Formalization of Measure Theory and Lebesgue Integration for Probabilistic Analysis in HOL

TAREK MHAMDI, OSMAN HASAN, and SOFIÈNE TAHAR, Concordia University

Dynamic systems that exhibit probabilistic behavior represent a large class of man-made systems such as communication networks, air traffic control, and other mission-critical systems. Evaluation of quantitative issues like performance and dependability of these systems is of paramount importance. In this paper, we propose a generalized methodology to formally reason about probabilistic systems within a theorem prover. We present a formalization of measure theory in the HOL theorem prover and use it to formalize basic concepts from the theory of probability. We also use the Lebesgue integration to formalize statistical properties of random variables. To illustrate the practical effectiveness of our methodology, we formally prove classical results from the theories of probability and information and use them in a data compression application in HOL.

Categories and Subject Descriptors: F.4.1 [**Mathematical Logic and Formal Languages**]: Mathematical Logic—*Proof Theory*

General Terms: Verification, Reliability

Additional Key Words and Phrases: Probabilistic systems, Lebesgue integration, measure theory, theorem proving, statistical properties, information theory

ACM Reference Format:
Mhamdi, T., Hasan, O., and Tahar, S. 2013. Formalization of measure theory and Lebesgue Integration for probabilistic analysis in HOL. *ACM Trans. Embed. Comput. Syst.* 12, 1, Article 13 (January 2013), 23 pages.
DOI:http://dx.doi.org/10.1145/2406336.2406349

1. INTRODUCTION

Hardware and software systems usually exhibit some random or unpredictable elements. Examples include failures due to environmental conditions or aging phenomena in hardware components and the execution of certain actions based on a probabilistic choice in randomized algorithms. Moreover, these systems act upon and within complex environments that themselves have certain elements of unpredictability, such as noise effects in hardware components and the unpredictable traffic pattern in the case of telecommunication protocols. Due to these random components, establishing the correctness of a system under all circumstances usually becomes impractically expensive. The engineering approach to analyze a system with these kind of unavoidable elements of randomness and uncertainty is to use probabilistic analysis. Even for hardware and software systems for which correctness may be unconditionally guaranteed, the study of system performance primarily relies on probabilistic analysis. In fact, the term *system performance* commonly refers to the average time required by a system to perform a given task, such as the average runtime of a computational algorithm

Author's address: T. Mhamdi (corresponding author), Concordia University; email: tarek.mhamdi@gmail.com.

DOI:http://dx.doi.org/10.1145/2406336.2406349

or the average message delay of a telecommunication protocol. These averages can be computed, based on a probabilistic analysis approach, by using appropriate random variables to model inputs for the system model.

Traditionally, computer simulation techniques are used to perform probabilistic analysis. However, they provide less accurate results and cannot handle large-scale problems due to their enormous computer processing time requirements. This unreliable nature of the results poses a serious problem in safety-critical applications, such as those in space travel, military, and medicine. A possible solution for overcoming the accuracy problem of simulation is to conduct probabilistic analysis within the sound core of a higher-order logic theorem prover. Higher-order logic [Gordon 1989] is a system of deduction with a precise semantics and is expressive enough to be used for the specification of almost all classical mathematics theories. Due to its high expressive nature, higher-order logic can be utilized to precisely model the behavior of any system, while expressing its random or unpredictable elements in terms of formalized random variables, and any kind of system property, including the probabilistic and statistical ones, as long as they can be expressed in a closed mathematical form. Interactive theorem proving [Harrison 2009] is the field of computer science and mathematical logic concerned with precise computer based formal proof tools that require some sort of human assistance. Due to its interactive nature, interactive theorem proving can be utilized to reason about the correctness of probabilistic or statistical properties of systems, which are usually undecidable.

The foremost criteria for conducting a theorem proving based probabilistic analysis is to be able to formalize the underlying mathematical theories of measure [Bogachev 2006], probability [Halmos 1944], and Lebesgue integration [Berberian 1998] in higher-order logic. Using measure theory to formalize probability has the advantage of providing a mathematically rigorous treatment of probability and a unified framework for discrete and continuous probability measures. In this context, a probability measure is a measure function, an event is a measurable set and a random variable is a measurable function. The expectation of a random variable is its integral with respect to the probability measure. Lebesgue integration is the natural choice for developing statistical properties on random variables from measure theory and is used in most textbooks.

In the recent past, most of the above three fundamentals have been formalized in higher-order logic. For instance, Hurd [2002] formalized some measure and probability theory in the HOL theorem prover [Gordon and Melham 1993]. Richter [2004] and Coble [2010] built upon Hurd's formalization of measure theory to formalize Lebesgue integration using the Isabelle/HOL [Paulson 1994] and HOL theorem provers, respectively. Lester [2007] also attempted the formalization of all the three fundamental concepts of measure, probability and Lebesgue integral in the PVS theorem prover [Owre et al. 1992]. But, unfortunately, none of these formalizations can be termed as complete and thus each approach has its own limitations. For example, the available formalizations of measure theory do not allow the manipulation of random variables defined on arbitrary topological spaces, the probability theory formalizations do not not allow us to work with the sum of random variables as a random variable itself, and finally the Lebesgue integration formalizations have a very limited support for the Borel algebra [Bogachev 2006], which is a sigma algebra generated by the open sets. These deficiencies restrict the formal reasoning about some very useful probabilistic and statistical properties, which in turn limits the scope of theorem proving based probabilistic analysis of systems.

In this article, we present a generalized formalization of the measure and probability theories and Lebesgue integration in order to exploit their full potential for the formal analysis of probabilistic systems. We extend the measure theory with a general formalization of the Borel sigma algebra that can be used for any topological space. We prove important properties of real-valued measurable functions and use them to define real-valued random variables and prove their properties. We also formalize the concept of independence of random variables in HOL and prove key properties of independent random variables. Additionally, we prove important properties of the Lebesgue integral and use it to define statistical properties of random variables such as the expectation and variance. To the best of our knowledge, the above capabilities are not shared by any other existing formalization of measure, probability and Lebesgue integration.

