Preface

Probabilistic analysis is a tool of fundamental importance for virtually all scientists and engineers as they often have to deal with systems that exhibit random or unpredictable elements. Traditionally, computer simulation techniques are used to perform probabilistic analysis. However, they provide less accurate results and cannot handle large-scale problems due to their enormous computer processing time requirements. To overcome these limitations, this book presents a rather novel idea to perform probabilistic analysis by formally specifying the behavior of random systems in higher-order logic and using these formal models for verifying the intended probabilistic and statistical properties in a computer based theorem prover. The analysis carried out in this way is free from any approximation or precision issues due to the mathematical nature of the models and the inherent soundness of the theorem proving approach.

The book presents the higher-order-logic formalizations of foundational mathematical theories for conducting probabilistic analysis. These foundations mainly include measure, Lebesgue integration, probability, Markov chain, information and reliability theories. The most important notions in these theories have been defined in higher-order logic and most of their commonly used characteristics are then formally verified within the sound core of the HOL4 theorem prover. This formalization can be utilized to conduct accurate probabilistic analysis of real-world systems and for illustration purposes the book presents several examples.

The book starts with a brief introduction to the foundations (Chapters 1-3). We can divide the contents of the rest of the book into five main formalizations: Probability Theory (Chapters 4-5), Discrete-Time Markov Chains (DTMCs) (Chapters 6-8), Information Theory (Chapters 9-11), Reliability Theory (Chapter 12), and Wireless Sensor Network (WSN) Analysis (Chapters 13-14). The last four formalizations do not have any inter-dependency and thus can be read in any order after reading the first five chapters of the book. More details of the chapters are as follows:
Chapter 1 provides some background information related to the domains of probabilistic analysis and traditional analysis methods, like paper-and-pencil methods, simulation, and computer algebra systems. The intent is to introduce the foundations that we build upon in the rest of the manuscript.

Chapter 2 provides a general overview of formal verification methods. In particular, the two most commonly used formal methods (i.e., model checking and theorem proving) are introduced along with some examples. The chapter also includes some convincing arguments about using higher-order-logic theorem proving for conducting probabilistic analysis.

Chapter 3 presents the proposed methodology followed throughout the book for conducting probabilistic analysis including an overview about the HOL4 theorem prover, which is the main tool of focus in this book. The main reasons for this choice include the availability of foundational probabilistic analysis formalizations in HOL4 along with a very comprehensive support for real and set theoretic reasoning. The chapter also provides some of the frequently used HOL4 symbols in this manuscript.

Next, Chapter 4 provides the higher-order-logic formalization of the foundational theories of measure and Lebesgue integration. These theories are based on extended-real numbers (real numbers with $\pm\infty$). This allows us to define sigma-finite and even infinite measures and handle extended-real-valued measurable functions. It also allows us to verify the properties of the Lebesgue integral and its convergence theorems for arbitrary functions.

We build upon the higher-order-logic foundations, presented in the last chapter, to formalize, in Chapter 5, the probability theory in higher-order logic. This chapter also includes the formalizations of statistical properties, like expectation and variance, as well as conditional probability, and provides our first example of formal probabilistic analysis of the Heavy Hitter problem. Here the Heavy Hitter problem is formalized in higher-order logic and based on this formalization; some of its commonly used properties are formally verified.

In Chapter 6, we build upon the formalizations of the last two chapters and provide the higher-order-logic formalizations of Discrete-Time Markov Chains (DTMCs) and stationary distributions. These results are then used to conduct the formal probabilistic analysis of a binary communication channel and the Automatic Mail Quality Measurement (AMQM) protocol. These examples illustrate how to construct formal Markovian models of the given system and how to analyze it within a theorem prover. A comprehensive discussion about the comparison of model checking and theorem proving for formal probabilistic analysis is also included in this chapter.

Chapter 7 extends the DTMC formalization of Chapter 6 and presents the formalizations of classified states and classified DTMCs. We then use these formalizations to formally verify long-term properties, such as positive transition probability and
convergence. These mathematical foundations are then used to analyze some real-world applications, namely a Least Recently Used (LRU) Stack model, the Birth-Death process, and a memory contention problem in microprocessors.

