: Section 2 provides a review of related

work. Some measure theory and Lebesgue integration preliminaries are given in Section 3. In Section 4, we give an overview of main definitions of the measure theory [2]. Section 4.2 presents our formalization of the Borel theory, which is used in Section 5 to prove the main properties of the Lebesgue integral and its convergence theorems. In Section 6, we use our formalization for verifying some important theorems from the theory of probability. Finally, Section 7 concludes the report and provides hints to future work.

2 Related Work

Coble [3] generalized the measure theory formalization by Hurd [11] and built on it to formalize the Lebesgue integration theory. He proved some properties of the Lebesgue integral but only for the class of positive simple functions. Besides, multiple theorems in Coble's work have the assumption that every set is measurable which is not correct in most cases of interest. We propose to prove the Lebesgue integral properties and convergence theorems for arbitrary functions by providing a formalization of the Borel sigma algebra, which has also been used to overcome the assumption of Cobles's work.

Based on the work of Hurd [11], Richter [17] also formalized the measure theory in Isabelle/HOL, where he restricts the measure spaces that can be constructed. In Richter's formalization, a measure space is the pair $(\mathcal{A}, \mu)$; $\mathcal{A}$ is a set of subsets of X, called the set of measurable sets and μ is a measure function. The space is implicitly the universal set of the appropriate type. This approach does not allow to construct a measure space where the space is not the universal set. The only way to apply this approach for an arbitrary space X is to define a new type for the elements of X, redefine operations on this set and prove properties of these operations. This requires considerable effort that needs to be done for every space of interest. The work we propose in this report is based on the formalization of Coble [3] where we define a measure space as a triplet $(X, \mathcal{A}, \mu)$; the set X being the space.

Richter [17] defined the Borel sets as being generated by the intervals. In the formalization we propose in this report, the Borel sigma algebra is generated by the open sets and is more general as it can be applied not only to the real numbers but to any metric space such as the complex numbers or $\mathbb{R}^n$, the n-dimensional Euclidean space. It provides a unified framework to prove the measurability theorems in these spaces. Besides, our formalization allows us to prove that any continuous function is measurable which is an important result to prove the measurability of a large class of functions, in particular, trigonometric and exponential functions. To prove this result we also formalize in this report key concepts of topology [15] in HOL.

In his work in topology in the PVS theorem prover, Lester [12] provided formalizations for measure and integration theories but did not prove the properties of the Lebesgue integral nor its convergence theorems such as the Lebesgue Monotone Convergence.

3 Preliminaries

3.1 Measure Theory

A measure is a way to assign a number to a set, interpreted as its size, a generalization of the concepts of length, area, volume, etc. Two important examples are the Lebesgue measure on a Euclidean space and the probability measure on a Borel space. The former assigns the conventional length, area and volume of Euclidean geometry to suitable subsets of $\mathbb{R}^n, n = 1, 2, 3$ and the latter assigns a probability to an event and satisfies the condition that the measure of the sample space is equal to 1.

A measure function is defined as a function that assigns a non-negative real number to all the sets over which it is defined. It must also satisfy the countable additivity condition. This condition states that the measure of the union of a collection of disjoint sets is equal to the sum of their measures.

After the discovery of paradoxes in the naive set theory, various axiomatic systems were proposed, the best known of which is the Zermelo-Fraenkel set theory [5] with the famous Axiom of Choice (ZFC). This set theory is the most common foundation of mathematics down to the present day. The Axiom of Choice, however, implies the existence of counter-intuitive sets and gives rise to paradoxes of its own, in particular, the Banach-Tarski paradox [18]. The solution to this is to tag some sets as non-measurable and to define the measure only on a class of subsets called the measurable sets.

As a consequence of the countable additivity condition, the measurable sets are required to form a sigma-algebra, meaning that unions, intersections and complements of sequences of measurable sets are measurable.

Definition 1. *(Sigma Algebra)*
Let $\mathcal{A}$ be a collection of subsets (or subset class) of a space X. $\mathcal{A}$ defines a sigma algebra on X iff *$\mathcal{A}$ contains the empty set $\emptyset$, and is closed under countable unions and complementation within the space X.*

The smallest sigma algebra on a space X is $\mathcal{A} = \{\emptyset, X\}$ and the largest is its powerset, $\mathcal{P}(X)$, the set of all subsets of X. The pair $(X, \mathcal{A})$ is called a σ-field or a measurable space, $\mathcal{A}$ is the set of measurable sets.

For any collection G of subsets of X we can construct $\sigma(X, G)$, the smallest sigma algebra on X containing G. $\sigma(X, G)$ is called the sigma algebra on X generated by G. There is at least one sigma algebra on X containing G, namely the power set of X. $\sigma(X, G)$ is the intersection of all those sigma algebras.

Definition 2. *(Measure Space)*
A triplet $(X, \mathcal{A}, \mu)$ is a measure space iff *$(X, \mathcal{A})$ is a measurable space and $\mu : \mathcal{A} \to \mathbb{R}$ is a non-negative and countably additive measure function.*

A probability space $(\Omega, \mathcal{A}, p)$ is a measure space satisfying $p(\Omega) = 1$.

There is a special class of functions, called measurable functions, that are structure preserving, in the sense that the inverse image of each measurable set is also measurable. This is analogous to continuous functions in metric spaces where the inverse image of an open set is open.

Definition 3. *(Measurable Functions)*
Let $(X_1, \mathcal{A}_1)$ and $(X_2, \mathcal{A}_2)$ be two measurable spaces. A function $f : X_1 \to X_2$ is called measurable with respect to $(\mathcal{A}_1, \mathcal{A}_2)$ (or $(\mathcal{A}_1, \mathcal{A}_2)$ measurable) iff *$f^{-1}(A) \in \mathcal{A}_1$ for all $A \in \mathcal{A}_2$.*

In this definition, we did not specify any structure on the measurable spaces. If we consider a function f that takes its values on a metric space, most commonly the set of real numbers or complex numbers, then the Borel sigma algebra on that space is used.

Definition 4. *(Borel Sigma Algebra)*
The Borel sigma algebra on a space X is the smallest sigma algebra generated by the open sets of X.

An important example, especially in the theory of probability, is the Borel sigma algebra on $\mathbb{R}$, denoted by $\mathcal{B}(\mathbb{R})$.

3.2 Lebesgue Integration

Lebesgue integration [1] is a fundamental concept in many mathematical theories, such as real analysis [6], probability [7] and information, which are widely used to model and reason about the continuous and unpredictable components of physical systems. The reasons for its extensive usage, compared to the commonly known Riemann integral, include the ability to handle a broader class of functions, which are defined over more general types than the real line, and its better behavior when it comes to interchanging limits and integrals, which is of prime importance, for instance, in the study of Fourier series. Similar to the way in which step functions are used in the development of the Riemann integral, the Lebesgue integral makes use of a special class of functions called positive simple functions. They are measurable functions taking finitely many values. In other words, a positive simple function g is represented by the triple (s, a, α) as a finite linear combination of indicator functions of measurable sets (a_i) that form a partition of the space X.

$$\forall x \in X,\ g(x) = \sum_{i \in s} \alpha_i I_{a_i}(x) \quad c_i \geq 0 \tag{1}$$

The Lebesgue integral is first defined for those functions then extended to non-negative functions and finally to arbitrary functions.

Definition 5. *Let $(X, \mathcal{A}, \mu)$ be a measure space. The integral of the positive simple function g with respect to the measure μ is defined as*

$$\int_X g \, d\mu = \sum_{i \in s} \alpha_i \mu(a_i) \tag{2}$$

While the choice of $((\alpha_i),(a_i),s)$ to represent g is not unique, the integral as defined above is independent of that choice.