Some of the possible applications of the proposed probability formalization include the verification of security protocols and communication systems. In this article, we formalize the Shannon entropy in higher-order logic as a measure of how much information was leaked [Smith 2009]. This result can be directly used to verify properties, such as the anonymity of classical security protocols like the dining cryptographers [Chaum 1988] and the crowds protocols [Reiter and Rubin 1998]. Similarly, our formalization can be used to verify properties of error-correcting codes. In order to illustrate the practical effectiveness of our work and its utilization to tackle such applications, we present in this article a data compression application in which we prove the Asymptotic Equipartition Property [Cover and Thomas 1991], a fundamental concept in information theory, and use it to prove the Shannon source coding theorem that establishes the limits of data compression [Cover and Thomas 1991]. The source coding theorem states that it is possible to compress the data and get a code rate that is arbitrarily close to the Shannon entropy without significant loss of information. Most of the infrastructure that we needed for this application, such as the properties of real valued measurable functions, properties of the expectation of arbitrary functions, variance, independence of random variables and the weak law of large numbers, was not available in the previous formalizations and thus the technical contributions of this article paved the path for the verification of Asymptotic Equipartition Property.

We use the HOL theorem prover for the above mentioned formalization and verification tasks. The main motivation behind this choice was to build on existing formalizations of measure [Hurd 2002] and Lebesgue integration [Coble 2010] theories in HOL

The rest of the article is organized as follows: Section 2 provides a review of related work. In Section 3, we give an overview of the main definitions of the measure theory and present our formalization of the Borel theory. Section 4 presents a formalization of probability spaces and random variables as well as independence of random variables and related properties. In Section 5 we prove the main properties of the Lebesgue integral that we use to define statistical properties of random variables. In this section we also prove important inequalities from the theory of probability as well as the Weak Law of Large Numbers. In Section 7, we illustrate the practical effectiveness of our formalization by proving fundamental results in information theory and data compression. Finally, Section 8 concludes the article and provides hints to future work.

2. RELATED WORK

The early foundations of probabilistic analysis in a higher-order-logic theorem prover were laid down by Nędzusiak [1989] and Bialas [1990] when they proposed a formalization of some measure and probability theories in higher-order logic. Hurd [2002] implemented their work and developed a formalization of measure theory in HOL, upon which he constructed definitions for probability spaces and functions on them. Despite important contributions, Hurd's formalization did not include basic concepts such as the expectation of random variables. Besides, in Hurd'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. Hence, 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.

Based on the work of Hurd [2002], Richter [2004] formalized the measure theory in Isabelle/HOL, where he has the same restriction on the the measure spaces that can be constructed. Richter [2004] defined the Borel sets as being generated by the intervals. In the formalization we propose in this article, 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 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.

Coble [2010] generalized the measure theory formalization by Hurd [2002] 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 based our work on the formalization of Coble [2010] where we define a measure space as a triplet $(X, \mathcal{A}, \mu)$; the set X being the space. We 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.

Hasan built upon Hurd and Coble's formalizations of measure, probability and Lebesgue integration to verify the probabilistic and statistical properties of some commonly used discrete [Hasan and Tahar 2007] and continuous [Hasan et al. 2009] random variables. The results were then utilized to formally reason about the correctness of many real-world systems that exhibit probabilistic behavior. Some examples include the analysis of the Coupon Collector's problem [Hasan and Tahar 2009a], the Stop-and-Wait protocol [Hasan and Tahar 2009b] and the repairability condition of hardware reconfigurable memory arrays in the presence of stuck-at and coupling faults [Hasan et al. 2009]. Hasan's work demonstrates the practical usefulness of formal probabilistic analysis using theorem proving but inherits the above mentioned limitations of Hurd and Coble's work. For example, separate frameworks for handling systems with discrete and continuous random variables are required and the inability to handle multiple continuous random variables. We believe that the formalization, presented in the current article, would allow us to utilize Hasan's formalized random variables for the analysis of a broader range of systems and properties.

In his work in topology on the PVS theorem prover, Lester [2007] provided formalizations for measure and integration theories but did not prove the properties of the Lebesgue integral nor its convergence theorems.

Based on the work of Coble [2010], we developed a formalization of the Lebesgue integration and verified its key properties and Lebesgue convergence theorems using the HOL theorem prover [Mhamdi et al. 2010b]. In the current article, we utilized some parts of this formalization to achieve a generalized methodology for analyzing systems with probabilistic behavior. The distinguishing features of the current article, when compared to this previous work of ours, include the formalization of random variables and their statistical properties as well as the formal proofs of classical results from the theories of probability, information, and communications technologies.

Besides theorem proving, probabilistic model checking is the second most widely used formal probabilistic analysis method [Baier et al. 2003; Rutten et al. 2004]. Like traditional model checking [Baier and Katoen 2008], probabilistic model checking involves the construction of a precise state-based mathematical model of the given probabilistic system, which is then subjected to exhaustive analysis to verify if it satisfies a set of probabilistic properties formally expressed in some appropriate logic. Numerous probabilistic model checking algorithms and methodologies have been proposed in the open literature, e.g., de Alfaro [1997] and Parker [2001], and based on these algorithms, a number of tools have been developed, e.g., PRISM [Kwiatkowska et al. 2005] and VESTA [Sen et al. 2005]. Besides the accuracy of the results, another promising feature of probabilistic model checking is the ability to perform the analysis automatically. On the other hand, probabilistic model checking is limited to systems that can only be expressed as probabilistic finite state machines or Markov chains. Another major limitation of the probabilistic model checking approach is state space explosion [Baier and Katoen 2008]. Similarly, to the best of our knowledge, it has not been possible to precisely reason about statistical relations, such as expectation and variance, using probabilistic model checking so far. Higher-order-logic theorem proving, on the other hand, overcomes the limitations of probabilistic model checking and thus allows conducting formal probabilistic analysis of algorithms but at the cost of significant user interaction.

