Formalization of Normal Random Variables

Muhammad Qasim

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Applied Science (Electrical & Computer Engineering)

Concordia University

Montréal, Québec, Canada

April 2016

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Muhammad Qasim

Entitled: "Formalization of Normal Random Variables"

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science

Complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

______________________________________ Chair
Dr. R. Raut

______________________________________ Examiner, External
Dr. N. Shiri (CSE) To the Program

______________________________________ Examiner
Dr. S. Hashtrudi Zad

______________________________________ Supervisor
Dr. S. Tahar

Approved by: ______________________________________
Dr. W. E. Lynch, Chair
Department of Electrical and Computer Engineering

____________20_____ ______________________________________
Dr. Amir Asif, Dean
Faculty of Engineering and Computer Science

Abstract

Formalization of Normal Random Variables

Muhammad Qasim

Engineering systems often have components that exhibit random behavior. This randomness in many cases is normally distributed. To verify such systems, probabilistic analysis is used. Such engineering systems have applications in domains like transportation, medicine and military. Despite the safety-critical nature of these applications, most of the analysis is done using informal techniques like simulation and paper-and-pencil analysis, and thus cannot be completely relied upon. The unreliable results produced by such methods may result in heavy financial loss or even the loss of a human life. To overcome the limitation of traditional methods, we propose to conduct the analysis of such systems within the trusted kernel of a higher-order-logic theorem prover HOL4. The soundness and the deduction style of the theorem prover guarantee the validity of the analysis and the results of this type of analysis are generic and valid for any instance of the system. For this purpose, we provide HOL4 formalization of Lebesgue measure and normal random variables along with the proof of their classical properties. We also ported the theory of Gauge integral and other required foundational concepts from HOL Light and Isabelle/HOL theorem provers. To illustrate the usefulness of our formalization, we conducted the formal analysis of two applications, i.e., error probability of binary transmission in the presence of Gaussian noise and probabilistic clock synchronization in wireless sensor networks.

To My Parents

Acknowledgments

Firstly, I would like to thank Dr. Sofiene Tahar for his help, guidance and encouragement throughout my Master's degree. Other than research, I have learned many practical and professional aspects from him. I am very grateful to Dr. Osman Hasan for his support in my research, sound advice, insightful criticisms, prompt feedback and encouragement.

Many thanks to Muhammad Umair Siddique, Maissa Elleuch and Donia Chaouch for their support in my research. I would like to thank Muhammad Shirjeel who helped me in reading and correcting the initial drafts of this thesis. My sincere thanks to all my friends in the Hardware Verification Group for their support and motivation, though I do not list all their names here.

Finally, I would like to thank my family for their perpetual love and encouragement. Without the love and support of my parents nothing would have been possible, they have provided me with countless opportunities for which I am eternally grateful. I would like to thank my wife for her support and encouragement and my siblings for their love and affection.

Contents

List of Figures

List of Tables

List of Acronyms

CAS	Computer Algebra System
CDF	Cumulative Distribution Function
HOL	Higher-Order Logic
FOL	First-Order Logic
ML	Meta Language
PDF	Probability Density Function
PRISM	PRobabilistIc Symbolic Model checker
SMC	Statistical Model Checking
SML	Standard Meta-Language
HK	Henstock Kurzweil
RBS	Reference Broadcast System
WSN	Wireless Sensor Network
TDMA	Time Division Multiple Access

Chapter 1

Introduction

1.1 Motivation

Modern world engineering systems need to interact with their environments. This interaction often becomes a cause of randomness and therefore the design and analysis of such systems becomes more challenging. Other causes of randomness in engineering systems include the aging phenomena of hardware components and the execution of certain actions based on a probabilistic choice in randomized algorithms. While random events are by definition unpredictable, it is often possible to predict the frequency of different outcomes over a large number of events. Usually, probabilistic approaches are used to analyze such systems. The main idea is to quantify randomness and assign a measure of probability to its frequency distribution. The quantification of randomness is done by a random variable, i.e., a function that will map randomness to a suitable set of numbers. Using the probabilistic properties of these random variables and mathematically modelling the unpredictable component of a given system along with its environment, we can judge the parameters of interest and provide the likelihood of them satisfying given specifications.

In many cases the randomness in engineering systems is *normally distributed.*

The noise in communication channel, length and weight of manufactured goods, message arrival time in wireless sensor network, blood pressure reading of a general population, lifetime of an electric bulb and maximum speed of a particular car model are some of the examples. It was in the beginning of the 19th century that Carl Friedrich Gauss, a German mathematician, presented the fundamentals of normal distribution and hence this distribution was also known as Gaussian distribution. Later Karl Pearson, an English mathematician, popularized the term normal as a designation for this distribution.

> *Many years ago I called the Laplace Gaussian curve the normal curve, which name, while it avoids an international question of priority, has the disadvantage of leading people to believe that all other distributions of frequency are in one sense or another 'abnormal'.*
>
> Karl Pearson, Notes on the History of Correlation (1920)

The importance of normal distribution is much more evident with the central limit theorem [8]. It states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterations of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution [44]. Therefore, if the sample size is large enough, the sample mean of other distributions may also be treated as normal.

Traditionally, paper-and-pencil based approaches are used for carrying out probabilistic analysis. This method however being prone to human error, fails when it comes to complex and large systems. Another widely used technique for the same purpose is simulation, which is quite efficient in many cases. However, it provides less accurate results due to approximations in numerical computations and might not be feasible for large applications due to enormous processing time requirements. Due to the usage of engineering systems in safety-critical applications, these methods cannot be relied upon.

Formal methods, which provide computerized mathematical proofs, overcome above mentioned limitation by providing accurate analysis and eliminating human error. The main idea is to develop a mathematical model and formally verify that it meets the required specifications. The two most widely used formal methods are model checking and theorem proving [23]. Model checking is an automatic verification approach for systems that can be expressed as a finite-state machine but its usage for probabilistic analysis is somewhat limited due to restricted system expressiveness. On the other hand, theorem proving is an interactive technique but is more powerful in terms of expressiveness.

An error in a system that is used in safety-critical domains may result in heavy financial loss or even the loss of human life. Therefore, it is imperative to verify such systems formally. This thesis presents the mathematical foundations that will allow the verification of systems that are used in safety and mission critical domains and exhibit normally distributed randomness. We use the HOL4 theorem prover [48] for the above mentioned formalization and verification tasks. The main motivation behind this choice is to build upon existing formalizations of measure, Lebesgue integration and probability theory in HOL4.

1.2 State of the Art

Systems exhibiting random behavior whether distributed normally or otherwise, are traditionally analyzed using paper-and-pencil based approaches. Such analysis is always prone to human error and so cannot guarantee the accuracy of present world complex engineering and scientific systems. Also, it is often the case that many key assumptions in the results obtained are in the mind of the engineer or scientist and not documented. Such missing assumptions may also lead to erroneous system design. Computer based analysis techniques are a good alternative to traditional approaches and are capable of analyzing large and complex systems with better accuracy.

1.2.1 Simulation

Simulation is one of the most widely used computer based probabilistic analysis technique. To apply this technique, a system model with random components is created and then analyzed by taking a large number of samples to approximate the values of desired parameters. Using the simulation results, predictions may be made about the behavior of the system. Many simulation softwares have been developed. MATLAB [33], Minitab [38], SPSS [49], SAS [46] and mathStatica [32] are some of the examples. All of them contain a large collection of discrete, continuous and multivariate distributions which can be used to model systems with random or unpredictable components. Simulation techniques are however, less accurate as they never exactly imitate real world systems. This inaccuracy comes mainly from the following two reasons:

1. Rounding-off and truncation of numeric values when performing computation in the computer due to the finite precision representation of numbers.

2. Use of heuristics and algorithms to approximate the result in order to reduce the huge processing time required when analyzing large systems.

Due to the limitations of accuracy, simulation cannot provide a reliable measure of confidence towards the satisfaction of design requirements of critical systems. Another limitation of simulation based probabilistic analysis is the enormous processing time required to attain meaningful results. Generally, there is a trade-off between speed and accuracy.

1.2.2 Computer Algebra Systems

To conduct a probabilistic analysis of engineering systems with better accuracy one may consider *computer algebra systems* (CAS) as a good alternative to simulation techniques [17]. These systems manipulate mathematical symbols in a way that is similar to traditional manual computations of mathematicians and scientists. Mathematica [31] and Maple [30] are the most popular CAS systems available today. They provide toolboxes with a variety of options that may be used to model and analyze probabilistic systems with different discrete and continuous distributions including normal distribution. Maxima [35] is another CAS system that supports probabilistic analysis but is limited to univariate distributions.

In some cases, computer algebra systems would use numerical computations and thus compromise on the accuracy of the result for the same reason mentioned for simulation based techniques. Also, these systems are not strictly logical as they neglect basic assumptions in certain cases. For example, Mathematica returns 1 as the answer when given x/x as the input, while $x/x = 1$ holds only when $x \neq 0$. Another serious analysis issue is caused by the use of complex symbolic manipulation algorithms, which have not been verified [13].

1.2.3 Probabilistic Model Checking

Model checking is one of the most widely used formal analysis technique. The main idea is to develop a precise state based mathematical model and specify system properties using temporal logic [6]. The model is then subjected to exhaustive analysis to verify if it satisfies the given formally represented properties. It provides strictly logical proofs and therefore overcomes the above mentioned limitations. Numerous probabilistic model checking algorithms and methodologies have been proposed e.g., [12, 42], and based on these algorithms, a number of tools have been developed, e.g., PRISM [43] and VESTA [1].

Besides the accuracy of the results, the most promising feature of probabilistic model checking is the ability to perform the analysis automatically. On the other hand, it is limited to systems that can be modelled as finite state machines and may also suffer from state space explosion. Also, the computed expected values are expressed in a computer based notation, such as fixed or floating point numbers, which also introduces some degree of approximation in the results.

1.2.4 Theorem Proving

Theorem proving is another widely used formal analysis technique. Unlike model checking, theorem proving is not limited to the size of the state space and therefore, can be used to analyse large systems. Also, the underlying logic of theorem provers (first-order or higher-order logic) provides a high level of expressiveness, which allows the analysis of a wider range of systems without any modeling limitations. The most widely used theorem provers are HOL4 [48], HOL Light [29], Coq [11] and Isabelle/HOL [26].

Over the past decade, many foundational mathematical theories have been formalized. Hurd [25] developed a probability theory and formalized the measure space as a pair (Σ, μ) in the HOL theorem prover [20], where Σ is a set of measurable sets and μ is a measure on sets belonging to Σ. However, in this formalization the space is implicitly the universal set of appropriate data-type. Hasan [22] built upon Hurd's work and formalized statistical properties of both discrete and continuous random variables and their *Cumulative Distribution Function* (CDF) in the HOL4 theorem prover [48]. However, Hasan's work inherits the same limitations as of Hurd. As a consequence, when the space is not the universal set, the definition of the arbitrary space becomes very complex. Later, Coble [10] defined probability space and random variables based on an enhanced formalization of measure space which is the triplet (X, Σ, μ), where X is a sample space, Σ is a set of measurable subsets of X and μ is a measure on sets belonging to Σ. This measure space overcomes the disadvantage of Hurd's work since it contains an arbitrary space. Coble's probability theory is built upon finite-valued (standard real numbers) measures and function. Specifically, the Borel Sigma spaces [50] cannot be defined on open intervals which constrains the verification of some applications. More recently, Mhamdi [36] used the axiomatic definition of probability proposed by Kolmogorov [27] to provide a significant formalization of both measure and probability theory for formally analyzing information theory in HOL4. His work overcomes the limitations of the above mentioned work by allowing the definition of sigma-finite and other infinite measures as well as the signed measures. Affeldt [2] simplified the formalization of probability theory in Coq [11]. Holzl [24] has also formalized three chapters of measure theory in Isabelle/HOL [39] and building on top of it, formalized some very useful notions of probability theory like probability mass function, independent random variables and convolution [4, 14]. Most recently, the central limit theorem has been proved by Avigad et al. [5], who also formalized the characteristic function of random variables [4] in Isabelle/HOL [26].

Table 1 provides a comparison of paper and pencil analysis, simulation, computer algebra systems, probabilistic model checking and theorem proving based on the following attributes.

1. Expressiveness: ability to describe complex mathematical models.

2. Scalability: can be used for larger systems.

3. Accuracy: provides precise results.

4. Efforts: ease of use.

5. Coverage: results are valid for every possible input.

Table 1: Theorem Proving Compared to Other Approaches

Criteria	Paper and Pencil Proofs	Simulation	Computer Algebra Systems	Probabilistic Model Checking	Theorem Proving
Expressive	++	+	-	- -	++
Scalable	-	++	+	-	+
Accuracy	++	- -	+	++	++
Efforts	- -	++	++	++	- -
Coverage	++	- -	++	++	++

Paper and Pencil Proofs are not scalable because they are prone to human error and a lot of efforts are required to analyze large systems. Simulation compromises on accuracy and does not provide 100% coverage. Computer Algebra Systems are limited to expressions that can be solved automatically and use approximations in certain cases. Probabilistic Model Checking is limited to systems that can be expressed as finite state machines and is not scalable because of state space explosion problem. While Theorem Proving is accurate and scalable, it requires a considerable amount of manual efforts. Also, Theorem Proving uses mainly higher-order logic, which is more expressive compared to the propositional temporal logic used in Model Checking and offers more capability than first-order logic that is used in automated provers. For example, propositional logic and first-order logic cannot be used for the analysis of systems that involve multivariate or complex calculus.