Definition 6. *Let $(X,\mathcal{A},\mu)$ be a measure space. The integral of a non-negative measurable function f is defined as*

$$\int_X f\,d\mu = \sup\{\int_X g\,d\mu \mid g \leq f \quad \textit{and } g \textit{ positive simple function}\} \tag{3}$$

Finally, the integral for arbitrary measurable functions is given in the following definition.

Definition 7. *Let $(X,\mathcal{A},\mu)$ be a measure space. The integral of an arbitrary measurable function f is defined as*

$$\int_X f\,d\mu = \int_X f^+\,d\mu - \int_X f^-\,d\mu \tag{4}$$

where f^+ and f^- are the non-negative functions defined by $f^+(x) = \max(f(x),0)$ and $f^-(x) = \max(-f(x),0)$.

We focus on Lebesgue integrable functions for which the integral exists and is finite.

Definition 8. *(Integrable Functions)*
Let $(X,\mathcal{A},\mu)$ be a measure space, a measurable function f is integrable iff $\int_X |f|\,d\mu < \infty$ or equivalently iff $\int_X f^+\,d\mu < \infty$ and $\int_X f^-\,d\mu < \infty$

The Lebesgue integral shares some key properties with the Reimann integral, such as, the linearity and monotonicity properties. Let f and g be integrable functions and $c \in \mathbb{R}$ then
- $\forall\, x,\ 0 \leq f(x) \Rightarrow 0 \leq \int_X f\,d\mu$
- $\forall\, x,\ f(x) \leq g(x) \Rightarrow \int_X f\,d\mu \leq \int_X g\,d\mu$
- $\int_X cf\,d\mu = c\int_X f\,d\mu$
- $\int_X f+g\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu$
- A and B disjoint sets $\Rightarrow \int_{A\cup B} f\,d\mu = \int_A f\,d\mu + \int_B f\,d\mu$

4 Measure Theory

Parts of the measure theory were formalized in [11] and [3]. We make use of these formalizations in our development and extend it by formalizing the Borel sigma algebra and Borel measurable functions. This will allow us to define and manipulate random variables defined on any topological space.

4.1 Basic Definitions

A subset class is collection of subsets of a space X and is formalized as

```
⊢ ∀X A. subset_class X A = ∀s. s ∈ A ⇒ s ⊆ X
```

A set S is countable if its elements can be counted one at a time, or in other words, if every element of the set can be associated with a natural number, i.e., there exists a surjective function $f : \mathbb{N} \to S$.

```
⊢ ∀s. countable s = ∃f. ∀x. x ∈ s ⇒ ∃n. f n = x
```

A subset class of a space X defines a sigma algebra on X *iff* it contains the empty set $\emptyset$, and is closed under countable unions and complementation within the space X.

```
⊢ ∀X A. sigma_algebra (X,A) =
    subset_class X A ∧ {} ∈ A ∧
    (∀s. s ∈ A ⇒ X\s ∈ A) ∧
    ∀c. countable c ∧ c ⊆ A ⇒ ⋃c ∈ A
```

where $X\backslash s$ denotes the complement of s within X, $\bigcup c$ the union of all elements of c. We define the `space` and `subsets` functions such that

```
⊢ ∀X A. space (X,A) = X
⊢ ∀X A. subsets (X,A) = A
```

The sigma algebra on X generated by a collection of subsets G is formalized in HOL as

```
⊢ ∀X G. sigma X G = (X,⋂{s | G ⊆ s ∧ sigma_algebra (X,s)})
```

where $\bigcap c$ denotes the intersection of all elements of c.
A triplet $(X, \mathcal{A}, \mu)$ is a measure space *iff* $(X, \mathcal{A})$ is a measurable space and $\mu : \mathcal{A} \to \mathbb{R}$ is a non-negative and countably additive measure function.

```
⊢ ∀X A mu. measure_space (X,A,mu) =
    sigma_algebra (X,A) ∧ positive (X,A,mu) ∧
    countably_additive (X,A,mu)
```

A measure function is countably additive when the measure of a countable union of pairwise disjoint measurable sets is the sum of their respective measures.

```
⊢ ∀X A mu. countably_additive (X,A,mu) =
    ∀f. f ∈ (UNIV → A) ∧
      (∀m n. m ≠ n ⇒ DISJOINT (f m) (f n)) ∧
      ⋃ (IMAGE f UNIV) ∈ A ⇒
      mu o f sums mu (⋃(IMAGE f UNIV))
```

Similarly, we define the functions `m_space`, `measurable_sets` and `measure` such that

```
⊢ ∀X A mu. m_space (X,A,mu) = X
⊢ ∀X A mu. measurable_sets (X,A,mu) = A
⊢ ∀X A mu. measure (X,A,mu) = mu
```

Measurable functions satisfy the condition that the inverse image of a measurable set is also measurable. The HOL formalization is the following.

```
⊢ ∀a b f. f ∈ measurable a b =
    sigma_algebra a ∧ sigma_algebra b ∧ f ∈ (space a → space b) ∧
    ∀s. s ∈ subsets b ⇒ PREIMAGE f s ∩ space a ∈ subsets a
```

Notice that unlike Definition 3, the inverse image in the formalization (`PREIMAGE f s`) needs to be intersected with `space a` because the functions in HOL are total, meaning that they map every value of a certain HOL type (even those outside `space a`) to a value of an appropriate type which may or may not be in `space b`. In other words, writing in HOL that f is a function from `space a` to `space b` (`f ∈ (space a -> space b)`), does not exclude values outside `space a` and hence the intersection is needed.

In this definition, we did not specify any structure on the measurable spaces. If we consider a function f that takes its values on a metric space, most commonly the set of real numbers or complex numbers, then the Borel sigma algebra on that space is used. In the following, we present our formalization of the Borel sigma algebra in HOL.

4.2 Borel Sigma Algebra

Working with the Borel sigma algebra makes the set of measurable functions a vector space. It also allows us to prove various properties of the measurable functions necessary for the formalization of the Lebesgue integral and its properties in HOL.

To formalize as well as prove in HOL various properties of the Borel sigma algebra on $\mathbb{R}$, we need to formalize some topology concepts of $\mathbb{R}$ and also provide a formalization of the set of rational numbers $\mathbb{Q}$. A theory for the rational numbers was developed in HOL but does not include the theorems that we need and is in fact unusable for our development because we need to work on rational numbers as a subset of the real numbers and not of a different HOL type. Some concepts of the topology of $\mathbb{R}$ were formalized in HOL by Harrison [8] but his formalization does not use the set theory and also lacks some of the important theorems that we need in our development. Harrison, later, developed an extensive topology theory [9] in HOL-Light. Details of our formalization of the topology of $\mathbb{R}$ as well as a theory of rational numbers can be found in [14].
The Borel sigma algebra on a space X is the smallest sigma algebra generated by the open sets of X.

```
⊢ borel X = sigma X (open_sets X)
```

An important example, especially in the theory of probability, is the Borel sigma algebra on $\mathbb{R}$, denoted by $\mathcal{B}(\mathbb{R})$ which we simply call *Borel* in the sequel.

```
⊢ Borel = sigma UNIV (open_sets UNIV)
```

where `UNIV` is the universal set of real numbers $\mathbb{R}$. Details about our formalization of the open sets and other aspects of the topology of the real line can be found in [14].
$\mathcal{B}(\mathbb{R})$ is, by definition, generated by the open sets of $\mathbb{R}$. In the following theorem we prove that it is also generated by the open intervals ((c, d) for $c, d \in \mathbb{R}$). This was

actually used in many textbooks as a starting definition for the Borel sigma algebra on $\mathbb{R}$. While we will prove that the two definitions are equivalent in the case of the real line, our formalization is vastly more general and can be used for any metric space such as the complex numbers or $\mathbb{R}^n$, the n-dimensional Euclidian space. We show that $\mathcal{B}(\mathbb{R})$ is also generated by any of the following classes of intervals: $(-\infty, c)$, $[c, +\infty)$, $(c, +\infty)$, $(-\infty, c]$, $[c, d)$, $(c, d]$, $[c, d]$, where $c, d \in \mathbb{R}$.

