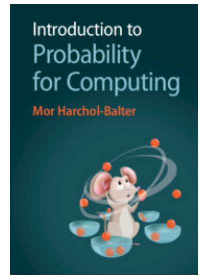
Review of [1]



Introduction to Probability for Computing
Mor Harchol-Balter
Cambridge University Press, 2023
Hardcover, 555 pages, $69.99

Review by
Sarvagya Upadhyay (supadhyay@fujitsu.com)
Fujitsu Research of America
4655 Great America Way, Suite 410
Santa Clara CA 95054, USA

1 Overview

Probability theory provides a rigorous mathematical foundation for analyzing and modeling uncertainty. It is a fundamental tool with widespread applications across numerous scientific disciplines. Its applications extend far beyond the multitude of science disciplines, playing a central role in two of the most prominent industrial sectors of today – finance and technology. In finance, sophisticated stochastic models are utilized for evaluating complex financial instruments and understanding market dynamics. In machine learning and statistics, probability is deeply embedded in models where decisions are not made deterministically, but instead are guided by the likelihoods of various outcomes – enabling systems to learn from data and adapt under uncertainty.

This book is a gentle introduction to probability theory from a computing perspective. It covers basic concepts along with several computing applications of the topic.

2 Summary of Contents

This book is structured into eight distinct parts. The initial three parts provide a comprehensive foundation in probability theory, encompassing topics such as discrete and continuous random variables. The subsequent five parts are dedicated to diverse applications of probability theory. As outlined in the preface, these sections may be utilized to develop curricula for four separate courses with an emphasis on applied probability.

Part I: Fundamentals and Probability of Events This part lays the mathematical groundwork essential for the study of probability theory. It is divided into two chapters. The first chapter revisits several fundamental mathematical concepts such as series, limits, integrals, and asymptotic notations. These tools are crucial for both understanding and solving problems in subsequent chapters. The second chapter delves into the basic principles of probability theory, introducing key topics such as the definition of sample spaces and events, the assignment of probabilities to events, conditional probability, independence, and Bayes' law. Throughout this part, carefully chosen examples are used to highlight important nuances. The structure and pedagogical approach adopted here perfectly sets up the presentation style of remainder of the book.

Part II: Discrete Random Variables This part of the book consists of four chapters, each dedicated to the study of discrete random variables. In Chapter 3, the formal definition of discrete random variables is introduced. This chapter also presents several commonly encountered discrete probability distributions such as the Bernoulli, Binomial, Geometric, and Poisson distributions. The next two chapters extend this foundation by examining moments of random variables, including expectations and linearity of expectations, variances, and higher-order moments. These concepts are essential for quantifying the central tendency and variability of random phenomena. Building upon this, the fifth chapter introduces the sum of random variables, particularly useful in contexts like algorithm analysis and queuing systems. It also introduces the fundamental tail inequalities such as Markov's and Chebyshev's inequalities, which provide probabilistic bounds on the likelihood of extreme outcomes. In addition, the concepts of stochastic dominance and Jensen's inequality are discussed, offering powerful tools for comparing random variables and analyzing convex functions of them. The final chapter of this part introduces the z-transform, an analytical tool widely used in the analysis of discrete-time systems. The author demonstrates how z-transforms can be effectively employed to compute moments and solve recurrence relations.

Part III: Continuous Random Variables As the author notes, much of the material presented in this part parallels the concepts introduced earlier in the context of discrete random variables, but is now extended to the continuous domain. This section comprises five chapters, each focusing on different aspects of continuous random variables and their applications. The first two chapters lay the theoretical foundation. The opening chapter explores continuous random variables drawn from a single distribution, introducing the probability density function (PDF), cumulative distribution function (CDF), and methods for computing expectations and variances. The following chapter expands on this by examining jointly distributed continuous random variables, discussing marginal and conditional distributions, and properties of moments. The proof of these properties are deferred to the exercises at the end of the chapter. Chapters 9 and 10 are devoted to two of the most widely used continuous distributions in both theory and practice: the Gaussian (Normal) and Pareto distributions. Chapter 9 also introduces one of the most important results in probability theory, the Central Limit Theorem (CLT), which underpins much of statistical inference and analysis. Chapter 10 begins with a particularly engaging introduction, a personal anecdote from the author. The story, centered around the author attending an operating systems course, creates a memorable entry point into the material. I must add that the author's sentiment resonated with me! The final chapter of this part introduces the Laplace transform, a powerful analytical tool for continuous random variables. The author demonstrates how Laplace transforms can be used to compute moments much like the z-transform is used for discrete random variables.