1.3 Thesis Contribution

In this thesis we present the formalization of normal random variable along with the required foundational theories. This will allow us to conduct the formal probabilistic analysis of real world systems that exhibit normally distributed randomness and also facilitate the formalization of some other kinds of random variables. As normal distribution is continuous, we first need to develop the means to reason about the properties of a continuous distribution. Usually such analysis would involve Lebesgue integration and the concepts of probability theory. Fortunately, they have been formalized in HOL4 by Mhamdi [36].

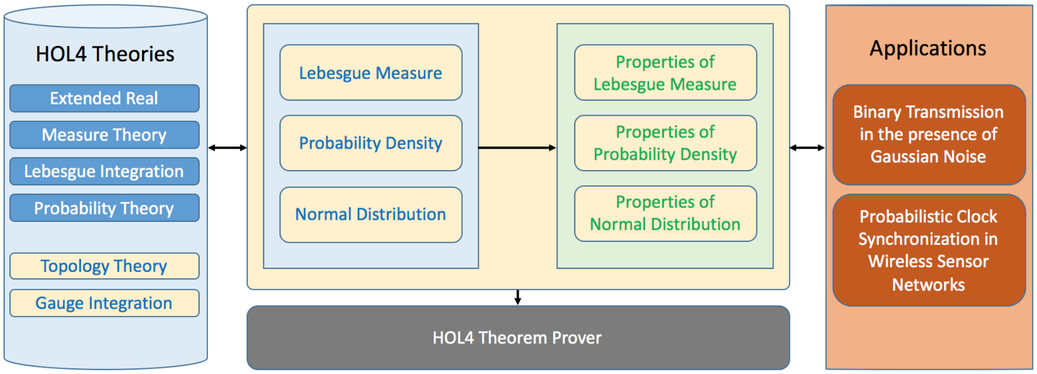
HOL4 Theories
Extended Real
Measure Theory
Lebesgue Integration
Probability Theory
Topology Theory
Gauge Integration
Lebesgue Measure
Probability Density
Normal Distribution
Properties of Lebesgue Measure
Properties of Probability Density
Properties of Normal Distribution
HOL4 Theorem Prover
Applications
Binary Transmission in the presence of Gaussian Noise
Probabilistic Clock Synchronization in Wireless Sensor Networks

Figure 1.1: Overview of the Proposed Framework

A general overview of our formalization is illustrated in Figure 1.1. In order to formalize the distribution of normal random variables, we first provide the formalization of Lebesgue-Borel measure [7] based on the Gauge integral formalization of Harrison [21] in HOL Light. The Lebesgue-Borel measure allows us to evaluate the integral of a measurable function using the Lebesgue integral formalization of Mhamdi [36]. Then, we formalize the probability density function as Radon Nikodym derivative [7] of probability measure with respect to Lebesgue-Borel measure. Finally, we define the normal random variable based on the probability density function and prove some classical properties related to its distribution. These proofs will be of great assistance in the analysis of real-world systems. To demonstrate the usefulness of our formalization, we used it to model and verify the error probability of binary signals transmission in the presence of Gaussian noise [47]. We also used it to analyze the probabilistic bounds on the accuracy of clock synchronization technique in wireless sensor network [41].

1.4 Thesis Organization

The rest of the thesis is organized as follows: In Chapter 2, we provide a brief introduction to the HOL theorem prover and an overview of the formalization of measure, Lebesgue integral and probability theory to equip the reader with some notation and concepts that are going to be used in the rest of this thesis. Chapter 3 describes the formalization of Lebesgue-Borel measure based on Gauge integral, it also presents the formalization of Gauge integral that is ported from HOL Light theorem prover. In Chapter 4, we present the formalization of normal random variables and discuss the proofs of some useful properties. Chapter 5 illustrates the usefulness of our formalization by presenting two example applications. We verify the probability of error in binary transmission systems having Gaussian noise. We also use it for a formal probabilistic analysis of clock synchronization error in wireless sensor network. Finally, Chapter 6 concludes the thesis and outlines some future research directions.

Chapter 2

Preliminaries

In this chapter, we provide a brief introduction to the HOL theorem prover and present an overview of Mhamdi's formalization of measure, Lebesgue integration and probability theory. The intent is to introduce the basic theories along with some notations that are going to be used in the rest of this thesis.

2.1 Theorem Proving

Theorem proving is a technique that is used to verify a system along with its desired properties. To do so, a system model is developed using mathematical logic. One may use propositional logic, first-order logic or higher-order logic depending on the expressibility requirements [3]. To prove that a system model satisfies the desired properties, theorems are developed and proved using inference rules. First-order logic (FOL) [16] can be significantly automated and as a consequence, theorems that are comprised of only FOL, are proved with comparative ease. On the other hand, theorems that require higher-order logic (HOL) [9] entail more efforts to prove as it is difficult to automate HOL proofs due to its undecidable nature. For probabilistic analysis, random variables are formalized as functions that map randomness to an appropriate set of real numbers. Also, the characteristics of random variables, such as PDF and expectation, etc., are formalized by quantifying over random variable functions. Because first-order logic does not allow quantification over predicates and functional variables, we need to use higher-order logic to conduct probabilistic analysis.

Many theorem-proving systems have been implemented and used for all kinds of verification problems. The most popular theorem provers include HOL4 [48], Isabelle/HOL [26], HOL Light [29] and Coq [11]. These systems are distinguished by, among other aspects, the underlying mathematical logic, the way automatic decision procedures are integrated into the system and the user interface. We use the HOL4 theorem prover for our work, mainly because we build on measure, Lebesgue integration and probability theory of Mhamdi [36].

2.2 HOL Theorem Prover

HOL is an interactive theorem prover developed by Mike Gordon at Cambridge University [19]. It is used to carry out mathematical proofs in higher-order logic. HOL4 is based on standard meta language (SML) [37], which is a functional programming language. There are two popular implementations of SML i.e. Moscow ML [45] and Poly ML [34]. Either one can be used for HOL. In the last three decades, there have been several versions of HOL. The first version was called HOL88 followed by HOL90 and HOL98. A successor to them is HOL4, it is mainly based on HOL98 and incorporates ideas and tools from HOL Light. The core of HOL4 consists of only four basic axioms and eight primitive inference rules [40], which are implemented as ML functions. HOL4 has been widely used for the formal verification of software and hardware systems along with the formalization of pure mathematical theories.

There are four types of terms in HOL4, i.e., variables, constants, function applications, and lambda-abstractions. Variables are sequences of digits or letters beginning with a letter. The syntax of the constants is similar to that of variables, but they cannot be bounded by quantifiers. Function applications in HOL represent the evaluation of a function for a certain argument and lambda abstractions are used to denote a function f that takes x as argument and returns $f(x)$. Every variable or constant in HOL should be given a type. In our work, we are mostly using boolean, (natural) number, real and extended real types denoted by bool, num, real and extreal.

A mathematical proof is in the form of a theorem, that is a combination of terms. A type checking algorithm is implemented in HOL to infer the type of terms when a theorem is created. In certain cases, when a term may have overloaded types, it is required to mention the type explicitly. Normally a theorem consists of a premise and a conclusion. To prove the theorem within the HOL environment, inference rules are applied until one reaches the conclusion. The eight primitive inference rules in HOL are Assumption introduction, Reflexivity, Beta-conversion, Substitution, Abstraction, Type instantiation, Discharging an assumption and Modus Ponens [40]. All other rules are derived from these inference rules and axioms.

A collection of axioms, definitions and proven theorems is called a *theory*. To prove a theorem, it can be subdivided into a list of smaller theorems and a subset of that list may be solved by reusing already proven theorems present in HOL4 theories. HOL theories are organized in a hierarchical fashion, i.e., a theory can have other theories as parents. In fact, one of the primary motivations of selecting the HOL theorem prover for our work was to benefit from the theories already present. The starting theory of HOL is the *min* thoery. In this theory, the type constant for booleans, the binary type operator for functions, and the type constant for individuals are declared. Building on these types, three primitive constants are declared: equality, implication, and a choice operator [40]. There is no theorem or axiom in this theory. The most basic thoery of HOL is the *bool* theory. It is a parent for all other theories and contains the four axioms of HOL. These axioms together with the eight inference rules are sufficient for developing all of standard mathematics.

The following table provides the mathematical interpretations of some frequent HOL symbols and functions used in this thesis.

Table 2: HOL Symbols and Functions

HOL Symbol	**Meaning**
$\wedge$	Logical *and*
$\vee$	Logical *or*
$\neg$	Logical *negation*
$\Rightarrow$	Logical implication
$=$	Logical equality
$\forall$x.f	for all x : f
$\exists$x.f	for some x : f
@x.t(x)	Some x such that t(x) is true
(&n : num)	type casting (&n:extended real)
x pow n	x^n
λx.f	Function that maps x to $f(x)$
{x|P(x)}	Set of all x such that $P(x)$
UNIV	Universal Set
DISJOINT A B	Sets A and B are disjoint
IMAGE f A	Set with elements $f(x)$ for all $x \in A$
PREIMAGE f B	Set with elements $x \in B$ for all $f(x) \in B$
{}	Empty Set
FINITE S	S is a finite set
sup S	Supremum of the set S
SUC n	Successor of natural number
exp x	Exponential function
max a b	if a $\leq$ b then b else a
abs x	|x|
dist (a, b)	|a - b|
BIGUNION P	Union of all sets in the set P
BIGINTER P	Intersection of all sets in the set P
suminf(λn.f(n))	$\lim\limits_{k \to \infty} \sum_{n=0}^{k} f(n)$
SIGMA(λn.f(n))s	$\sum_{n \in s} f(n)$
integral i f	$\int_{x \in i} fx \ \mathrm{d}x$

2.3 Measure Theory

A measure is used to assign a number to a set, that will represent its size. It is a generalized concept of length, area, volume, etc. Two important examples are the Lebesgue measure, i.e., a standard way of assigning a measure to the measurable subsets of Euclidean space and the probability measure, i.e., a measure defined on a set of events for assigning it a value to indicate its probability of occurrence. Formally, a function defined on a set X is called a measure if it satisfies the following properties:

1. Positive: Measure of every set belonging to sigma algebra over X is positive and the measure of an empty set is zero. A sigma algebra over X is a collection of subsets of X that includes the empty set, is closed under compliment and is closed under union or intersection of countably many subsets.

2. Countably additive: Measure of the union of a collection of disjoint sets is equal to the sum of their measures.

In HOL, a measure is represented by a triplet $(X, \mathcal{A}, \mu)$ where X is a sample space, $\mathcal{A}$ is a sigma algebra over X and μ is a measure on sets belonging to $\mathcal{A}$. The triplet is called a measure space if it satisfies above mentioned properties. The notion of sigma algebra is required in order to define a predicate for measure space.

Definition 2.1. *(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 $\{\}$, and is closed under countable unions and complementation within the space X.*

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

The predicate `subset_class` is used to test whether $\mathcal{A}$ is a collection of subsets (or subset class) of X and `countable` will test if there exists a surjective function $f : \mathbb{N} \to s$ such that every element of the set s can be associated with a natural number. The HOL formalization of `subset_class` and `countable` are as follows:

```
⊢ subset_class X A = ∀s. s ∈ A ⇒ s ⊆ X
⊢ countable s = ∃f. ∀x. x ∈ s ⇒ ∃(n:num). f n = x
```

For any collection G of subsets of X, there is at least one sigma algebra on X containing G, namely the powerset of X. The smallest sigma algebra on X containing G is an intersection of all those sigma algebras, and is called the sigma algebra on X generated by G. This notion is defined in HOL as:

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

The `sigma` function will be used in the next chapter to define a sigma algebra on Borel space (or Borel sigma algebra). Now the predicate for measure space can be defined using the definition of sigma algebra.

Definition 2.2. *(Measure Space)*
A triplet $(X, \mathcal{A}, \mu)$ *is a measure space* iff $(X, \mathcal{A})$ *is a measurable space and* $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ *(i.e., extended real numbers) is a non-negative and countably additive measure function.*

```
⊢ measure_space (X,A,μ) =
    sigma_algebra (X,A) ∧ positive (X,A,μ) ∧
    countably_additive (X,A,μ)
```

where the positive and countably additive properties are defined as:

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

```
⊢ positive (X,A,μ) =
    μ {} = 0 ∧ ∀s. s ∈ A ⇒ 0 ≤ μ s
```

The pair $(X, \mathcal{A})$ is called a σ-field or a measurable space and $\mathcal{A}$ is called a sigma algebra over X or a set of measurable sets. Mhamdi defined some auxiliary functions that will take a σ-field or a measure space and return individual components.

```
⊢ space (X,A) = X
⊢ subsets (X,A) = A
⊢ m_space (X,A,μ) = X
⊢ measurable_sets (X,A,μ) = A
⊢ measure (X,A,μ) = μ
```

There is a special class of functions called measurable functions. For these functions the inverse image of each measurable set is also measurable. Measurable functions are used in probability theory to define random variables.