Theorem 1. $\mathcal{B}(\mathbb{R})$ *is generated by the open intervals* (c, d) *where* $c, d \in \mathbb{R}$

```
⊢ Borel = sigma UNIV (open_intervals_set)
```

where the open intervals set is formalized as

```
⊢ open_intervals_set =
      {{x | a < x ∧ x < b} | a ∈ UNIV ∧ b ∈ UNIV}
```

Proof. The sigma algebra generated by the open intervals, σ_I, is by definition the intersection of all sigma algebras containing the open intervals. $\mathcal{B}(\mathbb{R})$ is one of them because the open intervals are open sets (Theorem 20). Hence, $\sigma_I \subset \mathcal{B}(\mathbb{R})$. Now let $A \in \mathcal{B}(\mathbb{R})$, then A is an open set and since every open set on the real line is the union of a countable set of open intervals (Theorem 22), $A \in \sigma_I$. Consequently $\mathcal{B}(\mathbb{R}) = \sigma_I$. To prove that $\mathcal{B}(\mathbb{R})$ is also generated by the other classes of intervals, it suffices to prove that any interval]a,b[is contained in the sigma algebra corresponding to each class. For the case of the intervals of type $[c, d[$, this follows from the following equation:

$$]a, b[= \bigcup_n [a + \frac{1}{2^n}, b[\tag{5}$$

For the open rays $] - \infty, c\ [$, the result follows from the fact that $[a, b[$ can be written as the difference of two rays, $[a, b[=] - \infty, b\ [\setminus] - \infty, a\ [$.
In a similar manner, we prove in HOL that all mentioned classes of intervals generate the Borel sigma algebra on $\mathbb{R}$.

Another useful result, asserts that the singleton sets are measurable sets of $\mathcal{B}(\mathbb{R})$.

Theorem 2. $\forall c \in \mathbb{R}, \quad \{c\} \in \mathcal{B}(\mathbb{R})$

```
⊢ ∀c. {c} ∈ subsets Borel
```

The proof of this theorem follows from the fact that a sigma algebra is closed under countable intersection and the equation

$$\forall c \in \mathbb{R} \quad \{c\} = \bigcap_n [c - \frac{1}{2^n}, c + \frac{1}{2^n}[\tag{6}$$

4.3 Real-Valued Measurable Functions

Recall that in order to check if a function f is measurable with respect to $(\mathcal{A}_1, \mathcal{A}_2)$, it is necessary to check that for any $A \in \mathcal{A}_2$, its inverse image $f^{-1}(A) \in \mathcal{A}_1$. The following theorem states that, for real-valued functions, it suffices to perform the check on the open rays $((-\infty, c),\ c \in \mathbb{R})$.

Theorem 3. *Let $(X, \mathcal{A})$ be a measurable space. A function $f : X \to \mathbb{R}$ is measurable with respect to $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ iff $\forall c \in \mathbb{R},\ f^{-1}((-\infty, c)) \in \mathcal{A}$*

```
⊢ ∀f a.
    f ∈ measurable a Borel =
    sigma_algebra a ∧ f ∈ (space a → UNIV) ∧
    ∀c. {x | f x < c} ∩ space a ∈ subsets a
```

Proof. Suppose that f is measurable with respect to $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$, we showed in the previous section that $\forall c \in \mathbb{R}$, $]-\infty, c[\in \mathcal{B}(\mathbb{R})$. Since f is measurable then $f^{-1}(]-\infty, c[) \in \mathcal{A}$. Now suppose that $\forall c \in \mathbb{R}$, $f^{-1}(]-\infty, c[) \in \mathcal{A}$, we need to prove $\forall A \in \mathcal{B}(\mathbb{R})$, $f^{-1}(A) \in \mathcal{A}$. This follows from Theorem 22 stating that A is a countable union of open intervals and the equalities $f^{-1}(\bigcup_{n\in\mathbb{N}} A_n) = \bigcup_{n\in\mathbb{N}} f^{-1}(A_n)$ and $f^{-1}(]-\infty, c[) = \bigcup_{n\in\mathbb{N}} f^{-1}(]-n, c[)$

In a similar manner, we prove in HOL that f is measurable with respect to $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ iff $\forall\, c, d \in \mathbb{R}$ the inverse image of any of the following classes of intervals is an element of $\mathcal{A}$: $]-\infty, c[$, $[c, +\infty[$, $]c, +\infty[$, $]-\infty, c]$, $[c, d[$, $]c, d]$, $[c, d]$.

Every constant real function on a space X is measurable. The indicator function on a set A is measurable iff A is measurable.
In the following, we prove in HOL various properties of the real-valued measurable functions.

Theorem 4. *If f and g are $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ measurable and $c \in \mathbb{R}$ then cf, $|f|$, f^n, $f + g$, $f * g$ and $max(f, g)$ are $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ measurable.*

```
⊢ ∀a f g h c. sigma_algebra a ∧ f ∈ measurable a Borel ∧
      g ∈ measurable a Borel ⇒
      ((\x. c * f x) ∈ measurable a Borel) ∧
      ((\x. abs(f x)) ∈ measurable a Borel) ∧
      ((\x. f x pow n) ∈ measurable a Borel) ∧
      ((\x. f x + g x) ∈ measurable a Borel) ∧
      ((\x. f x * g x) ∈ measurable a Borel) ∧
      ((\x. max (f x) (g x)) ∈ measurable a Borel)
```

The notation `(\x. f x)` is the lambda notation of f, used to represent the function $f : x \mapsto f(x)$.

Theorem 5. *If (f_n) is a monotonically increasing sequence of real-valued measurable functions with respect to $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$, such that $\forall n, x,\ f_n(x) \to f(x)$ then f is also $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ measurable.*

```
⊢ ∀a f fi. sigma_algebra a ∧ (∀i. fi i  ∈ measurable a Borel) ∧
      (∀x. mono_increasing (\i. fi i x)) ∧
      (∀x. x ∈ m_space m ⇒ (\i. fi i x) ⟶ f x) ⇒
      f ∈ measurable a Borel
```

Theorem 6. *Every continuous function $g : \mathbb{R} \to \mathbb{R}$ is $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$ measurable.*

```
⊢ ∀g. (∀x. g contl x) ⇒ g ∈ measurable Borel Borel
```

Theorem 7. *If $g : \mathbb{R} \to \mathbb{R}$ is continuous and f is $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ measurable then $g \circ f$ is also $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ measurable.*

```
⊢ ∀a f g. sigma_algebra a ∧ f ∈ measurable a Borel ∧
      (∀x. g contl x) ⇒ g o f ∈ measurable a Borel
```

Theorem 6 is a direct result of Theorem 23 stating that the inverse image of an open set by a continuous function is open. Theorem 7 guarantees, for instance, that if f is measurable then $exp(f)$, $Log(f)$, $cos(f)$ are measurable. This is derived using Theorem 6 and the equality $(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A))$. We now show how to prove that the sum of two measurable functions is measurable.