3. MEASURE THEORY

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 [Fraenkel et al. 1973] 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 [Wagon 1993], which says that it is possible to decompose a solid unit ball into finitely many pieces and reassemble them into two copies of the original ball, using only rotations and no scaling. This paradox shows that there is no way to define the volume in three dimensions in the context of the ZFC set theory and at the same time require that the rotation preserves the volume, and that the volume of two disjoint sets is the sum of their volumes. The solution to this is to tag some sets as nonmeasurable and to assign a volume only to a measurable set. This solution was adopted in the measure theory by defining the measure only on a class of subsets called the measurable sets.

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. We define the measure 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.

Parts of the measure theory were formalized in Hurd [2002] and Coble [2010]. 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.

3.1. General Definitions

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.

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.
We define the space and subsets functions such that

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

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 and it 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.

Definition 2 (*Measure Space*). A triplet $(X, \mathcal{A}, \mu)$ is a measure space *iff* $(X, \mathcal{A})$ is a measurable space and $\mu : \mathcal{A} \rightarrow \mathbb{R}$ is a nonnegative 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.
```

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 $A \in \mathcal{A}_2$.

$f^{-1}(A)$ 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.
```

Notice that unlike Definition 3.1, 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 that 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.

3.2. 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.

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.

```
⊢ 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 Mhamdi et al. [2010b].

We prove in HOL the following theorem, stating that $\mathcal{B}(\mathbb{R})$, which, by definition, is generated by the open sets of $\mathbb{R}$, 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 as proven in Mhamdi et al. [2010b]. 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, a result we proved in [Mhamdi et al. 2010b]. 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 following equation:

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

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

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

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. We have shown above that $\forall c \in \mathbb{R},]-\infty, c[\in \mathcal{B}(\mathbb{R})$. If f is measurable with respect to $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ 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 $\forall A \in \mathcal{B}(\mathbb{R}), f^{-1}(A) \in \mathcal{A}$. Since $\mathcal{B}(\mathbb{R})$ is generated by the open sets, it is sufficient to prove the result for an open set A. Any open set of $\mathbb{R}$ can be written as a countable union of open intervals [Mhamdi et al. 2010b]. The result is then derived from 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 $c \leq k,\ f^{-1}(]-\infty, c[) = \emptyset \in \mathcal{A}$. 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 $c \leq 0$, $c > 1$ or $0 < c \leq 1$ respectively.

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 the theorem stating that the inverse image of an open set by a continuous function is open [Mhamdi et al. 2010b]. 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 the density of $\mathbb{Q}$ in $\mathbb{R}$ [Mhamdi et al. 2010b] 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 measurable set as a countable union of measurable sets because $\mathbb{Q}$ is countable [Mhamdi et al. 2010b] and f and g are measurable functions. □

This concludes our formalization of the measure theory in HOL. This formalization will allow us to define random variables, events and probability measures in the next section.

4. PROBABILITY THEORY

Probability provides mathematical models for random phenomena and experiments. The purpose is to describe and predict relative frequencies (averages) of these experiments in terms of probabilities of events.

The classical approach to formalize probabilities, which was the prevailing definition for many centuries, defines the probability of an event A as $p(A) = \frac{N_A}{N}$, Where N_A is the number of outcomes favorable to the event A and N is the number of all possible outcomes of the experiment. Problems with this approach include the assumption that all outcomes are equally likely (equiprobable), a concept of probability used to define probability itself, and hence this cannot be used as a basis for a mathematical theory. Besides, for many random experiments the outcomes are not equally likely. Finally the definition does not work when the number of possible outcomes is infinite.

Kolmogorov later introduced the axiomatic definition of probability that provides a mathematically consistent way for assigning and deducing probabilities of events. This approach consists in defining a set of all possible outcomes, Ω, called the sample space, A set F of events that are subsets of Ω and a probability measure p such that (Ω, F, p) is a measure space with $p(\Omega) = 1$.

Using measure theory to formalize probability has the advantage of providing a mathematically rigorous treatment of probabilities and a unified framework for discrete and continuous probability measures. In this context, a probability measure is a measure function, an event is a measurable set and a random variable is a measurable function. The expectation of a random variable is its integral with respect to the probability measure.

Basic definitions in the formalization of the probability theory in HOL are based on the work of Coble [2010]. Our contributions consist in going beyond the definitions to provide important theorems that will allow us to operate with the basic concepts such us random variables and their expected values. For instance, the formalization of Coble [2010] does not allow us to work with the sum of random variables as a random variable itself; we would have to add it as an assumption. Another important shortcoming is the lack of the properties of the expected value of a random variable such as the linearity and monotonicity.

Definition 5 (*Probability Space*). (Ω, F, p) is a probability space iff it is a measure space and $p(\Omega) = 1$.

```
⊢ ∀p. prob_space p = measure_space p ∧ (measure p (p_space p) = 1)
```

A probability measure is a measure function and an event is a measurable set.

```
⊢ ∀p. p_space p = m_space p
⊢ ∀p. prob p = measure p
⊢ ∀p. events p = measurable_sets p.
```

Definition 6 (*Independent Events*). Two events A and B are independent iff $p(A \cap B) = p(A)p(B)$.

Here $A \cap B$ is the intersection of A and B, that is, it is the event that both events A and B occur.