Definition 2.3. *(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 ∈ 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
```

2.4 Lebesgue Integration Theory

Integration is the reverse of differentiation operation and is a fundamental concept in mathematics. There are many ways of formally defining an integral, distinguished by the ability to handle differing special cases which may not be integrable under other definitions. The two most used integrals that have also been formalized in HOL4 are *Riemann Integral* and *Lebesgue Integral*. In this thesis we use the Lebesgue integral because of the following reasons:

1. The Riemann integral is limited to intervals of real line.

2. The Lebesgue integration can handle a broader class of functions.

3. The Probability theory in HOL4 is based on Lebesgue integration.

Lebesgue integral was formalized by Mhamdi [36]. He proved many useful theorems including monotone convergence and Radon Nikodym Theorem but to apply his formalization to evaluate an integral, a suitable Lebesgue measure is required. We describe one such measure in the next chapter. Below we present the basic definitions of the Lebesgue integration theory that we have used in our formalization of normal random variables.

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. In HOL a positive simple function g is represented by the triplet (s, a, α) as a finite linear combination of indicator functions of measurable sets (a_i) that form a partition of the space X.

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

where s is a set of partition tags, a_i is a sequence of measurable sets, α_i is a sequence of real numbers and I_{a_i} is an indicator function on a_i. The indicator function is defined in HOL as:

```
⊢  indicator_fn A = (λx. if x ∈ A then 1 else 0)
```

The Lebesgue integral is first defined for positive simple functions and then extended to non-negative functions.

Definition 2.4. *(Lebesgue Integral of Positive Simple Functions)*
Let $(X, \mathcal{A}, \mu)$ be a measure space. The integral of the positive simple function g with respect to the measure μ is defined as

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

This is formalized in HOL as:

```
⊢  pos_simple_fn_integral m s a α =
     SIGMA (λi. α i * measure m (a i)) s
```

Definition 2.5. *(Lebesgue Integral of Non-Negative Measurable Functions)*
The definition of the Lebesgue integral of positive simple functions is used to define the integral of non-negative measurable functions using the supremum operator as follows:

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

Its formalization in HOL is the following:

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

where `psfis m g` represents the Lebesgue integral of the positive simple function g.

2.5 Probability Theory

The classical definition for probability that is prevailing for many centuries is given as the ratio of the number of favorable outcomes to the number of all possible outcomes. This definition is based on the assumption that all outcomes are equiprobable and that the number of possible outcomes is finite. These limitations were overcome by the axiomatic definition of probability by Kalmogorov. It provides a mathematically consistent way for assigning and deducing probabilities of events. Inspired from the work of Kalmogorov [27], Mhamdi has formalized probability in HOL4 [36]. He defined the probability space as a measure space, i.e., (Ω, F, p), where Ω is a set of all possible outcomes, called the sample space, F is a set of events and p is the probability measure. A probability theory is developed based on the following three axioms:

1. $\forall A.\ 0 \leq Pr(A)$
2. $Pr(\Omega) = 1$
3. For any countable collection $A0$, $A1$,... of mutually exclusive events, $Pr(\bigcap_{i \in \Omega} A_i) \ = \ \sum_{i \in \Omega} Pr(A_i)$

Below we present the basic definitions of probability thoery.

Definition 2.6. *(Probability Space)*
(Ω, F, p) is a probability space if 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.

```
⊢ prob = measure
⊢ events = measurable_sets
⊢ p_space = m_space
```

Definition 2.7. *(Random Variable)*
$X : \Omega \to \overline{\mathbb{R}}$ is a random variable iff X is $(F, \mathcal{B}(\overline{\mathbb{R}}))$ measurable

where F denotes 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 the general definition of the Borel sigma algebra.

```
⊢ random_variable X p s ⇔
    prob_space p ∧ X ∈ measurable (p_space p,events p) s
```

Definition 2.8. *(Probability Distribution)*
The probability distribution of a random variable X is defined as the function assigning to A the probability of the event $\{X \in A\}$.

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

```
⊢ distribution p X = (λA. prob p (PREIMAGE X A ∩ p_space p))
```

Because `PREIMAGE X A` does not exclude values that do not belong to the probability space p. Therefore, it is intersected with `p_space p`.

Joint distribution and conditional probability are two very useful notions in probability thoery. The joint distribution of two random variables X and Y assigns a probability to the event $\{(X, Y) \in a\}$ and conditional probability gives the distribution of the random variable X given the distribution of the random variable Y. These notions are required in the analysis of many engineering systems as in one of the applications presented in this thesis. Mhamdi formalized these notions in HOL as [36]:

```
⊢ joint_distribution p X Y =
          (λa. prob p (PREIMAGE (λx. (X x,Y x)) a ∩ p_space p))
```

```
⊢ conditional_distribution p X Y =
      (λa b. joint_distribution p X Y (a  b) / distribution p Y b)
```

Chapter 3

Lebesgue-Borel Measure

This chapter presents the formalization of Lebesgue-Borel measure. First, we present the formalization of Gauge integral that we ported from the HOL Light theorem prover and then use it to define a Lebesgue measure. Then, we restrict this measure to Borel measurable sets and call it a Lebesgue-Borel measure. This is because the extended real valued Borel measurable sets are not compatible with the real valued gauge integral. Therefore, we also formalize the real valued borel measurable sets and prove some useful properties required for our work.

3.1 Gauge Integral

Gauge integral, also known as the Henstock Kurzweil (HK) integral [51], is a generalization of Riemann integral and is suitable for a wider class of functions. The Riemann integral formalized in HOL4 is defined on intervals. In contrast, the Gauge integral is defined on a set that may or may not be an interval. Also, it allows the real line to be represented by a universal set. In this section, we discuss the formalization of Gauge integral that we ported from the HOL Light theorem prover and is originally formalized by Harrison [21]. Below, we present the definitions of Gauge integral along with all of its required components.

Definition 3.1. *(Open Set)*

A set s is called open if, given any point $x \in s$, there exists a real number $\epsilon > 0$ such that, given any point $y \in \mathbb{R}$ whose distance from x is smaller than ϵ, $y \in s$.

```
⊢ open s = ∀x. x ∈ s ⇒ ∃ϵ. ϵ > 0 ∧ ∀y. dist (y,x) < ϵ ⇒ y ∈ s
```

Definition 3.2. *(Tagged Partial Division)*

Tagged partial division of a set X is a set of nonempty subsets of X such that every element $x \in X$ is in exactly one of these subsets.

```
⊢ s tagged_partial_division_of X =
      FINITE s ∧
      ∀x k. (x,k) ∈ s ⇒ x ∈ k ∧ k ⊆ X ∧ ∃a b. k = interval [a,b] ∧
      ∀x1 k1 x2 k2. (x1,k1) ∈ s ∧ (x2,k2) ∈ s ∧ (x1,k1) ≠ (x2,k2)
         ⇒ interior k1 ∩ interior k2 = {}
```

where interior is used to exclude the intersection points and is defined in HOL as:

```
⊢ interior s = {x | ∃t. open t ∧ x ∈ t ∧ t SUBSET s}
```

Definition 3.3. *(Tagged Division)*

Tagged partial division of X is a tagged division when X is the union of all partitions.

```
⊢  s tagged_division_of X =
     s tagged_partial_division_of X ∧
     (BIGUNION {k | ∃x. (x,k) ∈ s} = X)
```

Definition 3.4. *(Gauge)*

A function is called a gauge if for every tag argument, it returns an open interval.

```
⊢  gauge d = ∀x. x ∈ d x ⇒ open (d x)
```

Definition 3.5. *(Fineness Property)*
A tagged division s is called δ-fine with respect to a Gauge function d if every partition of the division is a subset of (d x)

```
⊢ d fine s = ∀x k. (x,k) ∈ s ⇒ k ⊆ d x
```

Definition 3.6. *(Riemann Sum)*
Riemann sum is an approximate sum of areas of all partitions in a tagged division p.

```
⊢ sum p (λ(x,k). content k * f x)
```

where x denotes a tag, k denotes a partition and the `content` of partition k is defined in HOL as the difference of its supremum and infimum:

```
⊢ content k = if k = {} then 0 else (sup k - inf k)
```

All notions defined above are required to formalize the Gauge integral. While Harrison defined it for many dimensions, we could only define it for a single dimension ($\mathbb{R}$) due to the unavailability of multivariate theory in HOL4. However, for our work, it will suffice.

Definition 3.7. *(Gauge Integral)*
Let f:[a,b]$\rightarrow\mathbb{R}$ be some function, and let y be some number. We say that y is the Gauge integral of f over i written $y = \int_i f(x)\,\mathrm{d}\,x$, if for each number $e > 0$ there exists a Gauge d such that $|\sum_p f$ - $y| < e$, where, p is a tagged division of i and p is δ-fine with respect to p.

```
⊢ (f has_integral_compact_interval y) i =
    ∀e. 0 < e ⇒ ∃d. gauge d ∧
          ∀p. p tagged division of i ∧ d fine p ⇒
          abs (sum p (λ(x,k). content (k) * f(x)) - y) < e
```

An alternate definition of Gauge integral that simplifies the proof steps for integration over intervals is given as:

```
⊢ (f has_integral y) i =
    if ∃a b. i = interval [a,b] then (f has_integral_compact_interval y) i
    else ∀e. 0 < e ⇒ ∃B. 0 < B ∧
          ∀a b. ball (0,B) SUBSET interval [a,b] ⇒
          ∃z. ((λx. if x ∈ i then f x else 0)
             has_integral_compact_interval z) (interval [a,b]) ∧
          abs (z - y) < e
```

Also, a simplified definition using the Hilbert choice operator (`@`), where `@`$x.t(x) =$ some x such that $t(x)$ is true, is as follows:

```
⊢ integral i f = @y. (f has_integral y) i
```

The Gauge integral theory in HOL Light contains a lot of useful properties for the Gauge integral. We only used a few of them for our formalization of Lebesgue-Borel measure. These are given below.

Theorem 3.1. *Integral Addition of Two Functions.*

```
⊢ ∀f g s. f integrable_on s ∧ g integrable_on s ⇒
   integral s (λx. f x + g x) = integral s f + integral s g
```

Theorem 3.2. *If $s \subset t \wedge \forall x.\ 0 \leq f\,x$ then, integral over $s \leq$ integral over t.*

```
⊢ ∀f s t. s SUBSET t ∧ f integrable_on s ∧ f integrable_on t ∧
   (∀x. x ∈ t ⇒ 0 ≤ f x) ⇒
   integral s f ≤ integral t f
```

Theorem 3.3. *If $\forall x.\ g\,x \leq f\,x$ then, integral of $g \leq$ integral of f.*

```
⊢ ∀f g s. f integrable_on s ∧ g integrable_on s ∧
   (∀x. x ∈ s ⇒ f x ≤ g x) ⇒
   integral s f ≤ integral s g
```

Theorem 3.4. *Integral of a Function is Unique*

```
⊢ ∀f i k1 k2. f has_integral k1 ∧ f has_integral k2 ⇒ (k1 = k2)
```

Theorem 3.5. *Bounds on Integral over an Interval.*

```
⊢ ∀f a b. f integrable_on interval [a,b] ∧
   (∀x. x ∈ interval [a,b] ⇒ f x ≤ B) ⇒
   integral (interval [a,b]) f ≤ B * content (interval [a,b])
```

Theorem 3.6. *Integral of a Positive Function is Non-Negative.*

```
⊢ ∀f s. f integrable_on s ∧
   (∀x. x ∈ interval [a,b] ⇒ 0 ≤ f x) ⇒
   0 ≤ integral (interval [a,b]) f
```

Theorem 3.7. *Integral over a Singleton is 0.*

```
⊢ ∀f a. integral (interval [a,a]) f = 0
```

Theorem 3.8. *Integral of a Zero Function is 0.*

```
⊢ ∀s. integral s (λx. 0) = 0
```

Another very useful property for the Gauge integral is the Fundamental Theorem of Calculus [51].

Theorem 3.9. *If f is a real valued continuous function on an interval [a,b] and f' is an antiderivative of f in [a,b], then*

$$\int_a^b f(t), dt = f'(b) - f'(a)$$

```
⊢ ∀f f' a b. a ≤ b ∧ f continuous_on interval [a,b] ∧
   (∀x. x ∈ interval [a,b] ⇒
     (f has_derivative f'(x)) (at x within interval [a,b])) ⇒
   (f' has_integral (f(b) - f(a))) (interval [a,b])
```

where `continuous_on` and `has_derivative` are defined in HOL as:

```
⊢ f continuous_on s =
   (∀x. x ∈ s ⇒ ∀e. 0 < e ⇒
     ∃d. 0 < d ∧ !x. x’ ∈ s ∧ dist (x’,x) < d ⇒
         dist (f(x’),f(x)) < e
```

```
⊢ (f has_derivative x’) net =
   linear (λx. x * x’) ∧
   ((λy. inv (abs (y = netlimit net)) * (f(y) -
      (f(netlimit net) + (λx. x * x’) (y - netlimit net)))) ⟶ 0) net
```

Here, `dist(x’,x) = abs(x’ - x)` and the notions `linear`, `netlimit` and tends to (⟶) are given as:

```
⊢ linear f =
   (∀x y. f(x + y) = f(x) + f(y)) ∧
   (∀c x. f(c * x) = c * f(x))
```

```
⊢ netlimit net = @a. ∀x. ¬(netord net x a)
```

```
⊢ (f ⟶ l) net =
   ∀e. 0 < e ⇒ eventually (λx. dist(f(x), l) < e) net
```

where `netord net` denotes a net (i.e., a generalization of a sequence) and `eventually (λx. dist(f(x), l) < e) net`, when `net = (at x within interval [a,b])`, is equivalent to:

```
⊢ eventually (λx. dist(f(x), l) < e) (at x within interval [a,b]) =
   ∃d. 0 < d ∧ ∀x. x ∈ interval [a,b] ∧
      0 < dist(x,a) ∧ dist(x,a) < d ⇒ dist(f(x), l) < e
```

3.2 Borel Measurable Sets

A collection of all borel measurable sets on $\mathbb{R}$ forms a sigma algebra, called the Borel sigma algebra. It allows us to prove various properties of measurable functions which in our case are the random variables. The borel sigma algebra on $\mathbb{R}$ is defined as the smallest sigma algebra generated by the open sets of $\mathbb{R}$. To formalize this in HOL, we use the `sigma` function defined in Chapter 2 and the definition of open sets presented in previous section.