Proof. We need to prove that for any $c \in \mathbb{R}$, $(f+g)^{-1}]-\infty, c[$ is a measurable set. One way to solve this is to write it as a countable union of measurable sets. By definition of the inverse image, $(f+g)^{-1}]-\infty, c[= \{x : f(x)+g(x) < c\} = \{x : f(x) < c - g(x)\}$. Using Theorem 18 we prove that it is equal to $\bigcup_{r\in\mathbb{Q}}\{x : f(x) < r \text{ and } r < c - g(x)\}$. We deduce that $(f+g)^{-1}]-\infty, c[= \bigcup_{r\in\mathbb{Q}} f^{-1}]-\infty, r[\cap g^{-1}]-\infty, c-r[$. The right hand side is a countable union of measurable sets because $\mathbb{Q}$ is countable and f and g are measurable functions.

5 Lebesgue Integration

In this section we present the formalization of the Lebesgue integral and prove its main properties as well as the Lebesgue monotone convergence theorem.

5.1 Lebesgue Integral

The Lebesgue integral is first defined for positive simple functions, then extended to non-negative functions and finally to arbitrary functions. Let $(X, \mathcal{A}, \mu)$ be a measure space and g a positive simple function with respect to the measure μ that is represented by `(s,a,x)`. The integral of g is formalized as follows.

```
⊢ ∀m s a x. pos_simple_fn_integral m s a x =
      SIGMA (\i. x i * measure m (a i)) s
```

Next, the Lebesgue integral of a non-negative measurable function f is formalized as

```
⊢ ∀m f. pos_fn_integral m f =
      sup {r | ∃g. r ∈ psfis m g ∧ ∀x. g x ≤ f x}
```

where `r ∈ psfis m g` is equivalent to `r = pos_simple_fn_integral m s a x` and g is a positive simple function represented by `(s,a,x)`.
Finally, the integral for an arbitrary measurable function is formalized in terms of the integrals of f^+ and f^- where f^+ and f^- are the non-negative functions defined by $f^+(x) = \max(f(x), 0)$ and $f^-(x) = \max(-f(x), 0)$.

```
⊢ ∀m f. fn_integral m f =
      pos_fn_integral m (\x. if 0 < f x then f x else 0) -
      pos_fn_integral m (\x. if f x < 0 then -f x else 0)
```

Various properties of the Lebesgue integral for positive simple functions have been proven in HOL [3]. We mention in particular that the above integral is well-defined and independent of the choice of $(\alpha_i), (a_i), s$. Other properties include the linearity and monotonicity of the integral for positive simple functions. Another theorem that was widely used in [3] has however a serious constraint, as was discussed in the related work, where the author had to assume that every subset of the space X is measurable which is equivalent to assuming that every function defined on that space is measurable.

Utilizing our formalization of the Borel sigma algebra and functions measurable with respect to it, we have been able to prove that the functions used in the theorem are in fact measurable without having to assume that every function is measurable. For example we prove that a positive simple function is a measurable function as a linear combination of indicator functions on measurable sets. We also use Theorem 1 to prove that the sets used in the theorem are in fact measurable sets. The new theorem can be stated as

Theorem 8. *Let $(X, \mathcal{A}, \mu)$ be a measure space, f a non-negative function measurable with respect to $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ and (f_n) a monotonically increasing sequence of positive simple functions, pointwise convergent to f such that $\forall\, n, x,\ f_n(x) \leq f(x)$ then $\int_X f\, d\mu = lim_{n\to\infty} \int_X f_n\, d\mu$.*

```
⊢ ∀m f fi ri r. measure_space m ∧
        f ∈ measurable (m_space m,measurable_sets m) Borel ∧
        (∀x. mono_increasing (\i. fi i x)) ∧
        (∀x. x ∈ m_space m ⇒ (\i. fi i x) ⟶ f x) ∧
        (∀i. ri i ∈ psfis m (fi i)) ∧ ri ⟶ r ∧
        (∀i x. fi i x <= f x) ⇒
        (pos_fn_integral m f = r)
```

where the notation $x_n \longrightarrow x$ means that the sequence x_n converges to x.

5.2 Integrability

In this section, we provide the criteria of integrability of a measurable function and prove the integrability theorem which will play an important role in proving the properties of the Lebesgue integral.

Definition 9. *(Integrable Functions)*
Let $(X, \mathcal{A}, \mu)$ be a measure space, a measurable function f is integrable iff $\int_X |f|\, d\mu < \infty$ or equivalently iff $\int_X f^+\, d\mu < \infty$ and $\int_X f^-\, d\mu < \infty$

```
⊢ ∀m f. integrable m f =
  f ∈ measurable (m_space m,measurable_sets m) Borel ∧
  (∃z. pos_fn_integral m (\x. if 0 < f x then f x else 0) ≤ z) ∧
  (∃z. pos_fn_integral m (\x. if f x < 0 then -f x else 0) ≤ z)
```

Next, we prove the integrability theorem which allows us to prove the properties of the Lebesgue integral for arbitrary functions as an extension of the properties for positive simple functions. This theorem provides also an alternative definition of the Lebesgue integral.

Theorem 9. *For any non-negative integrable function f there exists a sequence of positive simple functions (f_n) such that $\forall\, n, x,\ f_n(x) \le f_{n+1}(x) \le f(x)$ and $\forall\, x,\ f_n(x) \to f(x)$. Besides*

$$\int_X f\, d\mu = \lim_n \int_X f_n\, d\mu$$

For arbitrary integrable functions, the theorem is applied to f^+ and f^- and results in a well-defined integral, given by

$$\int_X f\, d\mu = \lim_n \int_X f_n^+\, d\mu - \lim_n \int_X f_n^-\, d\mu$$

```
⊢ ∀m f. measure_space m ∧ integrable m f ⇒
    (∃fi ri r. (∀x. mono_increasing (\i. fi i x)) ∧
      (∀x. x ∈ m_space m ⇒ (\i. fi i x) ⟶ fn_plus f x) ∧
      (∀i. ri i ∈ psfis m (fi i)) ∧ ri ⟶ r ∧
      (∀i x. fi i x <= fn_plus f x) ∧
      (pos_fn_integral m (fn_plus f) = r)) ∧
    ∃gi vi v. (∀x. mono_increasing (\i. gi i x)) ∧
      (∀x. x ∈ m_space m ⇒ (\i. gi i x) ⟶ fn_minus f x) ∧
      (∀i. vi i ∈ psfis m (gi i)) ∧ vi ⟶ v ∧
      (∀i x. gi i x <= fn_minus f x) ∧
      (pos_fn_integral m (fn_minus f) = v)
```

we show that the sequence (f_n) satisfies the conditions of the theorem and use Theorem 8 to conclude that $\int_X f\, d\mu = \lim_n \int_X f_n\, d\mu$.

$$f_n(x) = \sum_{k=0}^{4^n-1} \frac{k}{2^n} I_{\{x:\frac{k}{2^n} \le f(x) < \frac{k+1}{2^n}\}} + 2^n I_{\{x:2^n \le f(x)\}} \tag{7}$$

First, we use the definition of (f_n) to prove in HOL the following lemmas

Lemma 1. $\forall\, n, x,\ f(x) \ge 2^n \Rightarrow f_n(x) = 2^n$

Lemma 2. $\forall\, n, x,\ and\ k < 4^n,\ \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n} \Rightarrow f_n(x) = \frac{k}{2^n}$

Lemma 3. $\forall\, x,\ (f(x) \geq 2^n) \vee (\exists k,\ k < 4^n\ and\ \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n})$

Next, we prove that the sequence is pointwise convergent to f, upper bounded by f and monotonically increasing.