```
⊢ ∀p a b. indep p a b =
      a ∈ events p ∧ b ∈ events p ∧
      (prob p (a ∩ b) = prob p a * prob p b).
```

Definition 7 (*Random Variable*). $X : \Omega \to \mathbb{R}$ is a random variable iff X is $(F, \mathcal{B}(\mathbb{R}))$ measurable

```
⊢ ∀X p. random_variable X p Borel =
      prob_space p ∧ X ∈ measurable (p_space p,events p) Borel.
```

where F denotes, as previously, the set of events. Here we focus on real-valued random variables but the definition can be adapted for random variables having values on any topological space thanks to our general definition of the Borel sigma algebra.

```
⊢ ∀X p s. random_variable X p s =
      prob_space p ∧ X ∈ measurable (p_space p,events p) s.
```

The properties we proved for measurable functions are obviously valid for real-valued random variables.

THEOREM 8. *If X and Y are random variables and* $c \in \mathbb{R}$, *then the following functions are also random variables:* $cX, |X|, X^n, X+Y, XY$ *and* $max(X,Y)$.

```
⊢ ∀X Y c n p.
      random_variable X p Borel ∧ random_variable Y p Borel ⇒
      random_variable (\x. c * X x) p Borel ∧
      random_variable (\x. abs (X x)) p Borel ∧
      random_variable (\x. (X x) pow n) p Borel ∧
      random_variable (\x. X x + Y x) p Borel ∧
      random_variable (\x. X x * Y x) p Borel ∧
      random_variable (\x. max (X x) (Y x)) p Borel.
```

THEOREM 9. *If X is a random variable, then* $exp(X)$ *is also a random variable.*

```
⊢ ∀X p. random_variable X p Borel ⇒
      random_variable (\x. exp (X x)) p Borel.
```

THEOREM 10. *If X is a positive random variable, then so is* $Log(X)$.

```
⊢ ∀X p. random_variable X p Borel ∧ (∀x. 0 < f x) ⇒
      random_variable (\x. ln (X x)) p Borel.
```

We prove the last two theorems by first proving that the functions `(\x. exp(x))` and `(\x. ln(x))` are continuous and then use Theorem 3.2.

Definition 8 (*Independent Random Variables*). Two random variables X and Y are independent iff $\forall A, B \in \mathcal{B}(\mathbb{R})$, the events $\{X \in A\}$ and $\{Y \in B\}$ are independent.

The set $\{X \in A\}$ denotes the set of outcomes ω for which $X(\omega) \in A$. In other words $\{X \in A\} = X^{-1}(A)$.

```
⊢ ∀p X Y. indep_rv p X Y =
     ∀A B. A ∈ subsets Borel ∧ B ∈ subsets Borel ⇒
     indep p (PREIMAGE X A ∩ p_space p) (PREIMAGE Y B ∩ p_space p).
```

Equivalently, two random variables X and Y are independent iff $\forall A, B \in \mathcal{B}(\mathbb{R})$, $p(\{X \in A\} \cap \{Y \in B\}) = p(\{X \in A\})p(\{Y \in B\})$.
The event $\{X \in A\}$ is used to define the probability mass function (PMF) of a random variable.

Definition 9 (*Probability Mass Function*). The probability mass function p_X of a random variable X is defined as the function assigning to A the probability of the event $\{X \in A\}$.

$$\forall A \in \mathcal{B}(\mathbb{R}),\ p_X(A) = p(\{X \in A\}) = p(X^{-1}(A))$$

```
⊢ ∀p X. pmf p X = (\A. prob p (PREIMAGE X A ∩ p_space p)).
```

We also define the joint PMF of two random variables and of a sequence of random variables as

$$\forall A, B \in \mathcal{B}(\mathbb{R}),\ p_{XY}(A,B) = p(\{X \in A\} \cap \{Y \in B\}) = p(X^{-1}(A) \cap Y^{-1}(B))$$

$$\forall A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R}),\ p_{X_1 \ldots X_n}(A_1, \ldots, A_n) = p(\bigcap_{i=1}^{n} \{X_i \in A_i\}) = p(\bigcap_{i=1}^{n} X_i^{-1}(A_i))$$

```
⊢ ∀p X Y. joint_pmf p X Y =
     (\(A,B). prob p (PREIMAGE X A ∩ PREIMAGE Y B ∩ p_space p)).

⊢ ∀p X s. joint_pmf_sequence p X s =
     (\V. prob p (⋂ (IMAGE (\i. PREIMAGE (X i) (V i)) s)
                 ∩ p_space p)).
```

THEOREM 11. *If X and Y are independent, then $\forall A, B \in \mathcal{B}(\mathbb{R})$, $p_{XY}(A,B) = p_X(A)p_Y(B)$*

```
⊢ ∀X Y A B p.
     random_variable X p Borel ∧ random_variable Y p Borel ∧
     indep_rv p X Y  ∧ A ∈ subsets Borel ∧ B ∈ subsets Borel
     ⇒ (joint_pmf p X Y (A,B) = pmf p X A * pmf p Y B).
```

THEOREM 12. *If $X_1, \ldots, X_n$ are pairwise independent, then $\forall A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})$, $p_{X_1 \ldots X_n}(A_1, \ldots, A_n) = \prod_{i=1}^{n} p_{X_i}(A_i)$*

