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

Abstract. Lebesgue integration is a fundamental concept in many mathematical theories, such as real analysis, probability and information theory. 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 paper, 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.

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 [5], probability [6] and information theory, 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

M. Kaufmann and L. Paulson (Eds.): ITP 2010, LNCS 6172, pp. 387–402, 2010.

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 [3,15]. 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 paper, 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) [14], 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 [10] and Lebesgue integration [3] theories.

The rest of the paper is organized as follows: Section 2 provides a review of related work. In Section 3, we give an overview of main definitions of the measure theory [2]. Section 4 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 paper and provides hints to future work.

2 Related Work

Coble [3] generalized the measure theory formalization by Hurd [10] 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 [10], Richter [15] 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 paper 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 [15] defined the Borel sets as being generated by the intervals. In the formalization we propose in this paper, 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 paper key concepts of topology [13] in HOL.

In his work in topology in the PVS theorem prover, Lester [11] 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 Measure Theory

The measure theory formalization in HOL was essentially done in [10] and [3]. We make use of this formalization in our development and hence will only mention the main definitions. 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. A measure is defined on a class of subsets called the measurable sets. One important condition for a measure function is countable additivity, meaning that the measure of a countable collection of disjoint sets is the sum of their measures. This leads to the requirement that the measurable sets should form a sigma algebra.

Definition 1. *Let $\mathcal{A}$ be a collection of subsets 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.*

Definition 1 is formalized in HOL as

```
⊢ ∀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 and `subset_class` is defined 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 $f : \mathbb{N} \rightarrow S$.

```
⊢ ∀s. countable s = ∃f. ∀x. x ∈ s ⇒ ∃n. f n = 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 the smallest sigma algebra on X containing G, we call it the sigma algebra on X generated by G, denoted by $\sigma(X, G)$. 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.

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

Definition 2. *A triplet* $(X, \mathcal{A}, \mu)$ *is a measure space* iff $(X, \mathcal{A})$ *is a measurable space and* $\mu : \mathcal{A} \rightarrow \mathbb{R}$ *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 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. *Let* $(X_1, \mathcal{A}_1)$ *and* $(X_2, \mathcal{A}_2)$ *be two measurable spaces. A function* $f : X_1 \rightarrow X_2$ *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* $A \in \mathcal{A}_2$.

$f^{-1}(A)$ denotes the inverse image of A. 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
```

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.

4 Borel Theory

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 development in HOL of the Lebesgue integral and its properties.

Definition 4. *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 $\mathcal{B}(\mathbb{R})$.

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

Clearly, to formalize as well as prove in HOL various properties of $\mathcal{B}(\mathbb{R})$, 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 real numbers and not of a different HOL type. We will prove later that $\mathcal{B}(\mathbb{R})$ is generated by the open intervals. 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.

4.1 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 non-negative ($\mathbb{Q}^+$) and non-positive ($\mathbb{Q}^-$) rational numbers.

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

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 1. $\mathbb{N} \subset \mathbb{Q}$ *and* $\forall x, y \in \mathbb{Q}$, $-x$, $x + y$, $x - y$, $x * y \in \mathbb{Q}$ *and* $\forall y \neq 0$, $\frac{1}{y}$ *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.

Theorem 2. *The set of rational numbers $\mathbb{Q}$ is countable.*

Proof. We prove that there exists a bijection f_1 from the set of natural numbers $\mathbb{N}$ to the cross product of $\mathbb{N}$ and $\mathbb{N}^*$ ($f_1 : \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}^*$). Let $f_2 : \mathbb{N} \times \mathbb{N}^* \rightarrow \mathbb{Q}^+$ 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 3. *($\mathbb{Q}$ dense in $\mathbb{R}$)*
$\forall x, y \in \mathbb{R}$ and $x < y$, there exists $r \in \mathbb{Q}$ such that $x < r < y$.

Proof. We start by defining the ceiling of x as the smallest natural number larger than x, denoted by $\lceil x \rceil$ and prove that $\forall x,\ x \leq \lceil x \rceil$ and $\forall x \geq 0, \lceil x \rceil < x + 1$. Let $x, y \in \mathbb{R}$ such that $x < y$. We use the ceiling function and the Archimedean property to construct r such that $x < r < y$. □

Another definition that will be useful in our development is the set of open intervals with rational end-points $I_r = \{]r_1, r_2[: r_1, r_2 \in \mathbb{Q}\}$.

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

We prove that I_r is countable by showing that the mapping $I_r \rightarrow \mathbb{Q}\times\mathbb{Q}$ 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.

4.2 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 [7] 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 [8] in HOL-Light. In the following, we define the concepts of neighborhood and open set in $\mathbb{R}$ and prove the required theorems.

Definition 5. *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 ∈ A
```