```
⊢ borel = sigma UNIV {s | open s}
```

where `UNIV` is the universal set of real numbers $\mathbb{R}$. Using above definition, we proved that all open and closed sets are in Borel sigma algebra.

Theorem 3.10. *All open sets of* $\mathbb{R}$ *are in* $\mathcal{B}(\mathbb{R})$*.*

```
⊢ ∀s. {open s} ∈ subsets borel
```

Theorem 3.11. *All closed sets of* $\mathbb{R}$ *are in* $\mathcal{B}(\mathbb{R})$*.*

```
⊢ ∀s. {closed s} ∈ subsets borel
```

where a set is closed if its complement is an open set. It is defined in HOL as:

```
⊢ closed s = open (UNIV DIFF s)
```

Another useful property is that all singleton sets are also Borel measurable.

Theorem 3.12. $\forall c \in \mathbb{R}$, $\{c\} \in \mathcal{B}(\mathbb{R})$

```
⊢ ∀c:real. {c} ∈ subsets borel
```

Mhamdi [36] formalized Borel sigma algebra in the Measure theory as a sigma algebra generated by open intervals of extended real numbers $\overline{\mathbb{R}}$. In order to reuse his proof steps for proving various properties of our Borel sigma algebra generated by open sets of real numbers $\mathbb{R}$, it is required that we prove that our Borel sigma algebra can also be generated by open intervals of real numbers $\mathbb{R}$.

Theorem 3.13. *$\mathcal{B}(\mathbb{R})$ is also generated by open intervals of real numbers.*

```
⊢ borel = sigma UNIV (IMAGE (λ(a,b). interval (a,b)) UNIV)
```

Proof. We start by proving that an open set is equivalent to the union of open intervals denoted by interval (a,b), where a ∈ $\mathbb{Q}$ ∧ b ∈ $\mathbb{Q}$. Then we prove above theorem using the countable union property of sigma algebras and antisymmetric property of sets. □

3.2.1 Real-Valued Borel Measurable Functions

For a function to be integrable over a Borel measurable set, it has to be Borel measurable, i.e., the inverse image of the function belongs to Borel sigma algebra. Theorem 3.13 allowed us to prove some useful properties of Borel sigma algebra following the proof steps of Mhamdi [36].

Theorem 3.14. *If f and g are $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ measurable and $c \in \overline{\mathbb{R}}$ then cf, $|f|$, f^n, $f + g$, $f \cdot g$ 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 ⇒
      ((λx. c * f x) ∈ measurable a Borel)   ∧
      ((λx. abs(f x)) ∈ measurable a Borel)  ∧
      ((λx. f x pow n) ∈ measurable a Borel) ∧
      ((λx. f x + g x) ∈ measurable a Borel) ∧
      ((λx. f x * g x) ∈ measurable a Borel) ∧
      ((λx. max (f x) (g x)) ∈ measurable a borel)
```

Another useful theorem we proved is that all continuous functions are $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\overline{\mathbb{R}}))$ measurable.

Theorem 3.15. *Every continuous functions is* $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\overline{\mathbb{R}}))$ *measurable.*

```
⊢ ∀g. g continuous UNIV(:real) ⇒ g ∈ measurable borel Borel
```

Notice that `borel` is our Borel sigma algebra generated by open sets of real numbers $\mathbb{R}$ and `Borel` is the Borel sigma algebra of Mhamdi [36] generated by open intervals of extended real numbers $\overline{\mathbb{R}}$. This is done for compatibility reasons and will be explained in the next chapter. Using Theorem 3.15 we can prove that if f is measurable then $exp(f)$ is also measurable. This is derived using the equality $(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A))$ and the `MEASURABLE_COMP` theorem from measure theory of Mhamdi [36], i.e.,

```
⊢ ∀f g a b c. f ∈ measurable a b ∧ g ∈ measurable b c ⇒
              g ∘ f ∈ measurable a c
```

Theorem 3.16. *If f is a real-valued Borel measurable function, then Normal (f x) is an extended-real-valued Borel measurable function.*

```
⊢ ∀f m. f ∈ measurable (m_space m, measurable_sets m) borel ⇒
    (λx. Normal (f x)) ∈ measurable (m_space m, measurable_sets m) Borel
```

where `Normal` is used to map real numbers to extended real numbers. Thoerem 3.16 allows us to exploit the properties of both sigma algebras for functions that do not return infinite values.

3.3 Formalization of Lebesgue Measure

Lebesgue measure is a way of assigning a number to subsets of n-dimensional Euclidean space. Sets that can be assigned a Lebesgue measure are called Lebesgue measurable. It is a generalization of the size of a set or length of an interval. Since, we are not working in multi-dimensions, it will assign a number to intervals for representing their length. For an interval [a,b] of a real line, it is simply "b - a". But

Lebesgue measure may also be assigned to more abstract and irregular sets. Lebesgue measure on a Lebesgue measurable set A is denoted by $\lambda(A)$. Formally, it is defined in two steps. First the Lebesgue outer measure is defined as

$$\lambda^* E = inf \ \left(\sum length(i_k)\right) \tag{4}$$

where i_k is a sequence of open intervals with E $\subset \mathbb{R}$. The Lebesgue measure of E is then given by its outer measure if for every A $\subset \mathbb{R}$,

$$\lambda E = \lambda^*(A \cup E) + \lambda^*(A \cup E^c) \tag{5}$$

Our formalization of Lebesgue measure is inspired from Isabelle/HOL [24], it is defined as the supremum of Gauge integrals of X_{a} for all interval [-n,n] (or `line n`), where X_{a} is the indicator function of set A. To keep it compatible with the Measure theory of HOL4, we define it as a triplet by pairing it with the Lebesgue Space ($\mathbb{R}$) and Lebesgue measurable sets, i.e., all sets for which its indicator function is integrable with respect to the inertval [-n,n].

Definition 3.8. *(Lebesgue Measure)*

```
⊢ lebesgue = (univ(:real), {A | ∀n. indicator A integrable_on line n},
    (λA. sup {Normal (integral (line n) (indicator A)) | n IN univ(:real)}))
```

To work with the Measure theory of HOL4, it is required that all measures satisfy the properties of measure space.

Theorem 3.17. *Lebesgue measure is positive and countably additive.*

```
⊢ measure_space lebesgue
```

Most sets that we shall be dealing with are Borel measurable. Therefore, we prove that Borel measurable sets are a subset of Lebesgue measurable sets.

Theorem 3.18. *All Borel measurable sets are also Lebesgue measurable.*

```
⊢ ∀s. s ∈ subsets borel ⇒ s ∈ measurable_sets lebesgue
```

3.4 Formalization of Lebesgue-Borel Measure

A Lebesgue measure assigned to Borel measurable sets is called a Lebesgue-Borel measure. It will be used in later chapters to evaluate integrals of Borel measurable (or Lebesgue measurable) functions. We work with Lebesgue-Borel measure to exploit the available proven properties of Borel sigma algebra and Borel measurable functions. In HOL, we define the triplet of Lebesgue-Borel measure by pairing Lebesgue measure with Borel space and Borel sigma algebra.

Definition 3.9. *(Lebesgue-Borel Measure)*

```
⊢ lborel = (space borel, subsets borel, measure lebesgue)
```

We prove that Lebesgue-Borel satisfies all properties of a measure space and that it is a sigma finite measure.

Theorem 3.19. *Lebesgue-Borel measure is positive and countably additive.*

```
⊢ measure_space lborel
```

Theorem 3.20. *Lebesgue-Borel measure is σ-finite.*

```
⊢ sigma_finite_measure lborel
```

where `sigma_finite_measure` is defined in HOL as:

Definition 3.10. *(Sigma Finite Measure)*
A measure $(X, \mathcal{A}, \mu)$ is called sigma finite if X is the countable union of measurable sets and for every $x \in X$, $\mu(x)$ is a finite number (i.e., $\mu(x) \neq \infty$).

```
⊢ sigma_finite_measure m =
    ∃A. countable A ∧ A SUBSET measurable_sets m ∧
    (BIGUNION A = m_space m) ∧ (∀a. a ∈ A ⇒ (measure m a ≠ PosInf))
```

We also prove that the Lebesgue integral with Lebesgue-Borel measure is affine.

Theorem 3.21. *If f is a measurable function and $c \neq 0$, then*

$$\int_{-\infty}^{\infty} f \, d\mu = |c| * \int_{-\infty}^{\infty} (\lambda x. f(t + c * x)) \, d\mu$$

```
⊢ ∀f c t. c ≠ 0 ∧ f ∈ measurable borel Borel ⇒
    pos_fn_integral lborel (λx. max 0 (f x)) =
    Normal (abs c) * pos_fn_integral lborel (λx. max 0 (f (t + c * x)))
```

Theorem 3.21 is a very useful property of Lebesgue-Borel measure and will be used in the next chapter to prove some important properties of normal random variables.

3.5 Summary

In this chapter we discussed the formalization of Gauge Integral also known as the Henstock Kurzweil integral. Then we defined the Borel sigma algebra and discussed some of its useful properties. We also discussed the limitation of Borel sigma algebra generated from open intervals of extended real numbers and added an axiom to prove some essential properties. Then we presented the formalization of Lebesgue measure and paired it with Borel space and Borel sigma algebra to define Lebesgue-Borel measure. Finally, we presented some theorems for Lebesgue-Borel measure along with the affine transformation property.

Chapter 4

Normal Random Variables

The most common probability distribution in science and engineering is the Gaussian (normal) distribution. In many situations, it is assumed that continuous random variables follow this distribution; in fact, it is so common that it is simply referred to as the normal distribution. Like any other continuous distribution, normal distribution is defined by its probability density function (PDF), given as:

$$N(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{(-\frac{(x-\mu)^2}{2\sigma^2})}$$

where N represents a normal random variable, μ is the mean and σ^2 is the variance of normal distribution. Figure 4.1 gives the probability density functions of normal random variable x with different means and variances.

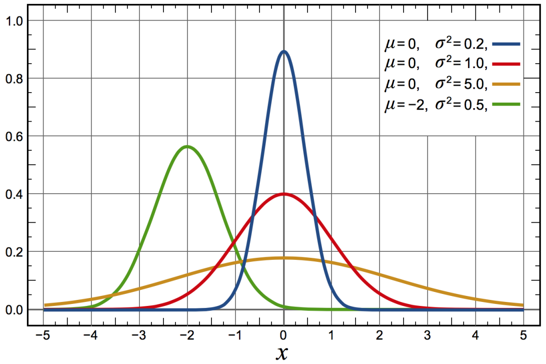

Figure 4.1: Probability Density Function of Normal Random Variable [52]

In this chapter, we provide the formalization of normal random variables. We start by formally defining the PDF as Radon-Nikodym derivative as well as generalizing the Radon-Nikodym theorem for sigma finite measures. Then, we define normal random variables based on their distributions defined by their PDFs. Finally, we present the formal proofs of some useful properties. This formalization will be used to analyze two real world applications in the next chapter.