Part IV: Computer Systems Modeling and Simulations This part consists of three chapters, each focused on foundational concepts in stochastic modeling and random process simulation. It begins in Chapter 12 with a review of the exponential distribution, a key continuous distribution widely used in modeling the time between independent events. Building on this, the chapter introduces the Poisson process, where the author presents a few of important properties of such processes. The subsequent chapter delves into two primary techniques for generating instances of random variables from arbitrary distributions. The inverse transform method is introduced first, demonstrating how to transform uniformly distributed random numbers into samples from a desired distribution. This is followed by the accept-reject method, which is particularly useful when the inverse of the cumulative distribution function is difficult or impossible to compute directly. The final chapter of this part shifts the focus to event-driven simulation, a technique used to model systems where state changes occur at discrete points in time, often triggered by stochastic events. Within this framework, the author introduces key definitions from queuing theory. While this chapter provides an introductory overview, the author provides a more in-depth treatment of the topic in Chapter 27.

Part V: Statistical Inference This part of the book is conceptually independent from the preceding parts and shifts the focus from probability theory to the domain of statistical inference. The central aim here is to estimate unknown parameters based on observed data—a fundamental task in data analysis, machine learning, and many areas of applied computing. This part comprises three chapters, each systematically building the foundation for understanding and applying different estimation techniques. The first chapter introduces the most widely used point estimators, particularly the sample mean and variance. These estimators serve as intuitive and mathematically tractable tools for summarizing data and making inferences about the underlying distribution. Their properties, such as unbiasedness and consistency, are discussed to highlight their effectiveness in various practical settings. The second chapter delves into Maximum Likelihood Estimation (MLE), one of the most powerful and widely applied methods for parameter estimation. The author explains the importance of MLE through various detailed examples. The final chapter in this part, Chapter 17, introduces estimation from a Bayesian perspective, where parameters are treated as random variables with prior distributions. The chapter presents two key Bayesian estimators: the Maximum A Posteriori (MAP) estimator, which identifies the most probable parameter value given the observed data and prior beliefs, and the Minimum Mean Squared Error (MMSE) estimator, which minimizes the expected squared error between the estimate and the true parameter.

Part VI: Tail Bounds and Applications This part of the book, composed of three chapters, provides a deeper and more rigorous exploration of the tail behavior of random variables, a topic that was introduced in a preliminary form back in Chapter 5 through general tail inequalities such as Markov's and Chebyshev's bounds. Here, the author significantly extends this discussion by introducing stronger and more refined probabilistic bounds that are essential in both theoretical analysis and practical applications. The first chapter in this part, Chapter 18, focuses on two of the most powerful concentration inequalities in probability theory: the Chernoff bound and the Hoeffding inequality. These bounds provide exponentially decreasing probabilities for large deviations of sums of independent random variables from their expected values. Building on this, the next chapter demonstrates how tail bounds can be used to derive confidence intervals for statistical estimates. These results are particularly valuable in situations where one needs to quantify uncertainty in empirical observations. To illustrate the implications of these techniques, the author presents a series of classic balls-and-bins problems (including exercises). The final chapter in this part applies the previously developed tools to the domain of hashing algorithms, a fundamental component in computer science. By leveraging both tail bounds and balls-and-bins analysis, the author shows how to estimate the probability of hash collisions and estimate the size of hash buckets.