Definition 6. *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 ∈ A ⇒ neighborhood_R A x
```

Theorem 4. *The empty set and the universal set are open.*

Theorem 5. *Every open interval is an open set.*

Theorem 6. *The union of any family of open sets is open. The intersection of a finite number of open sets is open.*

Theorem 7. *Every open set in* $\mathbb{R}$ *is the union of a countable family of open intervals.*

Proof. We only show the proof for Theorem 7. 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 8. *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 6, 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. □

4.3 Borel Measurable Sets

In this section, we prove in HOL that the Borel algebra on the real line $\mathcal{B}(\mathbb{R})$ is generated by the open intervals ($]c,d[$ for $c,d\in\mathbb{R}$). We show that it 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 9. $\mathcal{B}(\mathbb{R})$ *is generated by the open intervals* $]c,d[$ *where* $c,d\in\mathbb{R}$

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 5). Hence, $\sigma_I\subseteq\mathcal{B}(\mathbb{R})$. Conversely, $\mathcal{B}(\mathbb{R})$ is the intersection of all the sigma algebras containing the open sets. σ_I is one of them because every open set on the real line is the union of a countable collection of open intervals (Theorem 7). Consequently $\mathcal{B}(\mathbb{R})\subseteq\sigma_I$ and finally $\mathcal{B}(\mathbb{R})=\sigma_I$.

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

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 10. $\forall c\in\mathbb{R},\ \{c\}\in\mathcal{B}(\mathbb{R})$

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

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

4.4 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 11. *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 $\forall c \in \mathbb{R},\ f^{-1}(]-\infty, c[) \in \mathcal{A}$*

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 7 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. In fact, if $\forall x \in X,\ f(x) = k$, then if $c \le k,\ f^{-1}(]-\infty, c[) = \emptyset \in \mathcal{A}$. Otherwise $f^{-1}(]-\infty, c[) = X \in \mathcal{A}$. The indicator function on a set A is measurable iff A is measurable. In fact, $I_A^{-1}(]-\infty, c[) = \emptyset$, X or $X\backslash A$ when $c \le 0$, $c > 1$ or $0 < c \le 1$ respectively. We prove in HOL various properties of the real-valued measurable functions.

Theorem 12. *Let f and g be measurable functions and $c \in \mathbb{R}$ then the following functions are also measurable: $cf, |f|, f^n, f+g, fg$ and $max(f, g)$.*

Theorem 13. *If (f_n) is a sequence of real-valued measurable functions such that $\forall n, x,\ f_n(x) \to f(x)$ then f is a measurable function.*

Theorem 14. *Every continuous function $g : \mathbb{R} \to \mathbb{R}$ is measurable with respect to $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$.*

Theorem 15. *If $g : \mathbb{R} \to \mathbb{R}$ is continuous and f is measurable then $g \circ f$ is also measurable.*

Theorem 14 is a direct result of Theorem 8 stating that the inverse image of an open set by a continuous function is open. Theorem 15 guarantees, for instance, that if f is measurable then $exp(f)$, $Log(f)$, $cos(f)$ are measurable. This is derived using Theorem 14 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 3 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 Integral

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 can be written as a finite linear combination of indicator functions of measurable sets (a_i).

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

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

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

Various properties of the Lebesgue integral for positive simple functions have been proven in HOL [3]. We mention in particular that the integral above is well-defined and is 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. Utilizing our formalization of the Borel algebra, we have been able to overcome this problem. The new theorem can be stated as

Theorem 16. *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 $\int_X f\, d\mu = lim_{n\to\infty} \int_X f_n\, d\mu$.*

The next step towards the Lebesgue integration for arbitrary measurable functions is the definition of the Lebesgue integral of positive measurable functions which is given by

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

Finally, the integral for arbitrary measurable functions is given by

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

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

5.1 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 7. *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 $\int_X f^- \, d\mu < \infty$*

Theorem 17. *For any non-negative integrable function f there exists a sequence of positive simple functions (f_n) such that $\forall\, n, x,\ f_n(x) \leq f_{n+1}(x) \leq f(x)$ 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

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

Proof. Let the sequence (f_n) be defined as

$$f_n(x) = \sum_{k=0}^{4^n - 1} \frac{k}{2^n} I_{\{x : \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}\}} + 2^n I_{\{x : 2^n \leq f(x)\}} \tag{5}$$

We show that the sequence (f_n) satisfies the conditions of the theorem and use Theorem 16 to conclude that $\int_X f \, d\mu = \lim_n \int_X f_n \, d\mu$. First, we use the definition of (f_n) to prove in HOL the following lemmas

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

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

Using these lemmas we prove that the sequence (f_n) is pointwise convergent to f ($\forall\, x,\ f_n(x) \to f(x)$), upper bounded by f ($\forall\, n, x,\ f_n(x) \leq f(x)$) and monotonically increasing ($\forall\, n, x,\ f_n(x) \leq f_{n+1}(x)$). □

5.2 Lebesgue Integral Properties

We prove in HOL key properties of the Lebesgue integral, in particular the monotonicity and linearity. Let f and g be integrable functions and $c \in \mathbb{R}$ then

Theorem 18. $\forall x,\ 0 \leq f(x) \Rightarrow 0 \leq \int_X f\, d\mu$

Theorem 19. $\forall x,\ f(x) \leq g(x) \Rightarrow \int_X f\, d\mu \leq \int_X g\, d\mu$

Theorem 20. $\int_X cf\, d\mu = c \int_X f\, d\mu$

Theorem 21. $\int_X f + g\, d\mu = \int_X f\, d\mu + \int_X g\, d\mu$

Theorem 22. *A and B disjoint sets* $\Rightarrow \int_{A\cup B} f\, d\mu = \int_A f\, d\mu + \int_B f\, d\mu$

Proof. We only show the proof for Theorem 21. We start by proving the property for non-negative functions. Using the integrability property, given in Theorem 17, 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 16, $\int_X f + g\, d\mu = \lim_n \int_X h_n\, d\mu$. 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
$\int_X f+g\, d\mu = \int_X f_1\, d\mu - \int_X f_2\, d\mu = (\int_X f^+\, d\mu + \int_X g^+\, d\mu) - (\int_X f^+\, d\mu + \int_X g^+\, d\mu) = (\int_X f^+\, d\mu - \int_X f^-\, d\mu) + (\int_X g^+\, d\mu - \int_X g^-\, d\mu) = \int_X f\, d\mu + \int_X g\, d\mu$. □

5.3 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 23. *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 17, 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, $\int_X z\alpha_i I_{b_n}\, d\mu \leq \int_X f_n I_{b_n}\, d\mu \leq \int_X f_n I_{a_i}\, d\mu$. We conclude that $z\alpha_i\mu(a_i) \leq \lim_{n\to\infty} \int_X f_n I_{a_i}\, d\mu$. □

5.4 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 8. *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.