4.1 Radon Nikodym Theorem

The Radon-Nikodym derivative of a measure ν with respect to the measure μ is defined as a non-negative measurable function f, satisfying the following formula, for any measurable set A [18]:

$$\int_A f \, d\mu = \nu(A) \tag{6}$$

It has been formalized in HOL by Mhamdi [36] as:

```
⊢ RN_deriv m v =
     @f. f IN measurable (X,S) Borel ∧
     ∀x ∈ X, 0 ≤ f x ∧
     ∀a ∈ S, integral m (λx. f x × I_a x) = v a
```

where I_a denotes the indicator function of set a. The existence of the Radon-Nikodym derivative is guaranteed for absolutely continuous measures by the Radon-Nikodym theorem stating that if ν is absolutely continuous with respect to μ, then there exists a non-negative measurable function f satisfying Equation (6) for any measurable set A.

Mhamdi [36] proved Radon Nikodym theorem for finite measure. In the next section we will define the probability density function as a Radon Nikodym derivative of probability measure with respect to Lebesgue-Borel measure. Since the Lebesgue-Borel measure is not finite, it is required to generalize the Radon-Nikodym theorem for sigma finite measures.

Theorem 4.1. *Given a measurable space (X,S), if a measure ν on (X,S) is absolutely continuous with respect to a sigma-finite measure μ on (X,S), then there is a measurable function f, such that for any measurable subset $A \subset X$:*

$$\int_A f \, d\mu = \nu(A)$$

```
⊢ ∀u v X S.
    sigma_finite_measure (X,S,u) ∧
    measure_space (X,S,u) ∧
    measure_space (X,S,v) ∧
    measure_absolutely_continuous (X,S,u) (X,S,v) ⇒
      ∃f. f ∈ measurable (X,S) Borel ∧
      ∀x ∈ X, 0 ≤ f x ∧
      ∀a ∈ S, pos_fn_integral u (λx. f x × I_a x) = v a
```

where `measure_absolutely_continuous` is define in HOL by Mhamdi [36] as:

Definition 4.1. *(Absolutely Continuous Measures)*
If u and v are two measures on a measure space (X,S), then v is absolutely continuous with respect to u if v(A) = 0 for any $A \in S$ such that u(A) = 0.

```
⊢ ∀u v. measure_absolutely_continuous u v =
     ∀A. A ∈ measurable_sets u ∧ (measure v A = 0) ⇒
     (measure u A = 0)
```

Proof. We start by proving the existence of a finite integrable function h for sigma finite measures. We split the space into finite sets and rest (i.e., `UNIV` - finite spaces), then prove the existence of a function f such that it satisfies Equation (6) for finite μ and infinite ν. Finally, we prove that a function $g = (h \cdot f)$ exists even when μ is sigma-finite. Since g may also take infinite values, the axiom (singleton of infinity) was required to prove that g is Borel measurable. □

4.2 Probability Density Function

One way to define the distribution of a continuous random variable is the probability density function (PDF). It is analogous to the probability mass function of discrete random variables, which gives the probability of a discrete random variable acquiring a certain value. Since a continuous variable can have infinitely many values, its probability to have a certain value is zero. Instead, we determine the probability of acquiring values in a certain range. In order to find the probability that a continuous random variable x will be between x_1 and x_2, we integrate its probability density function $p(x)$.

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} p(x)\, dx$$

Formally, the PDF is defined as a Radon-Nikodym derivative and should satisfy the following properties.

1. $\forall$x. $0 \leq$ p(x)
2. $\int_{-\infty}^{\infty}$ p(x) dx = 1

We define the probability density function in HOL using the definition of Radon-Nikodym derivative of Mhamdi [36]. The distribution of random variables paired with Borel space and Borel sigma algebra gives the probability measure. The PDF of a random variable X is the derivative of the probability measure with respect to the Lebesgue-Borel measure. It is defined in HOL as:

Definition 4.2. *(Probability Density Function)*

```
⊢ PDF X p = RN_deriv lborel (space borel, subsets borel, measurable_distr p X)
```

where `measurable_distr` is the same as the distribution in the Probability theory but limited to sets measurable with respect to Lebesgue-Borel measure,

Definition 4.3. *(Measurable Distribution)*

```
⊢ measurable_distr p X =
    (λA. if A ∈ measurable_sets lborel then distribution p X A else 0)
```

With the help of Radon-Nikodym Theorem mentioned above, following properties of PDF were proved in HOL.

Theorem 4.2. *PDF of a random variable is always positive.*

```
⊢ ∀p X v. (v = (space borel, subsets borel, measurable_distr p X)) ∧
   measure_space v ∧ measure_absolutely_continuous v lborel ⇒
   ∀x. 0 ≤ PDF p X x
```

Proof. Using Theorem 3.20, we prove that Lebesgue-Borel measure is sigma-finite. We rewrite the goal using the Radon-Nikodym theorem, then used the Hibert choice elimination tactic and prove the existence and uniqueness of a non-negative measurable function, i.e., `PDF p X` □

Theorem 4.3. *Integral of PDF over whole space is equal to 1.*

```
⊢ ∀p X v. (v = (space borel, subsets borel, measurable_distr p X)) ∧
   prob_space v ∧ measure_absolutely_continuous v lborel ⇒
   integral m (PDF p X) = 1
```

Proof. Using the definition of probability space, we prove that the probability of an entire space is equal to 1. Then prove above theorem with the help of the Hilbert choice elimination tactic and the Radon-Nikodym Theorem. □

Both properties were easily proven because the probability measure is not defined for a certain random variable and therefore, absolutely continuous and probability space properties are taken as assumption. In the next section, we define the probability measure for a normal random variable and prove the validity of our assumptions.

4.3 Normal Random Variables

A random variable is a function that maps randomness to an appropriate set of numbers. It is called a normal random variable when its density is given as:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{(-\frac{(x-\mu)^2}{2\sigma^2})} \tag{7}$$

where $p(x)$ denotes the probability density function. From Equation (7), it is clear that the probability density of a normal random variable, called normal density, is completely defined by its mean μ and variance σ^2. We define the normal density in HOL as follows.

Definition 4.4. *(Normal Density)*

```
⊢ normal_density μ σ x =
  1 / sqrt (2 * π * σ pow 2) * exp (- (x - μ) pow 2 / 2 * σ pow 2)
```

In the limit when σ tends to zero, the `normal density` eventually tends to zero for any $\mathtt{x} \neq \mu$, but grows without limit if $\mathrm{x} = \mu$. Therefore, the normal density cannot be defined as an ordinary function when $\sigma = 0$. For this reason, most theorems involving normal density would require to add a premise, i.e., $0 < \sigma$. Using Definition 4.4, we prove some useful properties of normal density.

Theorem 4.4. *Normal density is always positive.*

```
  ⊢ ∀μ σ x. 0 ≤ normal_density μ σ x
```

Proof. Above theorem is easily proven, since multiplication, division and square root returns a positive number when given positive arguments and exponential function is always positive. □

Theorem 4.5. *If $0 < \sigma$, then normal density is also greater than 0.*

```
  ⊢ ∀μ σ x. 0 < σ ⇒ 0 < normal_density μ σ x
```

Proof. We use Theorem 4.4 and prove that normal density is not equal to 0 when variance is not equal to zero. □

Theorem 4.6. *Normal density is a Borel measurable function.*

```
⊢ ∀μ σ. (λx. Normal (normal_density μ σ x)) ∈
    measurable (m_space lborel, measurable_sets lborel) Borel
```

where `Normal` is used to maps real numbers to extended real numbers. To prove various properties of normal random variables, it is required to perform Lebesgue integration on normal density and since the Lebesgue Integral is defined for extended real valued functions, we add `Normal` to normal density.

Proof. We use Theorem 3.14 and prove that constant functions are Borel measurable using a property from the Lebesgue integration theory of HOL4. The exponential function is also proved Borel measurable by proving that it is continuous and using Theorem 3.16. □

Now we formalize the probability measure for normal random variables. It gives the probability that an event A (i.e., $P(X \in A)$) will occur for a normal random variable X.

Definition 4.5. *(Normal Probability Measure)*

```
⊢ normal_pmeasure μ σ A =
   if A ∈ measurable_sets lborel
   then pos_fn_integral lborel
    (λx. Normal (normal_density μ σ x) * indicator_fn A x)
   else 0
```

where `indicator_fn` is defined as:

```
⊢ indicator_fn A = (λx. if x ∈ A then 1 else 0)
```

Our definition is limited to measurable functions since, it is not possible to evaluate the integral of a function over non-measurable sets. Now we can define the normal random variable.

Definition 4.6. *(Normal Random Variable)*

```
⊢ normal_rv X p μ σ =
   random_variable X p borel ∧
   measurable_distr p X = normal_pmeasure μ σ
```

The first conjunct indicates that X is a real random variable, i.e., it is measurable from probability space to Borel space and the second conjunct ensures that it is a normal random variable as its distribution is given by normal probability measure. In the next section, we present some useful properties of normal random variables.

4.4 Properties of Normal Random Variable

In this section, we discuss some interesting properties of normal random variables. These properties would be useful while conducting formal verification of real world applications. The theorems related to affine transformation and independent random variables were ported from the Isabelle/HOL theorem prover [14].

4.4.1 PDF Properties

To prove the two properties of probability density functions for normal random variable, we first need to a validate the assumptions, i.e., the premises in Theorems 4.2 and 4.3.

Theorem 4.7. *Normal probability measure is absolutely continuous with respect to Lebesgue-Borel measure.*

```
⊢ ∀μ σ. measure_absolutely_continuous
         (space borel, subsets borel, normal_pmeasure μ σ
         (lborel)
```

Proof. We start by proving that integral of a non-negative Borel measurable function over a null set is equivalent to $\exists N.\ \lambda N = 0$ and $A \subset N$ and use it to rewrite our goal after unfolding the definition of `normal_pmeasure`. The goal is then proved using Definition 4.1. □

Using Theorem 4.7 and the properties of PDF, following theorems are proved.

Theorem 4.8. *PDF of a normal random variable is non-negative.*

```
⊢ ∀X p μ σ. normal_rv X p μ σ ⇒
     ∀x. 0 ≤ PDF p X x
```

Theorem 4.9. *PDF integral over the whole space is equal to 1*

```
⊢ ∀X p μ σ. normal_rv X p μ σ ⇒
    integral lborel (PDF p X) = 1
```

4.4.2 Symmetric Around Mean

Theorem 4.10. *For a normal random variable X with p(x) = N(μ, σ),*

$$\int_{-\infty}^{\mu} p(x)\,dx = \int_{\mu}^{\infty} p(x)\,dx$$

```
⊢ ∀X p μ σ. normal_rv X p μ σ ⇒
    pos_fn_integral lborel (λx. PDF p X x * indicator_fn {x | x ≤ μ} x) =
    pos_fn_integral lborel (λx. PDF p X x * indicator_fn {x | μ ≤ x} x)
```

Theorem 4.11. *For a normal random variable X with p(x) = N(μ, σ),*

$$\int_{\mu-a}^{\mu} p(x)\,dx = \int_{\mu}^{\mu+a} p(x)\,dx$$

```
⊢ ∀X p μ σ a. normal_rv X p μ σ ⇒
    pos_fn_integral lborel
      (λx. PDF p X x * indicator_fn {x | μ-a ≤ x ∧ x ≤ μ} x) =
    pos_fn_integral lborel
      (λx. PDF p X x * indicator_fn {x | μ ≤ x ∧ x ≤ μ+a } x)
```

Proof. For both properties, we use Theorem 3.21 to affine transform the right hand side value and prove that the normal density gives same values for x equi-distance from the mean μ on either side. □

4.4.3 Half Distribution

Using the symmetric property of normal distribution, we prove that for the normal probability density function, the integral of the first half is equal to the integral of the second half and since the full integral is equal to 1, the integral of the half distribution is $\frac{1}{2}$.