```
⊢ ∀X s V p.
     FINITE s ∧ (∀i. i ∈ s ⇒ random_variable (X i) p Borel) ∧
     (∀i j. i ∈ s ∧ j ∈ s ∧ i ≠ j ⇒ indep_rv p (X i) (X j)) ∧
     (∀i. i ∈ s ⇒ V i ∈ subsets Borel) ⇒
     (joint_pmf_sequence p X s V = PROD (\i. pmf p (X i) (V i)) s).
```

In this section we defined basic concepts of probability like the events, probability measures and random variables. Our main contributions in this section are the properties of real-valued random variables as well as the formalization of the notion of independence of random variables and properties related to the joint PMF of a sequence of mutually independent random variables. The next step towards a comprehensive formalization of probability in higher-order logic is the definition of main statistical properties of random variables, such as the expectation and the variance. The expectation of a random variable is its integral with respect to the probability measure. Lebesgue is the natural choice and will be discussed next.

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. The Lebesgue integral is first defined for those functions then extended to nonnegative functions and finally to arbitrary functions. A positive simple function g is a measurable function taking finitely many values. In other words, it can be written as a finite linear combination of indicator functions of measurable sets (a_i) that form a partition of X.

$$\forall x \in X,\ g(x) = \sum_{i \in s} \alpha_i I_{a_i}(x) \quad c_i \geq 0. \tag{3}$$

Definition 10. 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

$$\int_X g\, d\mu = \sum_{i \in s} \alpha_i \mu(a_i) \tag{4}$$

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

The choice of $((\alpha_i), (a_i), s)$ to represent g is not unique. However, the integral as defined above is independent of that choice.
Next, we define the Lebesgue integral of nonnegative measurable functions

Definition 11. Let $(X, \mathcal{A}, \mu)$ be a measure space. The integral of a nonnegative measurable function f is defined as

$$\int_X f\, d\mu = \sup \left\{ \int_X g\, d\mu \mid g \leq f \ \text{ and } g \text{ positive simple function} \right\} \tag{5}$$

```
  ⊢ ∀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 arbitrary measurable functions is given in the following definition.

Definition 12. Let $(X, \mathcal{A}, \mu)$ be a measure space. The integral of an arbitrary measurable function f is defined as

$$\int_X f\, d\mu = \int_X f^+\, d\mu - \int_X f^-\, d\mu, \tag{6}$$

where f^+ and f^- are the nonnegative 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 [Coble 2010]. 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 Coble [2010] 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 follows.

THEOREM 13. *Let $(X, \mathcal{A}, \mu)$ be a measure space, f a nonnegative 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.

5.1. Integrability

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

Definition 13 (*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 =
   measure_space m ∧
   f ∈ measurable (m_space m,measurable_sets m) Borel ∧
   (∃z. ∀r.
     r ∈ {r | ∃g. r ∈ psfis m g ∧ ∀x. g x ≤ fn_plus f x} ⇒ r ≤ z) ∧
   ∃z. ∀r.
      r ∈ {r | ∃g. r ∈ psfis m g ∧ ∀x. g x ≤ fn_minus f x} ⇒ r ≤ z.
```

THEOREM 14. *For any nonnegative 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

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

```
⊢ ∀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)
```

PROOF. Let the sequence (f_n) be defined as

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

We show that the sequence (f_n) satisfies the conditions of the theorem and use Theorem 5 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. $\forall n, x,\ f(x) \ge 2^n \Rightarrow f_n(x) = 2^n$.

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

LEMMA 3. $\forall x,\ (f(x) \ge 2^n) \vee \left(\exists k,\ k < 4^n \text{ and } \frac{k}{2^n} \le f(x) < \frac{k+1}{2^n}\right)$.

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

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

Monotonicity. $\forall n, x,\ f_n(x) \le f_{n+1}(x)$
If $f(x) \ge 2^{n+1}$ then $f_n(x) = 2^n$ and $f_{n+1}(x) = 2^{n+1}$. Hence $f_n(x) \le f_{n+1}(x)$. if $f(x) < 2^{n+1}$ then using Lemma 3, there exists a $k < 4^{n+1}$ such that $\frac{k}{2^{n+1}} \le f(x) < \frac{k+1}{2^{n+1}}$ 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 $f_n(x) \leq f_{n+1}(x)$. □

5.2. 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

— $\forall x,\ 0 \leq f(x) \Rightarrow 0 \leq \int_X f\,d\mu$
— $\forall x,\ f(x) \leq g(x) \Rightarrow \int_X f\,d\mu \leq \int_X g\,d\mu$
— $\int_X cf\,d\mu = c\int_X f\,d\mu$
— $\int_X f + g\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu$
— A and B disjoint sets $\Rightarrow \int_{A\cup B} f\,d\mu = \int_A f\,d\mu + \int_B f\,d\mu$

PROOF. We only show the proof for the linearity of the integral. We start by proving the property for nonnegative functions. Using the integrability property, given in Theorem 14, 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 13, $\int_X f + g\,d\mu = \lim_n \int_X h_n\,d\mu$. 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 nonnegative integrable functions satisfying $f + g = f_1 - f_2$. Using the lemma we conclude that $\int_X f + g\,d\mu = \int_X f_1\,d\mu - \int_X f_2\,d\mu = (\int_X f^+\,d\mu + \int_X g^+\,d\mu) - (\int_X f^+\,d\mu + \int_X g^+\,d\mu) = (\int_X f^+\,d\mu - \int_X f^-\,d\mu) + (\int_X g^+\,d\mu - \int_X g^-\,d\mu) = \int_X f\,d\mu + \int_X g\,d\mu$. □

In this section we presented a formalization of the Lebesgue integration in HOL. We also provided the criteria of integrability of a measurable function and proved the integrability theorem. We used this theorem to prove the properties of the Lebesgue integral for arbitrary measurable functions compared to only positive simple functions in the work of Coble [2010]. This formalization allows us to define the expectation and other statistical properties of random variables and prove their properties.

6. STATISTICAL PROPERTIES

We use our formalization of the Lebesgue integral to define the expected value of a random variable and prove its properties.

6.1. Expected Value

The expected value of a random value X is defined as the integral of X with respect to the probability measure.

Definition 14. $E[X] = \int_{\Omega} X\, dp$. ⊢ `expectation = fn_integral`.

The properties of the expectation are derived from the properties of the integral. We focus on random variables for which the expected value exists and prove, among other properties, the linearity and monotonicity of the expectation.

6.2. Variance

The variance and covariance of random variables are formalized as follow

```
⊢ variance p X = expectation p (\x. (X x - expectation p X) pow 2)
```

```
⊢ covariance p X Y =  expectation p (\x. (X x - expectation p X) *
                (Y x - expectation p Y)).
```

Two random variable X and Y are uncorrelated iff $Cov(X, Y) = 0$.