```
⊢ ∀m P. ae m P =
     {x | x ∈ m_space m ∧ ~P x} ∈ measurable_sets m ∧
     (measure m {x | x ∈ m_space m ∧ ~P x} = 0)
```

When $A = X$, we simply say "P a.e.". For example, $f = g$ a.e. means that the set $\{x \,|\, f(x) \neq g(x)\}$ 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 24. *If A is a null set then for any measurable function f, $\int_A f\, d\mu = 0$*

Theorem 25. *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 26. *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, $\int_A f\, d\mu = \int_X f I_A\, d\mu = \sup\{\int_X g\, d\mu \mid g \leq f I_A\}$ 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 $g \leq f I_A$ we show that $g(x) = 0$ outside of A. Hence $\int_X g\, d\mu = \int_A g\, d\mu = \int_X g I_A\, d\mu$. 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 $g I_A$ can be written as $g I_A = \sum_{i \in s} \alpha_i I_{A \cap a_i}$. This implies that $\int_X g I_A\, d\mu = \sum_{i \in s} \alpha_i \mu(A \cap a_i)$. 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 [14].

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 [14]. 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 $k > 0$,

$$P(|X - m| \geq k\sigma) \leq \frac{1}{k^2} \tag{6}$$

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

$$P(|X| \geq k) \leq \frac{m}{k} \tag{7}$$

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 27. *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,*

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

```
⊢ ∀m f g t.
    (let A = {x | x ∈ m_space m ∧ t ≤ f x} in
      measure_space m ∧
      f ∈ measurable (m_space m,measurable_sets m) Borel ∧
      (∀x. 0 ≤ g x) ∧ (∀x y. x ≤ y ⇒ g x ≤ g y) ∧
      integrable m (\x. g (f x)) ⇒
      measure m A ≤ (1 / (g t)) * fn_integral m (\x. g (f x)))
```

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. From the definition of A, $\forall x\ 0 \leq g(t)I_A(x)$ and $\forall x \in A\ t \leq f(x)$. Since g is non-decreasing, $\forall x,\ g(t)I_A(x) \leq g(f(x))I_A(x) \leq g(f(x))$. As a result, $\forall x\ g(t)I_A(x) \leq g(f(x))$. A is measurable because f is $(\mathcal{S}, \mathcal{B}(\mathbb{R}))$ measurable. Using the monotonicity of the integral, $\int_S g(t)I_A(x)d\mu \leq \int_S g(f(x))d\mu$. Finally from the linearity of the integral $g(t)\mu(A) \leq \int_S g \circ f d\mu$. □

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 28. *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 $\varepsilon > 0$,*

$$\lim_{n\to\infty} P(|\overline{X} - m| < \varepsilon) = 1 \tag{8}$$