Theorem 4.12. *For a normal random variable X with p(x) = $N(\mu, \sigma)$,*

$$\int_{-\infty}^{\infty} p(x)\,dx = \int_{-\infty}^{\mu} p(x)\,dx + \int_{\mu}^{\infty} p(x)\,dx.$$

```
⊢ ∀X p μ σ. normal_rv X p μ σ ⇒
    pos_fn_integral lborel (λx. PDF p X x) =
    pos_fn_integral lborel (λx. PDF p X x * indicator_fn {x | x ≤ μ} x) +
    pos_fn_integral lborel (λx. PDF p X x * indicator_fn {x | μ ≤ x} x)
```

Theorem 4.13. *For a normal random variable X with p(x) = $N(\mu, \sigma)$,*

$$\int_{-\infty}^{\mu} p(x)\,dx = \frac{1}{2}.$$

```
⊢ ∀X p μ σ. normal_rv X p μ σ ∧ A = {x | x ≤ μ} ⇒
    pos_fn_integral lborel (λx. PDF p X x * indicator_fn A x) = 1 / 2
```

Theorem 4.14. *For a normal random variable X with p(x) = $N(\mu, \sigma)$,*

$$\int_{\mu}^{\infty} p(x)\,dx = \frac{1}{2}.$$

```
⊢ ∀X p μ σ. normal_rv X p μ σ ∧ A = {x | μ ≤ x} ⇒
    pos_fn_integral lborel (λx. PDF p X x * indicator_fn A x) = 1 / 2
```

4.4.4 Affine Transformation

Theorem 4.15. *If X is a normal random variable with mean μ and standard deviation σ then $Y = b + a * X$ is also a normal random variable with mean b + a * μ and standard deviation $|\ a\ |$ * σ.*

```
⊢ ∀X p μ Y a b. a ≠ 0 ∧ 0 < σ ∧ normal_rv X p μ σ ∧
   (∀x. Y x = b + a * X x) ⇒
   normal_rv Y p (b + a * μ) (abs a * σ)
```

Proof. We unfold the definition of normal random variable and prove that Y is also a random variable using Definition 2.3. We prove $\lambda x.$ b + a * x is a continuous function and therefore, Borel measurable. Then we prove the equivalence of the probability distributions using Theorem 3.21. □

4.4.5 Convolution

Theorem 4.16. *Convolution of $p_1(x) = N(0, \sigma_1)$ and $p_2(x) = N(0, \sigma_2)$ is equal to $p(x) = N(0, \sqrt{\sigma_1^2 + \sigma_2^2})$,*

$$\int_{-\infty}^{\infty} p_1(z-x) \cdot p_2(x)\, dx = p(z)$$

```
⊢ ∀σ1 σ2 p X Y x. 0 < σ1 ∧ 0 < σ2 ∧ normal_rv X p 0 σ1 ⇒
    pos_fn_integral lborel
    (λy. Normal (normal_density 0 σ1 (x - y) *
         Normal (normal_density 0 σ2 y))) =
    Normal (normal_density 0 (sqrt (σ1 pow 2 + σ2 pow 2)) x)
```

Proof. We start by unfolding the definition of normal density and exploiting the properties of exponential function to separate a constant value equivalent to the right hand side. Then we use a property of the Lebesgue integration to bring the constant value out of the integral. The goal is then proved equal to 1 using real analysis and Theorem 4.9. □

4.4.6 Sum of Independent Random Variables

If X and Y are two normally distributed independent normal random variables. Then their sum is also a normally distributed random variable, i.e.,

$$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2), Z = X + Y \textit{ then, } Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

To prove this property, we start by formalizing independent variables in HOL4:

Definition 4.7. *(Independent Events)*
A finite set of events F_i is pairwise independent if and only if every pair of events is independent.

$$P(F_m \cap F_n) = P(F_m) \cdot P(F_n)$$

```
⊢ indep_sets p F I = prob_space p ∧
   (∀i. i ∈ I ⇒ F i SUBSET events p) ∧
   (∀J. J SUBSET I ∧ J ≠ {} ∧ FINITE J ⇒)
      ∀A. A ∈ (Pi J F) ⇒ (prob p (BIGINTER A j| j ∈ J) =
         Normal (product J (λj. real (prob p (A j)))))
```

where `Pi` is defined as:

```
⊢ Pi A B = {f | ∀x. x ∈ A ⇒ f x ∈ B x}
```

Having defined the independent events, we can now define independent random variables,

Definition 4.8. *(Independent Random Variables)*
A set of random variables is pairwise independent if and only if every pair of random variables is independent, i.e., a pair of random variables X and Y are independent if for every measurable set A and B, the events $\{X \in A\}$ and $\{Y \in B\}$ are independent.

```
⊢ indep_vars p M X I =
   (∀i. i ∈ I ⇒
     random_variable (X i) p (m_space (M i), measurable_sets (M i))) ∧
   indep_sets p
     (λi. PREIMAGE X A INTER p_space p | A ∈ measurable_sets (M i)) I
```

Using above definitions, we can define two independent events and two independent random variables as:

```
⊢ indep_set p A B = prob_space p ∧
   indep_sets p (λi. if i = 0 then A else B) UNIV
```

```
⊢ indep_var p M_a A M_b B =
   indep_vars p (λi. if i = 0 then M_a else M_b)
                (λi. if i = 0 then A else B) UNIV
```

To prove that the sum of independent normal random variables X and Y is also a normal random variable, we start by proving that the density f_z of $Z = X + Y$ is the convolution of f_x and f_y.

$$f_z(z) = \int_{-\infty}^{\infty} f_y(z-x) \cdot f_x(x)\, dx \tag{8}$$

For this purpose, we first formalize the notion of convolution for measures as:

Definition 4.9. *(Convolution of Measures)*
The convolution $M * N$ is the probability density of the sum $X + Y$ of two independent random variables X and Y whose respective PDFs are M and N.

```
⊢ convolution M N = (space borel, subsets borel,
   (λA. measure (pair_measure M N)
       (PREIMAGE (λ(x,y). x + y) A ∩ m_space (pair_measure M N))))
```

where `pair_measure` is defined as:

Definition 4.10. *(Product Measure)*
A product measure $\mu_1 \times \mu_2$ *is defined to be a measure on the measurable space* $(X_1 \times X_2, \Sigma_1 \otimes \Sigma_2)$ *satisfying the property,* $(\mu_1 \times \mu_2)(B_1 \times B_2) = \mu_1(B_1)\ \mu_2(B_2)$ *for all* $B_1 \in \Sigma_1$, $B_2 \in \Sigma_2$.

```
⊢ pair_measure A B = ((m_space A CROSS m_space B),
    {A CROSS B | a ∈ measurable_sets A ∧ B ∈ measurable_sets B},
    (λX. pos_fn_integral A (λx. pos_fn_integral B (λy. indicator_fn X (x,y))))
```

Now we formalize Equation (8):

Theorem 4.17. *For two random variables X and Y, the PDF* f_z *of* $Z = X + Y$ *is the convolution of* f_x *and* f_y*.*

```
⊢ ∀p X Y. prob_space p ∧
   indep_var p borel_triplet X borel_triplet Y ⇒
   (∀a. a ∈ measurable_sets (space borel, subsets borel, (λx. 0))) ⇒
   (distribution p (λx. X x + Y x)) a =
    measure (convolution (distr' p borel_triplet X)
                         (distr' p borel_triplet Y) a)
```

where `borel_triplet = (borel space, subsets borel, (λx. 0))` and `distr'` is defined as:

```
⊢ distr' M N X = (m_space N, measurable_sets N,
     (λs. if s ∈ measurable_sets N then distribution M X s else 0))
```

With the help of above theorem, we prove that the sum of two independent normal random variables is also normal. Theorems 4.9 and 3.21 were required in the proof steps.

Theorem 4.18. *If* $X \sim N(\mu 1,\sigma 1^2)$ *and* $Y \sim N(\mu 2, \sigma 2^2)$ *are two independent random variables, then* $Z = X + Y$ *is also normal with mean* $(\mu 1 + \mu 2)$ *and variance* $(\sigma 1^2 + \sigma 2^2)$*.*

```
⊢ ∀p X Y μ1 μ2 σ1 σ2.
    prob_space p ∧ 0 < σ1 ∧ 0 < σ2 ∧
    indep_var p borel_triplet X borel_triplet Y ∧
    normal_rv X p μ1 σ1 ∧ normal_rv Y p μ2 σ2 ⇒
    normal_rv (λx. X x + Y x) p (μ1 + μ2) (sqrt (σ1 pow 2 + σ2 pow 2))
```

Another useful theorem required in the proof of above property is the Fubini theorem for Lebesgue integration [7].

Theorem 4.19. *Suppose $M1$ and $M2$ are σ-finite measure spaces. If $f(x,y)$ is $M1 \times M2$ measurable, then*

$$\int_{M1}\left(\int_{M2} f(x,y), dy\right) dx = \int_{M2}\left(\int_{M1} f(x,y), dx\right) dy$$

```
⊢ ∀f M1 M2. measure_space M1 ∧ measure_space M2 ∧
     sigma_finite_measure M1 ∧ sigma_finite_measure M2 ∧
     (∀x. 0 ≤ f x) ∧ f ∈ measurable (m_space (pair_measure M1 M2),
         measurable_sets (pair_measure M1 M2)) Borel ⇒
     (pos_fn_integral M1 (λx. pos_fn_integral M2 (λy. f (x,y))) =
      pos_fn_integral M2 (λy. pos_fn_integral M1 (λx. f (x,y))))
```

With Theorem 4.18 and using induction, we also prove that the sum of a finite number of independent normal random variables is also a normal random variable.

Theorem 4.20. *If $X_i \sim N(\mu_i,\sigma_i^2)$ is a finite set of independent normal random variables, and $Z = \sum_i X_i$ then, $Z \sim N(\sum_i \mu_i, \sum_i \sigma_i^2)$.*

```
⊢ ∀p X μ σ I.
    prob_space p ∧ FINITE I ∧ ∧ I ≠ {} ∧
    indep_vars p (λi. borel_triplet X I) ∧
    (∀i, i ∈ I ⇒ 0 < σ i) ∧
    (∀i, i ∈ I ⇒ normal_rv (X i) p (μ i) (σ i)) ⇒
    normal_rv (λx. sum I (λx. X i x)) p (sum I μ)
               (sqrt (sum I (λi. (σ i) pow 2)))
```

4.5 Summary

In this chapter, we discussed how the Random-Nikodym theorem was generalized for sigma finite measures. We defined the probability density function and used Radon-Nikodym theorem to proved its basic properties. Then we defined normal random variables along with their probability measures. Finally, we discuss some useful properties of normal distribution that will help in applying our formalization to real world applications. Two applications are discussed in the next chapter.

Chapter 5

Applications

In previous chapters, we have provided the formalization of normal random variables along with the formal proof of some useful properties. These results can be used in the formal probabilistic analysis of a wide range of engineering systems. We illustrate the usefulness of conducting this analysis using theorem proving by tackling two real world applications, i.e., the error probability of binary signals transmission in the presence of Gaussian noise and the probabilistic clock synchronization in wireless sensor networks.

5.1 Binary Signals Transmission in the Presence of Gaussian Noise

Throughout history, communication has been a very useful tool to mankind. The need for fast and long distance communication keeps increasing and the technology is getting more advanced to satisfy those needs. Today, communication systems are being used almost everywhere including safety critical domains such as medical instruments and traffic. Hence, there is a dire need to verify such systems formally. Generally, communication systems involve transmitting data or information from a sender to a receiver. This is done mainly in two ways, i.e., analog transmission and digital (binary) transmission. Over time, binary transmission has become more popular, as analog is more prone to noise and attenuation. Also, analog transmission is more expensive as the required hardware transmitters and receivers are designed to fit a particular transmission. A simple binary transmission system consists of a transmitting device sending a stream of binary signals via a channel (e.g. copper wire) and a receiver. Figure 5.1 shows a simplified block diagram of such a system [47].

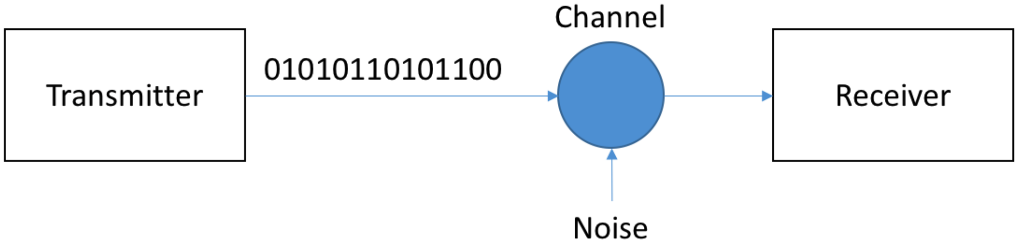

Figure 5.1: A Simple Binary Transmission System

In a binary transmission system, noise could cause 0 to be interpreted as 1 and vice versa, which makes data at the receiver different from the sender. A very common type of noise is Gaussian noise. It comes from many natural sources such as thermal vibration of atoms in conductors (e.g., copper wire), black body radiation from the earth and other warm objects, and from celestial sources such as the sun. It is called Gaussian because the values it can take are Gaussian (or normally) distributed. In a channel with Gaussian noise, the signal at the receiver would look as given in Figure 5.2.

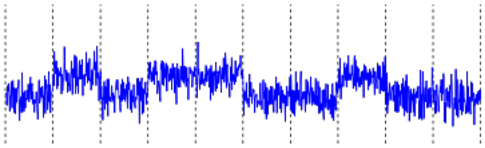

Figure 5.2: Binary Signals with Gaussian Noise [53]

If $f(t)$ is the transmitted signal and $N(t)$ is the added noise, then the signal at the receiver is

$$Y(t) = f(t) + N(t)$$

where $f(t)$ is a noise free signal and is given as

$$f(t) = \begin{cases} 0 & \text{signal absent} \\ A & \text{signal present} \end{cases}$$

Therefore, when the signal is absent only noise is received. Since the noise is random, the receiver cannot decide with certainty whether a signal was present or absent at a certain time. However, a reasonable rule for the decision is as follows,

$$Y(t) \leq \mu \quad sigmal\ absent$$

$$Y(t) > \mu \quad signal\ present$$

When the noise is Gaussian or normal distributed with a mean of zero and variance of σ^2. The probability density function of $Y(t)$ when the signal is absent is given as

$$P_0(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{(-\frac{y^2}{2\sigma^2})}$$

where y denotes $Y(t)$. Similarly when the signal is present,

$$P_1(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{(-\frac{(y-A)^2}{2\sigma^2})}$$

This is shown in Figure 5.3.