Convergence $\forall\, x,\ f_n(x) \to f(x)$
$\forall\, x,\ \exists N$ such that $f(x) < 2^N$. Then $\forall\, n \geq N,\ f(x) < 2^n$. Using Lemma 3, $\forall\, n \geq N$, there exists a $k < 4^n$ such that $\frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}$. Then using Lemma 2, $\forall\, n \geq N,\ f_n(x) = \frac{k}{2^n}$. Consequently, $\forall\, n \geq N,\ f_n(x) \leq f(x) < f_n(x) + \frac{1}{2^n}$ and $|f_n(x) - f(x)| < \frac{1}{2^n}$.

Upper Bound $\forall\, n, x,\ f_n(x) \leq f(x)$
if $f(x) \geq 2^n$ then by Lemma 1 $f_n(x) = 2^n$. Hence $f_n(x) \leq f(x)$
if $f(x) < 2^n$ then by Lemma 3 there exists a $k < 4^n$ such that $\frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}$ and by Lemma 2 $f_n(x) = \frac{k}{2^n}$. Hence $f_n(x) \leq f(x)$.

Monotonicity $\forall\, n, x,\ f_n(x) \leq f_{n+1}(x)$
If $f(x) \geq 2^{n+1}$ then $f_n(x) = 2^n$ and $f_{n+1}(x) = 2^{n+1}$. Hence $f_n(x) \leq f_{n+1}(x)$.
if $f(x) < 2^{n+1}$ then using Lemma 3, there exists a $k < 4^{n+1}$ such that $\frac{k}{2^{n+1}} \leq f(x) < \frac{k+1}{2^{n+1}}$ and using Lemma 2, $f_{n+1}(x) = \frac{k}{2^{n+1}}$. Now we need to determine $f_n(x)$ and compare it to $f_{n+1}(x)$.

$$\frac{k}{2^{n+1}} \leq f(x) < \frac{k+1}{2^{n+1}} \ \Rightarrow\ \frac{\frac{k}{2}}{2^n} \leq f(x) < \frac{\frac{k+1}{2}}{2^n}$$

- if k is even and $\frac{k}{2} < 4^n$ then $f_n(x) = \frac{k}{2^{n+1}} = f_{n+1}(x)$
- if k is even and $\frac{k}{2} \geq 4^n$ then $f_n(x) = 2^n$ and $f_n(x) \leq f_{n+1}(x)$
- if k is odd and $\frac{k-1}{2} < 4^n$ then $f_n(x) = \frac{k-1}{2^{n+1}} \leq f_{n+1}(x)$
- if k is odd and $\frac{k-1}{2} \geq 4^n$ then $f_n(x) = 2^n$ and $f_n(x) \leq f_{n+1}(x)$

5.3 Lebesgue Integral Properties

We formally verified in the HOL theorem prover some key properties of the Lebesgue integral, such as the monotonicity and linearity. Let f and g be integrable functions and $c \in \mathbb{R}$ then

```
⊢ ∀x. 0 ≤ f x ⇒ 0 ≤ fn_integral m f
⊢ ∀x. f x ≤ g x ⇒ fn_integral m f ≤ fn_integral m g
⊢ fn_integral m (\x. c * f x) = a * fn_integral m f
⊢ fn_integral m (\x. f x + g x) = fn_integral m f + fn_integral m g
```

We only show the proof for the linearity of the integral. We start by proving the property for non-negative functions. Using the integrability property, given in Theorem 9, there exists two sequences (f_n) and (g_n) that are pointwise convergent to f and g, respectively, such that $\int_X f\,d\mu = \lim_n \int_X f_n\,d\mu$ and $\int_X g\,d\mu = \lim_n \int_X g_n\,d\mu$. Let $h_n = f_n + g_n$ then the sequence h_n is monotonically increasing, pointwise convergent to $f+g$ and $\forall\, x\ h_n(x) \leq (f+g)(x)$ and using Theorem 8, $\int_X f+g\,d\mu = \lim_n \int_X h_n\,d\mu$. Finally, using the linearity of the integral for positive simple functions and the linearity of the limit, $\int_X f+g\,d\mu = \lim_n \int_X f_n\,d\mu + \lim_n \int_X g_n\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu$. Now we consider arbitrary integrable functions. We first prove in HOL the following lemma.

Lemma 4. *If f_1 and f_2 are positive integrable functions such that $f = f_1 - f_2$ then $\int_X f\,d\mu = \int_X f_1\,d\mu - \int_X f_2\,d\mu$*

The definition of the integral is a special case of this lemma where $f_1 = f^+$ and $f_2 = f^-$. Going back to our proof, let $f_1 = f^+ + g^+$ and $f_2 = f^- + g^-$ then f_1 and f_2 are non-negative integrable functions satisfying $f + g = f_1 - f_2$. Using the lemma we conclude that
$\int_X f+g\,d\mu = \int_X f_1\,d\mu - \int_X f_2\,d\mu = (\int_X f^+\,d\mu + \int_X g^+\,d\mu) - (\int_X f^+\,d\mu + \int_X g^+\,d\mu) =$
$(\int_X f^+\,d\mu - \int_X f^-\,d\mu) + (\int_X g^+\,d\mu - \int_X g^-\,d\mu) = \int_X f\,d\mu + \int_X g\,d\mu$.

5.4 Lebesgue Monotone Convergence

The monotone convergence is arguably the most important theorem of the Lebesgue integration theory and it plays a major role in the proof of the Radon Nikodym theorem [2]. In this section, we present a proof of the theorem in HOL.

Theorem 10. *Let f be an integrable function and (f_n) be a sequence of functions such that $\forall\, n, x,\ 0 \leq f_n(x) \leq f_{n+1}(x) \leq f(x)$ and $\forall\, x,\ f_n(x) \to f(x)$. Then*

$$\int_X f\,d\mu = \lim_{n\to\infty} \int_X f_n\,d\mu$$

Proof. By the monotonicity of the integral, we deduce that $\forall\, n,\ \int_X f_n\,d\mu \leq \int_X f\,d\mu$. Hence $\lim_{n\to\infty} \int_X f_n\,d\mu \leq \int_X f\,d\mu$. It remains to prove that $\int_X f\,d\mu \leq \lim_{n\to\infty} \int_X f_n\,d\mu$. From Theorem 9, there exists a sequence of positive simple functions (g_n) such that $\forall\, n, x,\ g_n(x) \leq g_{n+1}(x) \leq f(x)$ and $\forall\, x,\ g_n(x) \to f(x)$ satisfying $\int_X f\,d\mu = \lim_{n\to\infty} \int_X g_n\,d\mu$. It is sufficient to prove that $\forall k \in \mathbb{N}$, $\int_X g_k\,d\mu \leq \lim_{n\to\infty} \int_X f_n\,d\mu$. For a fixed k, since g_k is a positive simple function then there exists $(\alpha_i), (a_i)$ and a finite set s such that

$$\int_X g_k\,d\mu = \sum_{i\in s} \alpha_i \mu(a_i)$$

On the other hand, splitting the integral of f_n and using the properties of the integral and limit, we have

$$\lim_{n\to\infty} \int_X f_n\,d\mu = \lim_{n\to\infty} \sum_{i\in s} \int_X f_n I_{a_i}\,d\mu = \sum_{i\in s} \lim_{n\to\infty} \int_X f_n I_{a_i}\,d\mu$$