```
⊢ uncorrelated p X Y = (covariance p X Y = 0).
```

We prove the following properties in HOL.

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

Next, we use our formalization of the probability concepts in HOL to prove some important properties, namely, the Chebyshev and Markov inequalities and the Weak Law of Large Numbers [Papoulis 1984].

6.3. 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 values are close to the mean and it plays a major role in the derivation of the laws of large numbers [Papoulis 1984]. 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 $k > 0$,

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

```
⊢ ∀p X k.
     random_variable X p Borel ∧
     integrable p (\x. (X x - expectation p X) pow 2) ∧ 0 < k ⇒
     prob p {x | x ∈ p_space p ∧ k <= abs (X x - expectation p X)}
     <= variance p X / k pow 2.
```

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

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

```
⊢ ∀p X k.
     random_variable X p Borel ∧ 0 < k ⇒
     prob p {x | x ∈ p_space p ∧ k <= abs (X x)} <=
     expectation p X / k.
```

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 as well as for a number of similar inequalities. The probabilistic statement follows by considering a space of measure 1.

THEOREM 15. *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 $g(t)\mu(A) \leq \int_S g \circ f d\mu$. □

6.4. 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 [Cover and Thomas 1991], a fundamental concept in the field of information theory.

THEOREM 16. *Let $X_1, X_2, \ldots$ be an infinite sequence of independent, identically distributed random variables with finite expected value $E[X_1] = E[X_2] = \ldots = m$ and let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ then for any $\varepsilon > 0$,*

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