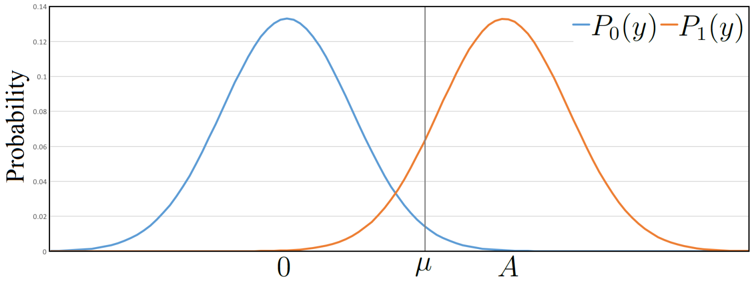

Figure 5.3: Distribution of Recieved Signals

where 0 denotes that a signal is absent and A denotes that a signal is present. μ is the threshold for taking decision at the receiver.

Using the decision rule mentioned above, it is clear that sometimes the receiver will decide a signal is present when it is absent and vice versa. The probability of error in both cases, is given as:

$$P_{\epsilon 0} = \int_{\mu}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp^{(-\frac{(y)^2}{2\sigma^2})} \, \mathrm{d}y$$
$$P_{\epsilon 1} = \int_{-\infty}^{\mu} \frac{1}{\sigma\sqrt{2\pi}} \exp^{(-\frac{(y-A)^2}{2\sigma^2})} \, \mathrm{d}y$$

We define this is HOL as the distribution of Y in the error region

```
⊢ prob_error_zero_sent p X Y u =
   conditional_distribution p Y X {x | u < x} {0} =
   distribution p Y {x | x ≤ u}
```

```
⊢ prob_error_zero_sent p X Y u =
   conditional_distribution p Y X {x | x ≤ u} {1} =
   distribution p Y {x | u < x}
```

where X is a random variable used to denote the transmitted signal which is either 0 or 1 with equal probability and random variable Y denotes the received signal. The definition of signal and decision made on the received signal are as follows

```
⊢ signal a b = (λx. if x = 0 then a else b)
```

```
⊢ decision u a b = (λx. if x ≤ u then a else b)
```

We define the total probability of error as the joint distribution of X and Y for which the decision is made is different from the original signal.

```
⊢ prob_error p X Y u a b =
   joint_distribution p X Y {(x,y)  | signal a b x <> decision u a b y}
```

We prove that the total probability of error is minimum when the decision threshold is at the exact middle of binary level 0 and binary level 1, i.e., $\mu = \frac{A}{2}$.

```
⊢ ∀ X Y N p μ σ u a b.
   noise N p 0 σ ∧ bern_rv X p (1/2) ∧ a ≤ b ∧
   (∀x y. Y y = N y + signal a b x) ∧ (0 < σ) ∧
   (∀u. prob_error_zero_sent p X Y u) ∧
   (∀u. prob_error_one_sent p X Y u) ⇒
   prob_error p X Y ((a + b) / 2) a b ≤ prob_error p X Y u a b
```

where noise is a normal random variable with a mean of zero and `bern_rv X p (1/2)` denotes that `X` is a bernoulli random variable with equal probability of failure and success. Finally, we prove that the probability of error is equal to the $Q(x)$ function when $x = \frac{a}{2\sigma}$.

```
⊢ ∀ X Y N p μ σ u a.
   noise N p 0 σ ∧ bern_rv X p (1/2) ∧
   (∀x y. Y y = N y + signal 0 a x) ∧
   (a ≠ 0) ∧ (0 < σ) ∧ (u = a / 2) ∧
   (∀u. prob_error_zero_sent p X Y u) ∧
   (∀u. prob_error_one_sent p X Y u) ⇒
   prob_error p X Y u 0 a = Q_func (a / 2 * σ)
```

where the Q function is given as:

```
⊢ Q_func z = pos_fn_integral lborel
    (λx. Normal (1 / sqrt (2 * π) * exp (-(x pow 2) / 2)) *
                     indicator_fn {x | z ≤ x} x)
```

5.2 Probabilistic Clock Synchronization in Wireless Sensor Network

A network of autonomous wireless sensors deployed to monitor a physical phenomena is called a wireless sensor network [54]. Wireless sensor networks have applications in many critical domains such as, health care monitoring, forest wire detection and natural disaster prevention. Recent advances in technology have made low-cost, low-power wireless sensors a reality. It is required that the network be energy efficient as the sensors have a limited power source. An important service in sensor networks is clock synchronization which is used for time division multiple access (TDMA) scheduling and power mode energy saving.

Wireless sensor networks have some unique characteristics due to which it is difficult to apply traditional approaches for clock synchronization. Highly accurate clock synchronization protocols require more processing and hence more energy consumption. Elson et al. [15] presented an analytical way to convert service specifications to protocol parameters, called reference broadcast sychronization (RBS). PalChaudhuri et al. [41] extended this work and provided probabilistic bounds on clock synchronization error for single and multi-hop networks. We conduct a formal analysis of both cases.

5.2.1 Sources of Clock Synchronization Error

The non-determinism in message delivery latency is the main cause of error. Kopetz et al. [28] have characterized the message delivery latency into four distinct components:

1. Send Time: time required to build the message at the sender node.

2. Access Time: waiting time required to get access to the transmission channel.

3. Propagation Time: time required for the message to reach the receiver.

4. Receive Time: processing time required at the receiver.

The RBS protocol entails synchronizing a set of receivers with each other, in contrast to synchronizing with the sender. For this reason, the send time and access time are not of consequence as they are identical for all receivers. The only variable times are the Propagation Time and Receive Time.

5.2.2 Single-Hop Network

PalChaudhuri et al. [41] presented the clock synchronization protocol as an extension to RBS. The following happens for every sender sensor in a single-hop broadcast region.

1. A sender broadcasts n reference packets to its neighbors. The interval between each packet is fixed and greater than some minimum, such that they are independent of each other.

2. Each receiver records the time according to its own local clock, when each of these reference packets are received. Using these time-stamps, the receiver uses linear regression to fit a line on these data. The slope of the line will approximate the relative clock skew between the receiver and the sender.

3. Each receiver sends back to the sender, a packet containing the slope of the line and one point on that line. The sending back of these packets are jittered over an interval so that the packets sent back by different receivers have less chance of colliding with each other.

4. The sender composes all these slopes together, and broadcasts a packet containing its relative clock skew slope to all receivers who have replied back.

5. Each receiver after receiving this packet, can now calculate its own slope relative to all receivers in the broadcast region of a particular sender. So, for every pair of receivers, within the broadcast region of the sender, the clock skew and clock offset are now known with some synchronization error. The Send Time and Access Time errors are factored out when calculating this relative slope, as that error is the same for any two receivers. The only error present will be that due to Propagation Time and Receive Time.

Elson et al. [15] found the distribution of the synchronization error among receivers. Multiple pulses are sent from the sender to the set of receivers. The difference in actual reception time at the receivers is plotted. As each of these pulses are independently distributed, the difference in reception times gives a normal distribution with zero mean. If the maximum error that is allowed between two sensors is ϵ_{max}, then the probability of synchronization with an error $\epsilon \leq \epsilon_{max}$ is given as

$$P(|\epsilon| \leq \epsilon_{max}) = \frac{\int_{-\epsilon_{max}}^{\epsilon_{max}} \exp^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

For n reference packets from the sender, the receivers exchange their observations. As explained earlier, the slope of the skew between the receivers is found by a least square linear estimation using the n data points. The calculated slope of the skew has an associated error in it. This error is the difference in phase between the calculated slope and the actual slope. As the points have a normal distribution, this error can be calculated as

$$P(|\epsilon| \leq \epsilon_{max}) = 2\ \mathit{erf}\left(\frac{\sqrt{n}\epsilon_{max}}{\sigma}\right)$$

where ϵ is the synchronization error, i.e., difference in packet reception time between two sensor, ϵ_{max} is the maximum allowable error, n is the minimum number of synchronization messages to guarantee the specified error, σ^2 is the variation of the distribution and `erf` is the error function given as,

$$\mathit{erf}(z) = \frac{\int_0^z \exp^{-\frac{x^2}{2}} \mathrm{d}x}{\sqrt{2\pi}}$$

We formalize this in HOL as follows,

Definition 5.1. *(Error Function)*

```
⊢  err_func z = pos_fn_integral lborel
      (λx. Normal (1 / sqrt (2 * π) * exp (-(x pow 2) / 2)) *
            indicator_fn {x | 0 ≤ x / x ≤ z} x)
```

We then define probability of synchronization within the maximum allowable error ϵ_{max} as the distribution of random variable Z.

```
⊢  prob_sync_error p Z = measurable_distr p Z
```

Now we calculate the probability of synchronization error for n reference packets.

Theorem 5.1. *For n reference packets, the probability of synchronization error is calculated as*

$$P(\mid \epsilon \mid \leq \epsilon_{max}) = 2\ erf\left(\frac{\sqrt{n}\epsilon_{max}}{\sigma}\right)$$

```
⊢  ∀p X μ σ n Emax.
   prob_space p ∧ (I = (1 .. n)) ∧ (0 < σ) ∧
   (0 < n) ∧ (∀i. i ∈ I ⇒ sync_error (X i) p μ σ) ∧
   (Z = (λx. sum I (λi. X i x) / n)) ∧ (μ = 0) ∧ 0 ≤ Emax ⇒
   (prob_sync_error p Z {x | -Emax ≤ x ∧ x ≤ Emax} =
    2 * err_func (Emax * sqrt n / σ))
```

where `sync_error` is equivalent to a normal random variable, Z is the average error for n reference packets and `Emax` is the maximum allowable synchronization error.

Proof. First we find the probability density function of Z using Theorems 4.15 and 4.19. Then we use the symmetry property, i.e., Theorem 4.11 to eliminate 2 from the right hand side. The goal is achieved with the help of Theorem 3.21. □

5.2.3 Multi-Hop Network

A multi-hop wireless network has several benefits over networks with single wireless link. It can extend the coverage of the network and improve connectivity. Also, transmission over multiple short links might consume less energy. They enable higher data rates resulting in higher throughput and more efficient use of the wireless medium. In case of dense multi-hop networks, several paths might become available that can be used to increase the reliability of the network. A multi-hop network also introduces some challenges, for instance, clock synchronization error would increase with every hop. PalChaudhuri et al. [41] extended the single-hop protocol to also provide probabilistic bounds on clock synchronization error for this type of network. For this protocol, senders are considered at various levels. A sender which does not need any synchronization is called a sender at level 0. A sensor node which is within the broadcast region of a sender at level 0 can behave as a sender in order to synchronize sensor nodes which are two hops away from the sender at level 0. Such a sender is called a sender at level 1. Receivers within the broadcast region of the sender at level 0 are synchronized using the same method discussed in previous section. Once these receivers get synchronized, each receiver starts behaving as a sender at level 1. Consider the scenario presented in Figure 5.4. Nodes R1, R2, R3 and R4 are within the broadcast region of the sender S. Using the single hop synchronization protocol nodes R1, R2, R3 and R4 are synchronized among themselves. Suppose R2 gets to be the first node to end the reference broadcast, that is R2 starts behaving as a sender at level 1. By a similar synchronization procedure, R1, R3, R5, R6 and R7 get synchronized among themselves. Now suppose R6 needs to send a message to R4. The message would have to be routed through a node which is synchronized with R6, say R3. The assumption here is that due to the relative high density of sensor nodes, a node, such as R3 as shown in Figure 5.4, will exist in the broadcast region of two senders; the two senders might be at the same level or they might be separated by a single level. Now since R3 is synchronized with R4, R3 can transform the time reported by R6. Finally, since R3 is synchronized with R4, R4 can transform the time reported by R3. Hence, all along the routing path of the message suitable time transformations can be performed.