```
⊢ ∀p X m v e.
  prob_space p ∧ 0 < e ∧
  (∀i j. i ≠ j ⇒ uncorrelated p (X i) (X j)) ∧
  (∀i. expectation p (X i) = m) ∧ (∀i. variance p (X i) = v) ⇒
    lim (\n. prob p {x | x ∈ p_space p ∧
      abs ((\x. 1/n * SIGMA (\i. X i x) (count n))x-m) < e}) = 1
```

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: $Var(X) = E[X^2] - \mu_X^2$; $Cov(X,Y) = E[XY] - \mu_X\mu_Y$; $Var(X) \geq 0$ and $Var(aX) = a^2Var(X)$. $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$ and $Var(X+Y) = Var(X) + Var(Y)$ if X and Y are uncorrelated. Finally if $\forall i \neq j, X_i, X_j$ are uncorrelated, then $Var(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} Var(X_i)$.

Proof. Using the linearity property of the Lebesgue integral as well as the properties of the variance we prove that $E[\overline{X}] = \frac{1}{N}\sum_{i=1}^{N} m = m$ and $Var(\overline{X}) = \frac{\sigma^2}{N}$. Applying the Chebyshev inequality to $\overline{X}$, we get $P(|\overline{X} - m| \geq \varepsilon) \leq \frac{\sigma^2}{N\varepsilon^2}$. Equivalently, $1 - \frac{\sigma^2}{N\varepsilon^2} \leq P(|\overline{X} - m| < \varepsilon) \leq 1$. 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. The above 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 [9] 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 [15] only allows random variables defined on the whole universe of a certain type. The above mentioned formalizations do not include the definition of variance and proofs of its properties and hence cannot be used to verify the WLLN.

7 Conclusions

In this paper, 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 paper, are available in [12].

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 theory. 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].

References

1. Berberian, S.K.: Fundamentals of Real Analysis. Springer, Heidelberg (1998)
2. Bogachev, V.I.: Measure Theory. Springer, Heidelberg (2006)
3. Coble, A.R.: Anonymity, Information, and Machine-Assisted Proof. PhD thesis, University of Cambridge, UK (2009)
4. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley-Interscience, Hoboken (1991)
5. Goldberg, R.R.: Methods of Real Analysis. Wiley, Chichester (1976)
6. Halmos, P.R.: The foundations of probability. The American Mathematical Monthly 51(9), 493–510 (1944)
7. Harrison, J.: Theorem Proving with the Real Numbers. Springer, Heidelberg (1998)
8. Harrison, J.: A HOL Theory of Euclidean Space. In: Hurd, J., Melham, T. (eds.) TPHOLs 2005. LNCS, vol. 3603, pp. 114–129. Springer, Heidelberg (2005)
9. Hasan, A., Tahar, S.: Formal Verification of Tail Distribution Bounds in the HOL Theorem Prover. Mathematical Methods in the Applied Sciences 32(4), 480–504 (2009)
10. Hurd, J.: Formal Verifcation of Probabilistic Algorithms. PhD thesis, University of Cambridge, UK (2002)
11. Lester, D.: Topology in PVS: Continuous Mathematics with Applications. In: workshop on Automated Formal Methods, pp. 11–20. ACM, New York (2007)
12. Mhamdi, T., Hasan, O., Tahar, S.: Formalization of the Lebesgue Integration Theory in HOL. Technical Report, ECE Dept., Concordia University (April 2009), `http://hvg.ece.concordia.ca/Publications/TECH_REP/MLP_TR10/`
13. Munkres, J.: Topology. Prentice Hall, Englewood Cliffs (1999)
14. Papoulis, A.: Probability, Random Variables, and Stochastic Processes. Mc-Graw Hill, New York (1984)
15. Richter, S.: Formalizing Integration Theory with an Application to Probabilistic Algorithms. In: Slind, K., Bunker, A., Gopalakrishnan, G.C. (eds.) TPHOLs 2004. LNCS, vol. 3223, pp. 271–286. Springer, Heidelberg (2004)