```
⊢ ∀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.
```

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 $Var(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$.
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 $\lim_{n\to\infty} P(|\overline{X} - m| < \varepsilon) = 1$.
Notice that in the proof we did not need to assume that the random variables are independent and identically distributed. We simply assume that they are uncorrelated and have the same expected value and variance. □

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 these results are not available in the work of Coble [2010] 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 [2009a] 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 [Richter 2004] 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. SHANNON SOURCE CODING THEOREM

The source coding theorem establishes the limits of data compression. It states that n independent and identically distributed (iid) random variables with entropy $H(X)$ can be expressed on the average by $nH(X)$ bits without significant risk of information loss, as n tends to infinity.

A proof of this result consists in proposing an encoding scheme for which the average codeword length can be made arbitrarily close to $nH(X)$ with negligible probability of loss. We start by proving the Asymptotic Equipartition Property (AEP) [Cover and Thomas 1991] and use it to define the *typical set* that will be the basis of the encoding scheme.

7.1. Asymptotic Equipartition Property (AEP)

The Asymptotic Equipartition Property is the Information Theory analog of the WLLN. It states that for a stochastic source X, if its time series $X_1, X_2, \ldots$ is a sequence of *iid* random variables with entropy $H(X)$, then $-\frac{1}{n}log(p_{X_1 \ldots X_n})$ converges in

probability to $H(X)$.
We define the entropy of a random source as $H(X) = E[-log(p_X)]$.

THEOREM 17. *(AEP): if $X_1, X_2, \ldots$ are iid, then*

$$-\frac{1}{n} log\big(p_{X_1 \ldots X_n}\big) \longrightarrow H(X) \textit{ in probability.}$$

PROOF. Let $X_1, X_2, \ldots$ be *iid* random variables and let $Y_i = -log(p_{X_i})$. Then $Y_1, Y_2, \ldots$ are *iid* random variables and $\forall i$, $E[Y_i] = H(X)$. Using Theorem 16, we have

$$\lim_{n \to \infty} P\big(|\frac{1}{n}\sum_{i=1}^{n} Y_i - H(X)| < \varepsilon\big) = 1.$$

Furthermore,

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n} -log\big(p_{X_i}\big) = -\frac{1}{n} log\big(\prod_{i=1}^{n} p_{X_i}\big).$$

Using Theorem 12, since $X_1, \ldots, X_n$ are mutually independent,

$$-\frac{1}{n} log\big(\prod_{i=1}^{n} p_{X_i}\big) = -\frac{1}{n} log\big(p_{X_1 \ldots X_n}\big).$$

Consequently,

$$\lim_{n \to \infty} P\big(| -\frac{1}{n} log\big(p_{X_1 \ldots X_n}\big) - H(X)| < \varepsilon\big) = 1. \tag{11}$$

□

7.2. Typical Set

A consequence of the AEP is the fact that the set of observed sequences $(x_1, \ldots, x_n)$ which joint probabilities $p(x_1, x_2, \ldots, x_n)$ are close to $2^{-nH(X)}$ has a total probability equal to 1. This set is called the *typical set* and such sequences are called the *typical sequences*. In other words, out of all possible sequences, only a small number of sequences will actually be observed and those sequences are nearly equally probable. The AEP guarantees that any property that is proved for the typical sequences will then be true with high probability and will determine the average behavior of a large sample.

Definition 15. The typical set A_ϵ^n with respect to $p(x)$ is the set of sequences $(x_1, \ldots, x_n)$ satisfying

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, \ldots, x_n) \leq 2^{-n(H(X)-\epsilon)}. \tag{12}$$

We prove in HOL various properties of the typical set.

THEOREM 18. If $(x_1, \ldots, x_n) \in A_\epsilon^n$, *then*

$$H(X) - \epsilon \leq -\frac{1}{n} log(p(x_1, \ldots, x_n)) \leq H(X) + \epsilon \tag{13}$$

This theorem is a direct consequence of Definition 15.

THEOREM 19. $\forall \epsilon > 0,\ \exists N,\ \forall n \geq N,\ p(A_\epsilon^n) > 1 - \epsilon.$

The proof of this theorem is derived from the formally verified AEP. The next two theorems give upper and lower bounds for the number of typical sequences $|A_\epsilon^n|$.

THEOREM 20. $|A_\epsilon^n| \leq 2^{n(H(X)+\epsilon)}$.

PROOF. Let $\underline{x} = (x_1, \ldots, x_n)$, then $\sum_{\underline{x} \in A_\epsilon^n} p(\underline{x}) \leq 1$. From Equation (12), $\forall \underline{x} \in A_\epsilon^n$, $2^{-n(H(X)+\epsilon)} \leq p(\underline{x})$. Hence $\sum_{\underline{x} \in A_\epsilon^n} 2^{-n(H(X)+\epsilon)} \leq \sum_{\underline{x} \in A_\epsilon^n} p(\underline{x}) \leq 1$. Consequently, $2^{-n(H(X)+\epsilon)} |A_\epsilon^n| \leq 1$ proving the theorem. □

THEOREM 21. $\forall \epsilon > 0, \exists N. \forall n \geq N,\ (1-\epsilon) 2^{n(H(X)-\epsilon)} \leq |A_\epsilon^n|$.

PROOF. Let $\underline{x} = (x_1, \ldots, x_n)$. From Theorem 19, $\exists N. \forall n \geq N,\ 1-\epsilon < \sum_{\underline{x} \in A_\epsilon^n} p(\underline{x})$. From Equation 12, $\forall \underline{x} \in A_\epsilon^n, p(\underline{x}) \leq 2^{-n(H(X)-\epsilon)}$. Hence, $\exists N. \forall n \geq N,\ 1-\epsilon < \sum_{\underline{x} \in A_\epsilon^n} p(\underline{x}) \leq \sum_{\underline{x} \in A_\epsilon^n} 2^{-n(H(X)-\epsilon)}$. Consequently, $\exists N. \forall n \geq N,\ 1-\epsilon < 2^{-n(H(X)-\epsilon)} |A_\epsilon^n|$ proving the theorem. □

7.3. Shannon Source Coding Theorem

The main idea behind the proof of the source coding theorem is that the average codeword length for all sequences is close to the average codeword length considering only the typical sequences. This is true because according to Theorem 19, for a sufficiently large n, the typical set has a total probability close to 1. In other words, for any $\epsilon > 0$, and sufficiently large n, the probability of observing a nontypical sequence is less than ϵ. Furthermore, according to Theorem 20, the number of typical sequences is smaller than $2^{n(H(X)+\epsilon)}$ and hence no more than $n(H(X)+\epsilon)+1$ bits are needed to represent all typical sequences. If we encode each typical sequence by simply enumerating its position within an ordered list of typical sequences and add 0 as a prefix, the total number of bits needed is no more than $n(H(X)+\epsilon)+2$. We also encode each nontypical sequence by enumerating its position within an ordered list of all possible sequences and prefix it by 1. The total number of bits needed for the nontypical sequences is less than $nlog(|\Omega|)+2$. If we denote by Y the random variable defined over all the possible sequences and returns the corresponding codeword length. The expectation of the Y is equal to the average codeword length $\overline{L}$. Besides, $\forall x \in A_\epsilon^n$, $Y(x) \leq n(H(X)+\epsilon)+2$ otherwise $Y(x) \leq nlog(|\Omega|)+2$.

$$\overline{L} = E[Y] = \sum_{\underline{x} \in A_\epsilon^n} p(\underline{x}) Y(\underline{x}) + \sum_{\underline{x} \in \overline{A_\epsilon^n}} p(\underline{x}) Y(\underline{x}) \tag{14}$$

$$\overline{L} \leq p(A_\epsilon^n)(n(H(X)+\epsilon)+2) + (1-p(A_\epsilon^n))(nlog(|\Omega|)+2) \tag{15}$$

$$\overline{L} \leq p(A_\epsilon^n) n(H(X)+\epsilon) + (1-p(A_\epsilon^n)) nlog(|\Omega|)+2. \tag{16}$$

Using Theorem 19, $\exists N,\ \forall n \geq N,\ p(A_\epsilon^n) > 1-\epsilon$. Hence,

$$\overline{L} \leq n(H(X)+\epsilon) + \epsilon nlog(|\Omega|)+2 \tag{17}$$

$$\overline{L} \leq n(H(X)+\epsilon^{'}), \tag{18}$$

where $\epsilon^{'} = \epsilon + \epsilon log(|\Omega|) + \frac{2}{n}$.