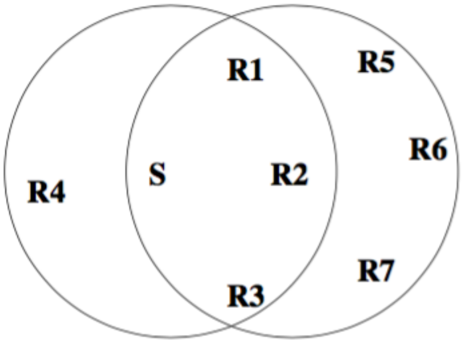

Figure 5.4: Multihop Network [41]

We define transformation in HOL as the sum of synchronization error and find the maximum synchronization possible along with the probability that error will stay within bounds for k hops,

Definition 5.2. *(Transformation)*

```
⊢ transformation X k = (λx. sum (1 .. k) (λi. X i x))
```

Theorem 5.2. *If Emax is the max allowable error for a single hop, then maximum error possible between two sensor nodes which are k hops apart is k * Emax.*

```
⊢  ∀X Emax k. 0 ≤ Emax ⇒
     (∀x. (∀i. (X i) x ∈ {x:real | -Emax ≤ x ∧ x ≤ Emax}) ⇒
       transformation X k x ∈ {x:real | -Emax * &k ≤ x ∧ x ≤ Emax * &k}
```

Proof. The theorem is easily proved by unfolding the definition of `transformation` and using properties of summation and real analysis. □

Theorem 5.3. *If we consider the error over a single hop to Emax then error over k hops will be* $\sqrt{k}$ ** Emax.*

```
⊢  ∀p X μ σ k Emax.
   prob_space p ∧ (I = (1 .. n)) ∧ (0 < σ) ∧
   indep_vars p (λi. (space borel, subsets borel, (λx. 0))) λ
   (0 < k) ∧ (∀i. i ∈ I ⇒ sync_error (X i) p μ σ) ∧
   (Z = (λx. sum I (λi. X i x))) ∧ (μ = 0) ∧ (0 ≤ Emax) ⇒
   (prob_sync_error p Z {x | -Emax * sqrt(k) ≤ x ∧ x ≤ Emax * sqrt(k)} =
    prob_sync_error p (X k) {x | -Emax ≤ x ∧ x ≤ Emax})
```

Proof. First we find the probability density function of Z using Theorem 4.19. Then we apply affine transformation on the left hand side using Theorem 3.21. Finally, we prove the goal using real analysis. □

5.3 Summary

In this chapter, we presented two example applications. In the first example, we formalized a binary transmission system and verified the probability of error in the presence of Gaussian noise. In the second example, we analysed the probabilistic clock synchronization in wireless sensor networks presented in a prominent paper [41]. These applications illustrate how our formalization of Lebesgue measure and normal random variables can be used to reason about engineering applications. Conducting the analysis within the sound core of a theorem prover helped to add more trust to the proved results. The soundness and the deduction style of the theorem prover guarantee the validity of the analysis when deriving these proofs. Besides, the results of this type of analysis are generic and valid for any instance of the system. We argue that these benefits are even more significant when dealing with larger and more complex systems as is the case for now-a-days parallel and distributed systems.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

The analysis of engineering systems used in safety critical domains such as transportation and medicine, is usually done using informal techniques. The unreliable results produced using such techniques may lead to heavy financial loss, or even the loss of a human life. Therefore, in this thesis we propose to conduct the probabilistic analysis of engineering systems exhibiting normally distributed randomness using higher-order logic theorem proving. To do so, we have provided the formalization of the mathematical notions required to formalized normal random variables. Compared to the standard techniques of computer simulation and paper-and-pencil analysis, our approach provides more accurate and trusted results by exploiting the soundness of theorem proving. It also allows to provide generic results instead of proving the properties for specific instances of the system.

The main purpose of this thesis is to develop a formalization of normal random variables along with other mathematical notions that are required to reach this goal. Normal random variables as any other continuous random variables are defined by their probability density functions (PDF). We formalized the PDF based on the Radon-Nikodym derivative, i.e., the derivative of probability measure with respect to a reference measure. To serve as a reference measure, we formalized the Lebesgue-Borel measure based on Gauge integral (or Henstock Kurzweil integral). We require the Radon-Nikodym theorem for the proof of some basic properties of PDF. This theorem has been proved by Mhamdi [36] but is limited to finite measures. Because the Lebesgue-Borel is not a finite measure, we generalized the Radon-Nikodym theorem for sigma-finite measures. The borel sigma algebra is another important notion required for proving many useful theorems. Due to the unavailability of some topological notions for extended real numbers, Mhamdi [36] defined it as open intervals instead of open sets. For this reasom some essential theorems could not be proved. We added an axiom to overcome this limitation. Also, for the purpose of compatibility with real valued Gauge integral, we formalized Borel measurable sets or Borel sigma algebra for real numbers as open sets and ported all required topological notions from HOL Light.

This work was conducted using the HOL4 kananaskis 10 version of the theorem prover and the main reason behind this choice was to be able to utilize available higher-order-logic formalizations of measure, Lebesgue integration and probability theories. Unfortunately, the theory of Gauge integral was not available, therefore, it was ported from the HOL Light theorem prover to HOL4. Some other notions were also ported from the Isabelle/HOL theorem prover, that helped us prove many useful properties of the Lebesgue-Borel measure and normal random variables. We also introduced a measurable distribution that is the same as the distribution definition in the probability thoery of Mhamdi [36], but is limited to Borel measurable sets. Other than the theories ported from the HOL Light theorem prover, the proof script of the formalization and verification of the notions presented in this thesis require around 18000 lines of HOL4 code available at `http://hvg.ece.concordia.ca/projects/prob-it/pr7.html`.

To prove the usefulness of our formalization, we conducted the formal analysis of two example applications. First, we modeled and verified a binary transmission system in the presence of Gaussian noise where Gaussian noise was modeled as normal random variable. Next we analyzed the probabilistic clock synchronization in wireless sensor networks. In this application, the main challenge was to verify the mean and variance of sum of random variables as well as the affinity property. These applications highlight the feasibility and benefits of conducting the probabilistic analysis using a higher-order-logic theorem prover. In fact, the added trust provided by the deduction style of theorem proving, is a crucial requirement when dealing with safety-critical applications.

6.2 Future Work

Some of the worth mentioning extensions of our formalization are outlined as follows:

1. The formalization of multivariate random variables will strengthen the probabilistic analysis framework of HOL4 and allow the analysis of many large and complex engineering systems. Also, porting the multivariate theories of the HOL Light theorem prover to HOL4 will facilitate the formalization of multivariate random variables. They will also be useful while conducting formal analyses involving multivariate real or complex numbers.

2. The formalization of a generalized notion of open and closed sets will allow a generalized definition of extended real Borel sigma algebra. With this we can prove that all closed sets of $\overline{\mathbb{R}}$ belong to Borel sets and since all singletons are closed sets, the singletons of positive and negative infinity also belong to Borel sets. Hence, nullifying the need to take the axiom of infinity.

3. Different types of integrals are distinguished by their ability to handle differing special cases. However, a large set of functions can be handled by all of them. Proving the equivalence of deferent integrals for a certain class of functions will allow us to better exploit the proven properties of these integrals.

Bibliography

[1] *Vesta: A Statistical Model-Checker and Analyzer for Probabilistic Systems*, 2005.

[2] R. Affeldt and M. Hagiwara. Formalization of Shannon's Theorems in SSReflectcoq. In *Interactive Theorem Proving*, volume 7406 of LNCS, pages 233–249. Springer, 2012.

[3] P. B. Andrews. *An Introduction to Mathematical Logic and Type Theory*, volume 27. Springer Science & Business Media, 2002.

[4] K. B. Athreya and S. N. Lahiri. *Measure Theory and Probability Theory*. Springer Science & Business Media, 2006.

[5] J. Avigad, J. Hölzl, and L. Serafin. A Formally Verified Proof of the Central Limit Theorem (preliminary report). *arXiv preprint arXiv:1405.7012*, 2014.

[6] C. Baier and J.P. Katoen. *Principles of Model Checking*. MIT Press, 2008.

[7] H. Bauer. *Measure and Integration Theory*. de Gruyter, 2001.

[8] P. Billingsley. *Probability and Measure*. John Wiley & Sons Inc., 1995.

[9] C. E. Brown. *Automated Reasoning in Higher-order Logic*. College Publications, 2007.

[10] A. Coble. *Anonymity, Information, and Machine-Assisted Proof*. PhD thesis, University of Cambridge, UK, 2009.

[11] Coq. `http://coq.inria.fr/`, 2015.

[12] L. de Alfaro. *Formal Verification of Probabilistic Systems.* PhD thesis, Stanford University, USA, December 1997.

[13] A. J. Durán, M. Pérez, and J. L. Varona. The Misfortunes of a Trio of Mathematicians Using Computer Algebra Systems. can we trust in them? *Notices of the AMS*, 61(10), 2014.

[14] M. Eberl, J. Hölzl, and T. Nipkow. A Verified Compiler for Probability Density Functions. In *European Symposium on Programming*, volume 9032 of *LNCS*, pages 80–104. Springer, 2015.

[15] J. Elson, L. Girod, and D. Estrin. Fine-Grained Network Time Synchronization Using Reference Broadcasts. *ACM SIGOPS Operating Systems Review*, 36(SI):147–163, 2002.

[16] M. Fitting. *First-Order Logic and Automated Theorem Proving.* Springer Science & Business Media, 2012.

[17] J. V. Z. Gathen and J. Gerhard. *Modern Computer Algebra.* Cambridge University Press, 2013.

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

[19] M. J. C. Gordon. Introduction to the HOL system. In *Proceedings of the International Workshop on the HOL Theorem Proving System and its Applications*, pages 2–3, Davis, California, USA, 1991.

[20] M. J. C. Gordon and T. F. Melham, editors. *Introduction to HOL: A Theorem Proving Environment for Higher-Order Logic.* Cambridge University Press, 1993.

[21] J. Harrison. *Theorem Proving with the Real Numbers.* Springer, 2012.

[22] O. Hasan. *Formal Probabilistic Analysis using Theorem Proving.* PhD thesis, Concordia University, Montreal, QC, Canada, 2008.

[23] O. Hasan and S. Tahar. Formal Verification Methods. In *Encyclopedia of Information Science and Technology*, chapter 705, pages 7162–7170. IGI Global, 2015.

[24] J. Holzl and A. Heller. Three Chapters of Measure Theory in Isabelle/HOL. In *Interactive Theorem Proving*, volume 6898 of LNCS, pages 135–151. Springer, 2011.

[25] J. Hurd. *Formal Verification of Probabilistic Algorithms.* PhD thesis, University of Cambridge, UK, 2002.

[26] Isabelle/HOL. `https://isabelle.in.tum.de/`, 2015.

[27] A. N. Kolmogorov. *Foundations of Probability.* 1933. Second English Edition, *Foundations of Probability*, Chelsea Publishing Co, 1950.

[28] H. Kopetz and W. Ochsenreiter. Clock Synchronization in Distributed Real-Time Systems. *IEEE Transaction on Computers*, 36(8):933–940, August 1987.

[29] HOL Light. `http://www.cl.cam.ac.uk/~jrh13/hol-light/`, 2015.

[30] Maplesoft. `http://www.maplesoft.com/`, 2015.

[31] Mathematica. `http://www.wolfram.com`, 2015.

[32] mathStatica. `www.wolfram.com/products/applications/mathstatica`, 2008.

[33] MATLAB. `http://www.mathworks.com/products/statistics`, 2008.

[34] D. C. J. Matthews. Poly manual. *SIGPLAN Notes*, 20(9):52–76, September 1985.

[35] Maxima. `http://maxima.sourceforge.net`, 2015.

[36] T. Mhamdi. *Information-Theoretic Analysis using Theorem Proving.* PhD thesis, Concordia University, Montreal, QC, Canada, 2012.

[37] R. Milner, M. Tofte, and D. Macqueen. *The Definition of Standard ML*. MIT Press, Cambridge, MA, USA, 1997.

[38] Minitab. `http://www.minitab.com`, 2008.

[39] T. Nipkow, L. C. Paulson, and M. Wenzel. *Isabelle/HOL: A Proof Assistant for Higher-Order Logic*, volume 2283 of *LNCS*. Springer, 2002.

[40] M. Norrish and K. Slind. The HOL System Description. Technical report, http://hol.sourceforge.net/documentation.html, 2014.

[41] S. PalChaudhuri, A. K. Saha, and D. B. Johnson. Adaptive Clock Synchronization in Sensor Networks. In *Proceedings of International Symposium on Information Processing in Sensor Networks*, pages 340–348, New York, NY, USA, 2004. ACM.

[42] D. Parker. *Implementation of Symbolic Model Checking for Probabilistic System*. PhD thesis, University of Birmingham, UK, 2001.

[43] PRISM. `http://www.prismmodelchecker.org`, 2015.

[44] J. A. Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, 1995.

[45] S. Romanenko, C. Russo, and P. Sestoft. Moscow ML Language Overview. `http://mosml.org/mosmlref.pdf`, 2000.

[46] SAS. `http://sas.com/technologies/analytics/statistics/stat/index.html`, 2008.

[47] B. Sklar. *Digital Communications: Fundamentals and Applications*. Pearson Education Limited, 2001.

[48] K. Slind and M. Norrish. A Brief Overview of HOL4. In *Theorem Proving in Higher Order Logics*, volume 5170 of *LNCS*, pages 28–32. Springer, 2008.

[49] SPSS. `http://www.spss.com/`, 2008.

[50] S. M. Srivastava. *A Course on Borel Sets*, volume 180. Springer Science & Business Media, 2008.

[51] C. Swartz. *Introduction to Gauge Integrals*. World Scientific, 2001.

[52] Wikipedia. `https://en.wikipedia.org/wiki/Normal_distribution`, 2008.

[53] Wikipedia. `https://en.wikipedia.org/wiki/Matched_filter`, 2012.

[54] J. Yick, B. Mukherjee, and D. Ghosal. Wireless Sensor Network Survey. *Computer Networks*, 52(12):2292 – 2330, 2008.