Part VII: Randomized Algorithms This part of the book is devoted to the study of randomized algorithms, a powerful class of computational techniques that incorporate randomness as a core component in algorithm design. The part comprises three chapters, each focusing on a distinct type of randomized algorithm and its applications. The discussion begins in Chapter 21, which introduces Las Vegas algorithms – randomized algorithms that always produce the correct result, but whose runtime is a random variable that depends on the outcomes of internal coin tosses. These algorithms are particularly useful when correctness is of paramount importance. The chapter presents two classic examples of this algorithmic variety: randomized quicksort and randomized median-finding. The subsequent chapter turns to Monte Carlo algorithms, which differ from Las Vegas algorithms in that their runtime is typically fixed or bounded, but the correct output can only be obtained with high probability. The author provides a variety of examples throughout the main text and exercises, including randomized matrix multiplication checking, the MAX-CUT and MIN-CUT problems. The final chapter in this part addresses a classic problem in theoretical computer science: primality testing. The chapter culminates with a detailed presentation of the Miller-Rabin primality test, a probabilistic algorithm that determines whether a number is prime with high probability.

Part VIII: Discrete-Time Markov Chains The final part of the book is dedicated to one of the most versatile and mathematically rich topics in probability theory: Markov chains. These models, which describe systems that transition between states in a memoryless fashion, have profound applications across a wide array of disciplines—including computer science, operations research, economics, and biology. While the book primarily emphasizes application-oriented learning, this section provides a meaningful foray into the theory of Markov chains to support and deepen that practical understanding. Of all the sections in the book, this one is arguably the most mathematically intensive. In Chapter 24, finite-state discrete-time Markov chains are introduced. Key concepts such as transition probability matrix, stationary distributions, and limiting distributions are developed. The author also shows that, under certain conditions, the stationary and limiting distributions coincide in the finite-state case. In the next chapter, we delve more deeply into the theoretical properties of finite-state Markov chains. It introduces crucial concepts such as irreducibility and aperiodicity, and explains how these properties link to the existence of limiting distributions. The chapter also explores topics such as mean return times to a given state and long-run time averages. A highlight of this chapter is the application of Markov chains to the PageRank algorithm used by Google. The next chapter shifts focus to infinite-state discrete-time Markov chains. Since the theoretical framework is more complex, the author focuses on prioritizing intuition and conceptual clarity over technical rigor. The concluding chapter of both this part and the book, Chapter 27, is devoted to queueing theory, an application area for Markov chains. This chapter synthesizes concepts developed throughout the part to show how stochastic processes can model and analyze waiting lines and service systems. Topics such as Little's Law and performance metrics for queues are introduced. Admittedly, the material here is dense and can be challenging, especially for readers like me who are unfamiliar with the subject. However, the author's pedagogical style helps in going through this chapter.

3 Evaluation and Opinion

In contrast to many traditional probability texts, this book deliberately steers away from overly abstract or purely theoretical treatments. Instead, it emphasizes practical insight and real-world intuition, aligning closely with the author's goal of delivering an application-oriented approach to probability. Drawing on more than two decades of teaching experience, the author brings a distinct pedagogical style to the material—one that is rich in illustrative examples, deeply intuitive, and highly accessible. The book is written in a conversational yet rigorous manner, making it suitable for a broad audience without compromising on the mathematical depth required for serious study.

For readers used to traditional mathematical books, this approach of mixing mathematical concepts with real-world applications and examples might seem a bit odd. However, as one progresses through the chapters, these practical illustrations become invaluable. There are plenty of exercises and examples, which are thoughtfully designed to reinforce the material covered in the chapters. They are worth doing!

One of the standout features of the book is its thoughtfully modular structure. For instance, rather than covering discrete and continuous random variables into two big chapters, the author opts to divide them into nine concise chapters. This approach makes the material more digestible, especially for beginners and practitioners who may be revisiting probability concepts. It seems that the aim of such conciseness is to allow the readers to focus on specific concepts without feeling overwhelmed. The content of most chapters and their exercises can be covered easily in a day. This structure not only aids in building a solid foundational understanding but also encourages readers to progress through the material with confidence.

This book is a great resource for undergraduate students seeking an applied perspective on probability. Among the probability books that I've read, this one has been the most engaging one. I thoroughly enjoyed reading the book. This book is borne out of a *Probability and Computing* class that Harchol-Balter has offered for over two decades. Given the presentation style of the book, I can only wish that I had the opportunity to attend one of her classes.