Consequently, for any $\epsilon > 0$ and n sufficiently large, the code rate $\frac{\overline{L}}{n}$ can be made as close as needed to the entropy $H(X)$ while maintaining a probability of error of the encoder that is bounded by ϵ.

The coding scheme we used is one-to-one, as the first bit in the codeword indicates its length and the remaining bits determine its position in the corresponding ordered set. We formally verified that the average codeword length of this code can be made arbitrarily close to $nH(X)$ without significant loss of information.

8. CONCLUSIONS

In this article we have presented a comprehensive methodology to reason about probabilistic systems in a theorem prover. We provided a formalization of the measure theory including the Borel sigma algebra defined for any topological space. We then focused on real-valued measurable functions and proved their key properties. Using this formalization we defined a theory of probability in Higher-order logic. Main concepts of probability in our formalization include random variables, probability mass functions (pmf), joint pmf, and independence of random variables. To formalize statistical properties of random variables, we presented a Lebesgue integration infrastructure including important properties such as the linearity and monotonicity of the Lebesgue integral. We applied the Lebesgue integral properties to the expectation operator and used the whole formalization to prove classical results from the theory of probability, namely, the Chebyshev and Markov inequalities as well as the WLLN.

The proposed work paves the path for the usage of formal verification for analyzing probabilistic aspects in many critical domains, such as security protocols, transportation and communication systems. We illustrated the effectiveness of our formalization by proving a fundamental result in information theory, namely the Asymptotic Equipartition Property, which we used to prove the classical source coding theorem. The HOL codes corresponding to all the formalization and proofs, presented in this article, are available in Mhamdi et al. [2010a] and can be built upon to formally analyze other interesting applications.

Overall our formalization required more than 9000 lines of code. Only 700 lines were required to verify the key properties of the application section. This shows the significance of our work in terms of simplifying the formal analysis of probabilistic systems. 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 monotone convergence theorem and the Lebesgue integral properties to prove the Radon Nikodym theorem [Bogachev 2006], paving the way to defining the probability density functions for continuous random variables as well as the Kullback-Leibler divergence [Cover and Thomas 1991], which is related to the mutual information, entropy and conditional entropy.

REFERENCES

Baier, C. and Katoen, J. 2008. *Principles of Model Checking*. MIT Press.

Baier, C., Haverkort, B., Hermanns, H., and Katoen, J. 2003. Model checking algorithms for continuous time Markov chains. *IEEE Trans. Softw. Engin 29,* 4, 524–541.

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

Bialas, J. 1990. The σ-additive measure theory. *J. Formal. Math. 2*.

Bogachev, V. I. 2006. *Measure Theory*. Springer.

Chaum, D. 1988. The dining cryptographers problem: Unconditional sender and recipient untraceability. *J. Cryptology 1,* 1, 65–75.

Coble, A. R. 2010. Anonymity, information, and machine-assisted proof. Ph.D. thesis, University of Cambridge.

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

de Alfaro, L. 1997. Ph.D. thesis, Stanford University.

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

Gordon, M. 1989. Mechanizing programming logics in higher-order logic. In *Current Trends in Hardware Verification and Automated Theorem Proving*. Springer, 387–439.

Gordon, M. and Melham, T. 1993. *Introduction to HOL: A theorem proving environment for higher-order logic*. Cambridge University Press.

Halmos, P. R. 1944. The foundations of probability. *Amer. Math. Monthly 51,* 9, 493–510.

Harrison, J. 2009. *Handbook of Practical Logic and Automated Reasoning*. Cambridge University Press.

Hasan, O. and Tahar, S. 2007. Verification of expectation properties for discrete random variables in HOL. In *Theorem Proving in Higher-Order Logics*. Lecture Notes in Computer Science, vol. 4732. Springer, 119–134.

Hasan, O. and Tahar, S. 2009a. Formal verification of tail distribution bounds in the HOL theorem prover. *Math. Methods Appl. Sci. 32,* 4 (March), 480–504.

Hasan, O. and Tahar, S. 2009b. Performance analysis and functional verification of the stop-and-wait protocol in HOL. *J. Autom. Reasoning 42,* 1, 1–33.

Hasan, O., Abbasi, N., Akbarpour, B., Tahar, S., and Akbarpour, R. 2009. Formal reasoning about expectation properties for continuous random variables. In *Proceedings of the 2nd World Congress on Formal Methods*. Lecture Notes in Computer Science, vol. 5850. 435–450.

Hasan, O., Tahar, S., and Abbasi, N. 2009. Formal reliability analysis using theorem proving. *Trans. Comput. 59*, 579–592.

Hurd, J. 2002. Formal verifcation of probabilistic algorithms. Ph.D. thesis, University of Cambridge.

Kwiatkowska, M., Norman, G., and Parker, D. 2005. Quantitative analysis with the probabilistic model checker PRISM. *Electron. Notes in Theor Comput Sci. 153,* 2, 5–31. Elsevier.

Lester, D. 2007. Topology in PVS: Continuous mathematics with applications. In *Proceedings of the Workshop on Automated Formal Methods*. ACM, 11–20.

Mhamdi, T., Hasan, O., and Tahar, S. 2010a. Formal analysis of systems with probabilistic behavior in HOL. http://users.encs.concordia.ca/˜mhamdi/hol/probability/.

Mhamdi, T., Hasan, O., and Tahar, S. 2010b. On the formalization of the Lebesgue integration theory in HOL. In *Proceedings of the Conference on Interactive Theorem Proving*. 387–402.

Nędzusiak, A. 1989. σ-fields and Probability. *J. Formal. Math. 1*.

Owre, S., Rushby, J. M., and Shankar, N. 1992. PVS: A prototype verification system. In *Proceedings of the 11th International Conference on Automated Deduction*. Lecture Notes in Computer Science, vol. 607. 748–752.

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

Parker, D. 2001. Ph.D. thesis, University of Birmingham, Birmingham, UK.

Paulson, L. C. 1994. *Isabelle: A Generic Theorem Prover*. Springer.

Reiter, M. K. and Rubin, A. D. 1998. Crowds: Anonymity for web transactions. *ACM Trans. Inf. Syst. Secur. 1,* 1, 66–92.

Richter, S. 2004. Formalizing integration theory with an application to probabilistic algorithms. In *Proceedings of the 17the International Conference on Theorem Proving in Higher Order Logics.* Lecture Notes in Computer Science. vol. 3223. 271–286.

Rutten, J., Kwaiatkowska, M., Normal, G., and Parker, D. 2004. *Mathematical Techniques for Analyzing Concurrent and Probabilisitc Systems*. CRM Monograph Series, vol. 23. American Mathematical Society.

Sen, K., Viswanathan, M., and Agha, G. 2005. VESTA: A statistical model-checker and analyzer for probabilistic systems. In *Proceedings of the IEEE International Conference on the Quantitative Evaluation of Systems*. 251–252.

Smith, G. 2009. On the foundations of quantitative information flow. In *Proceedings of the Conference on Foundations of Software Science and Computational Structures*. 288–302.

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

Received March 2010; revised October 2010; accepted June 2011