In Chapter 8, we again utilize the DTMC formalization of Chapter 6 to formalize Hidden Markov Models (HMM), which are the core concept for formally evaluating the probability of the occurrence of a particular observed sequence and finding the best state sequence to generate given observation. Besides the formalization of HMMs and the formal verification of their well-known properties, we also introduce some ideas about automating the formal reasoning about HMM-related properties. In order to present the usefulness of the formalization of HMM and automatic verification ideas, we provide the formal analysis of a DNA (Deoxyribon Nucleic Acid) sequence in this chapter as well.

Chapter 9 presents the formalization of information measures by building upon the formalizations of measure and probability theories, presented in Chapters 4 and 5. In particular, Chapter 9 presents the formalization of the Radon-Nikodym derivative, the Kullback-Leibler divergence, mutual information, and conditional mutual information. We also use these foundations to formalize two new measures of information leakage (i.e., information leakage degree and conditional information leakage degree).

Chapter 10 extends the formalization of Chapter 9 to present the formalization of information flow metrics, namely Min-Entropy and Belief Min-Entropy. We then utilize these formalizations to provide an approach for the formal analysis of information flow within the sound core of a theorem prover. As an illustrative example, the chapter also includes the formal information flow analysis for a set of channels in cascade.

In Chapter 11, we build upon the mathematical foundations presented in Chapters 9 and 10 to present the formalization of data compression or source encoding. Moreover, the chapter also presents the formalization of Chebyshev and Markov inequalities that provide estimates of tail probabilities, and the weak law of large numbers. These formalizations are followed by the information theoretic analysis of a single mix channel and one-time pad, which are used as illustrative examples for the formal information theoretic analysis of real-world systems in a theorem prover.

Chapter 12 presents the formalization of reliability theory foundations, based on the foundations, presented in Chapters 4 and 5. In particular, we formalize the Cumulative Distribution Function (CDF) and survival function and their respective properties. Moreover, the chapter provides the formalization of Reliability Block Diagrams (RBDs), such as series, parallel, series-parallel, and parallel-series. These reliability foundations are then used for the formal reliability analysis of two applications (i.e., a cyber network and an oil and gas pipeline).
Chapter 13 focuses on the formal probabilistic analysis of Wireless Sensor Networks (WSNs) and presents the formalization of their coverage-based k-set randomized scheduling. We first formalize the random variables, which are required for this formalization. This is followed by the formalization of the network coverage intensity. As an illustrative case study, we present the formal analysis of a WSN for forest fire detection.

Chapter 14 provides the formalization of the detection aspect of WSNs using the randomized scheduling of nodes. Based on the probability theory, described in Chapter 5, we first formally reason about the intrusion period of any occurring event. This characteristic is then built upon to develop the fundamental formalizations of the key detection metrics: the detection probability and the detection delay. For illustration purposes, we formally analyze the detection performance of a WSN deployed for border security monitoring.

Finally, a conclusion concludes the manuscript by presenting a summary of all chapters and sharing some interesting future directions of research.

The target readers of this book are engineers and scientists working in the domains of system analysis and formal methods. These system analysis experts would be able to learn an emerging trend and apply it to their problems at hand to improve analysis results with more accuracy. On the other hand, the formal methods users would be able to learn the foundations of probabilistic analysis and extend them to broaden the scope of formal probabilistic analysis using theorem proving. The target audience is expected to be familiar with the notions of logic and some familiarity with the domains of theorem proving and automated reasoning is also recommended.

The whole idea of using theorem proving for conducting probabilistic analysis of engineering systems with continuous elements was generated during the Ph.D. research (2004-08) of Dr. Osman Hasan, who was working under the supervision of Prof. Sofiène Tahar. Since then, there have been many new developments in this area by both Prof. Tahar and Dr. Hasan resulting in over 30 research publications and 3 completed and 4 ongoing Ph.D. theses. This book thus summarizes all of this decade-long work in a concise manner.

Osman Hasan  
National University of Sciences and Technology (NUST), Pakistan

Sofiène Tahar  
Concordia University, Canada

February 6, 2015