Consequently, it suffices to prove that $\forall\, i \in s$

$$\alpha_i \mu(a_i) \leq \lim_{n\to\infty} \int_X f_n I_{a_i}\, d\mu$$

Or, equivalently, that $\forall\, i \in s$ and z such that $0 < z < 1$

$$z\alpha_i \mu(a_i) \leq \lim_{n\to\infty} \int_X f_n I_{a_i}\, d\mu$$

Let $b_n = \{t \in a_i : z\alpha_i \leq f_n(t)\}$ then $\bigcup_n b_n = a_i$ and

$$z\alpha_i \mu(a_i) = z\alpha_i \mu(\bigcup_n b_n) = z\alpha_i \lim_n \mu(b_n) = \lim_n z\alpha_i \mu(b_n) = \lim_n \int_X z\alpha_i I_{b_n}\, d\mu$$

Furthermore, from the definition of b_n and the monotonicity of the integral

$$\int_X z\alpha_i I_{b_n}\, d\mu \leq \int_X f_n I_{b_n}\, d\mu \leq \int_X f_n I_{a_i}\, d\mu$$

Concluding that

$$z\alpha_i \mu(a_i) \leq \lim_{n\to\infty} \int_X f_n I_{a_i}\, d\mu$$

5.5 Almost Everywhere

In this section we will define the "almost everywhere" relation [1] and prove in HOL some properties of the Lebesgue integral based on this relation. Consider a measure space $(X, \mathcal{A}, \mu)$. A null set E is a measurable set of measure zero.

Definition 10. *Almost Everywhere*
Let A be a subset of X and P be a property about elements of A. We say that P is true almost everywhere in A, abbreviated as "P a.e. in A", relative to the measure μ, if the subset of A where the property does not hold is a null set.

When $A = X$, we simply say "P a.e.".
For example, $f = g$ a.e. means that the set $\{x \mid f(x) \neq g(x)\}$ is a null set.
Similarly, $f_n \to f$ a.e. means that there exists a null set E such that $\forall\, x \in X \backslash E \quad f_n(x) \to f(x)$.

Theorem 11. *If A is a null set then for any measurable function f, $\int_A f\, d\mu = 0$*

Theorem 12. *If f and g are two integrable functions such that $f = g$ almost everywhere, then $\int_X f\, d\mu = \int_X g\, d\mu$*

Theorem 13. *If f and g are two integrable functions such that $f \leq g$ almost everywhere, then $\int_X f\, d\mu \leq \int_X g\, d\mu$*

We provide the proof of the first theorem as it is used to prove the last two.

Proof. It suffices to prove the theorem for positive measurable functions as the integral of an arbitrary function g is the difference of the integrals of g^+ and g^-. By definition, $\int_A f\,d\mu = \int_X fI_A\,d\mu = \sup\{\int_X g\,d\mu \mid g \leq fI_A\}$ where the functions g are positive simple functions.
We will show that the set over which the supremum is taken is equal to $\{0\}$. For a positive simple function g such that $g \leq fI_A$ we show that $g(x) = 0$ outside of A. Hence $\int_X g\,d\mu = \int_A g\,d\mu = \int_X gI_A\,d\mu$. Furthermore, there exists $(\alpha_i), (a_i)$ and a finite set s such that $\forall\, x \in X,\ g(x) = \sum_{i\in s} \alpha_i I_{a_i}(x)$. The indicator function of A can be split as $I_A = \sum_{i\in s} I_{A\cap a_i}$. Hence gI_A can be written as $gI_A = \sum_{i\in s} \alpha_i I_{A\cap a_i}$. This implies that $\int_X gI_A\,d\mu = \sum_{i\in s} \alpha_i \mu(A \cap a_i)$. Since $0 \leq \mu(A \cap a_i) \leq \mu(A) = 0$ and s is finite, then $\int_X g\,d\mu = 0$

6 Applications

In this section, we use our formalized Lebesgue integration theory to prove in HOL some important properties from the theory of probability, namely, the Chebyshev and Markov inequalities and the Weak Law of Large Numbers [16].

6.1 Chebyshev and Markov Inequalities

In probability theory, both the Chebyshev and Markov inequalities provide estimates of tail probabilities. The Chebyshev inequality guarantees, for any probability distribution, that nearly all the values are close to the mean and it plays a major role in the derivation of the laws of large numbers [16]. The Markov inequality provides loose yet useful bounds for the cumulative distribution function of a random variable.

Let X be a random variable with expected value m and finite variance σ^2. The Chebyshev inequality states that for any real number $k > 0$,

$$P(|X - m| \geq k\sigma) \leq \frac{1}{k^2} \tag{8}$$

The Markov inequality states that for any real number $k > 0$,

$$P(|X| \geq k) \leq \frac{E[X]}{k} \tag{9}$$

Instead of proving directly these inequalities, we provide a more general proof using measure theory and Lebesgue integrals in HOL that can be used for both and a number of similar inequalities. The probabilistic statement follows by considering a space of measure 1.

Theorem 14. *Let $(S, \mathcal{S}, \mu)$ be a measure space, and let f be a measurable function defined on S. Then for any nonnegative function g, nondecreasing on the range of f,*

$$\mu(\{x \in S\ :\ f(x) \geq t\}) \leq \frac{1}{g(t)} \int_S g \circ f\,d\mu\,.$$

The Chebyshev inequality is derived by letting $t = k\sigma$, $f = |X - m|$ and g defined as $g(t) = t^2$ if $t \geq 0$ and 0 otherwise. The Markov inequality is derived by letting $t = k$, $f = |X|$ and and g defined as $g(t) = t^2$ if $t \geq 0$ and 0 otherwise.

Proof. Let $A = \{x \in S : t \leq f(x)\}$ and I_A be the indicator function of A.

	Statement	Justification
(1)	$\forall x \quad 0 \leq g(t)I_A(x)$	g non-negative + definition of I_A
(2)	$\forall x \in A \quad t \leq f(x)$	definition of A
(3)	$\forall x \quad g(t)I_A(x) \leq g(f(x))I_A(x)$	g nondecreasing and (2)
(4)	$\forall x \quad g(f(x))I_A(x) \leq g(f(x))$	definition of I_A
(5)	$\forall x \quad g(t)I_A(x) \leq g(f(x))$	(3) + (4) + transitivity
(6)	$A \in \mathcal{S}$	f is $(\mathcal{S}, \mathcal{B}(\mathbb{R}))$ measurable
(7)	$x \to g(t)I_A(x)$ is integrable	integrability properties
(8)	$\int_S g(t)I_A(x)d\mu \leq \int_S g(f(x))d\mu$	(5) + monotonicity of the integral
(9)	$g(t)\mu(A) \leq \int_S g \circ f d\mu$	linearity of the integral

6.2 Weak Law of Large Numbers (WLLN)

The WLLN states that the average of a large number of independent measurements of a random quantity converges in probability towards the theoretical average of that quantity. Interpreting this result, the WLLN states that for a sufficiently large sample, there will be a very high probability that the average will be close to the expected value. This law is used in a multitude of fields. It is used, for instance, to prove the asymptotic equipartition property [4], a fundamental concept in the field of information theory.

Theorem 15. *Let $X_1, X_2, ...$ be an infinite sequence of independent, identically distributed random variables with finite expected value $E[X_1] = E[X_2] = ... = m$ and let $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ then for any $\varepsilon > 0$,*

$$\lim_{n \to \infty} P(|\overline{X} - m| < \varepsilon) = 1 \tag{10}$$

Besides the Chebyshev inequality, to prove this theorem in HOL, we need to formalize and prove some key properties of the variance of a random variable. The main property being that the variance of a sum of uncorrelated random variables is the sum of their variances. Notice that the requirement of the random variables being independent in the WLLN can be relaxed to simply requiring them to be uncorrelated.

Let X and Y be random variables with expected values μ_X and μ_Y, respectively. The variance of X is given by

$$Var(X) = E[|X - \mu_X|^2]$$

and the covariance between X and Y is given by

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

X and Y are uncorrelated iff $Cov(X, Y) = 0$.
We prove the following properties in HOL.

$Var(X) = E[X^2] - \mu_X^2$	
$Cov(X, Y) = E[XY] - \mu_X \mu_Y$	
$Var(X) \geq 0$	
$Var(aX) = a^2 Var(X)$	$\forall a \in \mathbb{R}$
$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$	
$Var(X + Y) = Var(X) + Var(Y)$	if X, Y uncorrelated
$Var(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} Var(X_i)$	$\forall i \neq j, X_i, X_j$ uncorrelated

Proof. Using the linearity property of the Lebesgue integral as well as the properties of the variance we prove that

$$E[\overline{X}] = \frac{1}{N} \sum_{i=1}^{N} m = m \tag{11}$$

$$Var(\overline{X}) = \frac{1}{N^2} \sum_{i=1}^{N} Var(X_i) = \frac{\sigma^2}{N} \tag{12}$$

Applying the Chebyshev inequality to $\overline{X}$, we get

$$P(|\overline{X} - m| \geq \varepsilon) \leq \frac{\sigma^2}{N\varepsilon^2} \tag{13}$$

Equivalently

$$1 - \frac{\sigma^2}{N\varepsilon^2} \leq P(|\overline{X} - m| < \varepsilon) \leq 1 \tag{14}$$

It then follows that

$$\lim_{n \to \infty} P(|\overline{X} - m| < \varepsilon) = 1$$

To prove the results of this section in HOL we used the Lebesgue integral properties, in particular, the monotonicity and the linearity, as well as the properties of real-valued measurable functions. All of this framework is not available in the work of Coble [3] because his formalization does not include the Borel sets so he cannot prove

the Lebesgue properties and the theorems of this section. The Markov and Chebyshev inequalities were previously proven by Hasan and Tahar [10] but only for discrete random variables. Our formalization allows us to provide a proof valid for both the discrete and continuous cases. Richter's formalization [17] only allows random variables defined on the whole universe of a certain type. All of the mentioned formalizations do not include the definition of variance and proofs of its properties and hence cannot be used to verify the WLLN.

7 Conclusion

In this report, we have presented a formalization in HOL of the Borel algebra to fill the gap in previous formalizations in higher-order-logic of the Lebesgue integral. Our formalization is general as it can be applied on functions defined on any metric space. Building on this framework, we proved important properties of the Lebesgue integral, in particular, the monotonicity and linearity properties. We also proved in HOL the Lebesgue monotone convergence, a key result of the Lebesgue integration theory. Additionally, we formalized the concept of "almost everywhere" and proved that the Lebesgue integral does not distinguish between functions which differ on a null set as well as other important results based on the "almost everywhere" relation. These features of the proposed approach facilitate the formal reasoning process for the continuous and unpredictable components of a wide range of physical systems. For illustration purposes, we proved in HOL key theorems from the theory of probability, namely the Chebyshev and Markov inequalities as well as the WLLN. The HOL codes corresponding to all the formalization and proofs, presented in this report, are available in [13].

Overall our formalization required more than 7000 lines of code. Only 250 lines were required to verify the key properties of the applications section. This shows the significance of our work in terms of simplifying the formal proof of properties using the Lebesgue integration theory. The main difficulties encountered were the multidisciplinary nature of this work, requiring deep knowledge of measure and integration theories, topology, set theory, real analysis and probability and information theories. Some of the mathematical proofs also posed challenges to be implemented in HOL.

Our future plans include using the Lebesgue integral development to formalize key concepts of the information theory. We will use the Lebesgue monotone convergence theorem and the Lebesgue integral properties to prove the Radon Nikodym theorem [2], paving the way to defining the probability density functions as well as the Kullback-Leibler divergence [4], which is related to the mutual information, entropy and conditional entropy [4].

8 Appendix

8.1 Rational Numbers

A rational number is any number that can be expressed as the quotient of two integers, the denominator of which is positive. We use natural numbers and express $\mathbb{Q}$, the set of rational numbers, as the union of positive ($\mathbb{Q}^+$) and negative ($\mathbb{Q}^-$) rational numbers.

$$\mathbb{Q} = \{r \mid \exists n, m.\ r = \frac{n}{m} \textit{ and } m > 0\} \cup \{r \mid \exists n, m.\ r = \frac{-n}{m} \textit{ and } m > 0\}.$$

We prove in HOL an extensive number of reassuring properties on the set $\mathbb{Q}$ as well as few other less straightforward ones, namely, $\mathbb{Q}$ is countable, infinite and dense in $\mathbb{R}$.

Theorem 16. $\forall n \in \mathbb{N},\ n \in \mathbb{Q}$ *and* $\forall x, y \in \mathbb{Q},\ -x,\ \frac{1}{x},\ x + y,\ x - y,\ x * y$ *and* $\frac{x}{y} \in \mathbb{Q}$

A proof of this theorem in HOL is at the same time straightforward and tedious but it is necessary to manipulate elements of the newly defined set of rational numbers and prove their membership to $\mathbb{Q}$ in the following theorems.

A set S is countable if its elements can be counted one at a time, or in other words, if every element of the set can be associated with a natural numbers, i.e. there exists an injective function $f : S \to \mathbb{N}$, or equivalently, there exists a surjective function from $g : \mathbb{N} \to S$. The next theorem states that the set of rational numbers $\mathbb{Q}$ is a countable set. To prove this theorem, we start by proving the following lemma in HOL.

Lemma 5. *There exists a bijection f from the set of natural numbers $\mathbb{N}$ to the cross product of $\mathbb{N}$ and $\mathbb{N}^*$ ($f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}^*$).*

$\mathbb{N}^*$ being $\mathbb{N}\backslash\{0\}$. This lemma is equivalent to the statement that there exists a bijection $g : \mathbb{N} \times \mathbb{N}^* \to \mathbb{N}$

Proof. Using the Schroeder Berstein theorem stating that if there exists injective functions $f : s \to t$ and $g : t \to s$ then there exists a bijective function $h : s \to t$, it suffices to prove that there exists an injective function $f : \mathbb{N} \times \mathbb{N}^* \to \mathbb{N}$ and a surjective function $g : \mathbb{N} \times \mathbb{N}^* \to \mathbb{N}$.
Surjection: Let $g : \mathbb{N} \times \mathbb{N}^* \to \mathbb{N}$ such that $g(x, y) = x$. g is clearly a surjective function because for any $n \in \mathbb{N}$, $g(n, 1) = n$ and $(n, 1) \in \mathbb{N} \times \mathbb{N}^*$.
Injection: Let $f : \mathbb{N} \times \mathbb{N}^* \to \mathbb{N}$ such that $f(x, y) = 2^x(2y + 1)$. We prove in HOL that f is an injection.

Theorem 17. *The set of rational numbers $\mathbb{Q}$ is countable.*

Proof. From Lemma 5, there exists a bijection f_1 from the set of natural numbers $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N}^*$ ($f_1 : \mathbb{N} \to \mathbb{N} \times \mathbb{N}^*$). Let $f_2 : \mathbb{N} \times \mathbb{N}^* \to \mathbb{Q}^+$ such that $f_2(a, b) = \frac{a}{b}$. and $f = f_2 \circ f_1$. Then $\forall x \in \mathbb{Q}^+$, there exists $n \in \mathbb{N}$ such that $f(n) = x$. This proves that $\mathbb{Q}^+$ is countable. Similarly, we prove that $\mathbb{Q}^-$ is countable and that the union of two countable sets is countable.

Theorem 18. $\mathbb{Q}$ *is dense in* $\mathbb{R}$*, i.e.,* $\forall x, y \in \mathbb{R}$ *and* $x < y$*, there exists* $r \in \mathbb{Q}$ *such that* $x < r < y$.

Proof. Let $x, y \in \mathbb{R}$ such that $x < y$. Define the ceiling of x as the smallest natural number larger than x, denoted by $\lceil x \rceil$. The ceiling satisfies the following two properties:

$$\forall x.\ x \leq \lceil x \rceil$$

$$\forall x \geq 0,\ \lceil x \rceil < x + 1$$

We first consider the case where $0 \leq x$.
If $1 < y - x$ then $r = \lceil x \rceil - 1 \in \mathbb{Q}$ and using the properties of the ceiling above, we prove that $x < r < y$.
If on the other hand $y - x \leq 1$, using the Archimedean property for $0 < y - x$, there exists a natural number such that $1 < n(y-x)$. From the first case, there exists $r_2 \in \mathbb{Q}$ such that $nx < r_2 < ny$. Let $r = \frac{r_2}{n}$ then $r \in \mathbb{Q}$ and $x < r < y$.
Finally, we consider the case $x < 0$.
Using the properties of the ceiling $0 \leq x + \lceil -x \rceil$, and since $x < y$, we also have $0 \leq x + \lceil -x \rceil < y + \lceil -x \rceil$. Then using the first case, there exists $r_2 \in \mathbb{Q}$ such that $x + \lceil -x \rceil < r_2 < y + \lceil -x \rceil$. Finally, $r = r_2 - \lceil -x \rceil$ satisfies $r \in \mathbb{Q}$ and $x < r < y$.

Another definition that will be useful in our development is the set of open intervals with rational end-points $I_r = \{]r_1, r_2[\colon r_1, r_2 \in \mathbb{Q}\}$. We prove that I_r is countable by showing that the mapping $I_r \to \mathbb{Q} \times \mathbb{Q}$ that sends an open interval $]r_1, r_2[\in I_r$ to the ordered pair of rational numbers $(r_1, r_2) \in \mathbb{Q} \times \mathbb{Q}$ is injective, and that the cross product of two countable sets, $\mathbb{Q}$ in this case, is countable.

8.2 Topology

To define the Borel sigma algebra on $\mathbb{R}$, we need some concepts of the topology of $\mathbb{R}$ formalized in HOL. Some of this was already developed by Harrison [8] but his formalization in HOL does not use the set theory and also lacks some of the important theorems that we need in our development. Harrison, later, developed an extensive topology theory [9] in HOL-Light. In the following, we define the concepts of neighborhood and open set in $\mathbb{R}$ and prove the required theorems.

Definition 11. *Let $a \in A \subset \mathbb{R}$. A is a neighborhood of a* iff *there exists a real number $d > 0$ such that $\forall x.\ |x - a| < d \Rightarrow\ x \in A$. In other words, a is an interior point of A.*

```
|- !A a.
     neighborhood_R A a <=>
     ?d. 0 < d /\ !y. a - d < y /\ y < a + d ==> y IN A
```

Definition 12. *A set that is a neighborhood to all of its points in an open set. Equivalently, if every point of a set is an interior point then the set is open.*

```
|- !A. open_set_R A <=> !x. x IN A ==> neighborhood_R A x
```

Theorem 19. *The empty set and the Universe are open.*

Theorem 20. *Every open interval is an open set.*

Theorem 21. *The union of any family of open sets is open. The intersection of a finite number of open sets is open.*

Theorem 22. *Every open set in* $\mathbb{R}$ *is the union of a countable family of open intervals.*

Proof. We only show the proof for Theorem 22. Let A be an open set in $\mathbb{R}$, then by the definition of open set, for all x in A there exists an open interval containing x such that $]a, b[\subset A$. Using the property of density of $\mathbb{Q}$ in $\mathbb{R}$, there exists $]a_r, b_r[\subset A$ containing x, a_r and b_r being rational numbers. A is the union of family of elements of I_r which is then countable because I_r is countable.

Theorem 23. *The inverse image of an open set by a continuous function is open.*

Proof. Let A be an open set in $\mathbb{R}$. From the previous theorem, A is a countable union of open intervals (A_i). $f^{-1}(A) = f^{-1}(\bigcup A_i) = \bigcup f^{-1}(A_i)$. Using Theorem 21, it suffices to prove that the inverse image of an open interval is open. For this we use the definition of a continuous function and the limit of a function to prove that any point of $f^{-1}(A_i)$ is an interior point.

References

[1] S. K. Berberian. *Fundamentals of Real Analysis*. Springer, 1998.

[2] V. I. Bogachev. *Measure Theory*. Springer, 2006.

[3] A. R. Coble. *Anonymity, Information, and Machine-Assisted Proof*. PhD thesis, University of Cambridge, 2010.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991.

[5] A. A. Fraenkel, Y. Bar-Hillel, and A. Levy. *Foundations of Set Theory*. North Holland, 1973.

[6] R. R. Goldberg. *Methods of Real Analysis*. Wiley, 1976.

[7] P. R. Halmos. The foundations of probability. *The American Mathematical Monthly*, 51(9):493–510, 1944.

[8] J. Harrison. *Theorem Proving with the Real Numbers*. Springer, 1998.

[9] J. Harrison. A HOL Theory of Euclidean Space. In *Theorem Proving in Higher Order Logics*, volume 3603 of *LNCS*, pages 114–129. Springer, 2005.

[10] O. Hasan and S. Tahar. Formal Verification of Tail Distribution Bounds in the HOL Theorem Prover. *Mathematical Methods in the Applied Sciences*, 32(4):480–504, March 2009.

[11] J. Hurd. *Formal Verifcation of Probabilistic Algorithms*. PhD thesis, University of Cambridge, 2002.

[12] D. Lester. Topology in PVS: Continuous Mathematics with Applications. In *workshop on Automated Formal Methods*, pages 11–20. ACM, 2007.

[13] T. Mhamdi, O. Hasan, and S. Tahar. Formalization of the Lebesgue Integration Theory in HOL. http://users.encs.concordia.ca/~mhamdi/hol/lebesgue/, 2010.

[14] T. Mhamdi, O. Hasan, and S. Tahar. On the Formalization of the Lebesgue Integration Theory in HOL. In *Interactive Theorem Proving*, volume 6172 of *LNCS*, pages 387–402. Springer, 2010.

[15] J. Munkres. *Topology (2nd Edition)*. Prentice Hall, 1999.

[16] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. Mc-Graw Hill, 1984.

[17] S. Richter. Formalizing Integration Theory, with an Application to Probabilistic Algorithms. Master's thesis, Technische Universität München, 2003.

[18] S. Wagon. *The Banach-Tarski Paradox*. Cambridge University Press, 1993.