Preface

Embarking on the study of probability is both an opportunity and a challenge. This course represents a critical foundation for all subsequent studies in econometrics and data science. As we navigate an era of big data and artificial intelligence, proficiency in probability theory is indispensable. It not only builds the foundations for higher level courses but also cultivates a mindset geared towards probabilistic reasoning, which is essential for mitigating cognitive biases and making more informed decisions. Human cognition is often not well-equipped to handle probabilistic thinking. As we will see, human intuition and heuristics are mostly wrong about probabilistic events. This course seeks to provide a mathematical framework for properly understanding and applying probability.

Probability theory is more than just a mathematical discipline; it is a vital tool for making sense of uncertainty in the real world. Consider the myriad of questions that we encounter daily: Will it rain tomorrow? What is the expected return on an investment? What are the odds of a particular political party winning an election? How can a business optimize its customer service strategy when customer arrival times are unpredictable? Probability theory provides a scientific approach to answering these questions, enabling us to model and analyze uncertainties with mathematical tools.

However, this journey is not without its difficulties. For freshmen, particularly those new to calculus and linear algebra, the course presents a steep learning curve. The breadth of new concepts—such as random variables, expectations, and various distributions—can be overwhelming if encountered for the first time. Additionally, the use of advanced calculus, particularly integrals, may pose challenges for those who are still familiarizing themselves with these mathematical tools.

Despite these challenges, the rewards of studying probability are substantial.

Gaining a deep understanding of probability will not only enhance your knowledge base but also fundamentally transform your approach to problem-solving. The principles you will learn are applicable to a wide range of fields beyond econometrics and data science, including engineering, finance, and social sciences. You will learn the tools to approach these problems systematically and make informed decisions based on statistical evidence and probability. So be prepared for a challenging and rewarding journey!

Learning objectives

- **Review fundamental concepts:** Revisit the probability and calculus concepts learned in high school to ensure a solid foundation for more advanced topics.
- **Understand core probability theory:** Gain a thorough understanding of key concepts and theorems in probability theory, including random variables, expectations, covariances, and so on.
- **Develop probabilistic thinking:** Learn to approach problems with a probabilistic mindset and use random variables to describe and analyze uncertain outcomes.
- **Model real-world events:** Identify and apply important probability distributions to model and interpret real-world phenomena effectively.
- **Enjoy and have fun:** Discover and appreciate the inherent elegance of mathematics and the beauty of probability theory.

Study tips for new college students

- **Limit electronic distractions:** While digital tools like slides and tablets are convenient, traditional paper and pencil methods remain the most effective way to engage with and learn mathematics. Writing out problems and solutions helps reinforce concepts and improve retention.
- **Focus on key concepts:** College courses are often much more intensive than high school classes, and it is not feasible to master every detail. Concentrate on understanding the core ideas and principles, and don't get overwhelmed by the technical details.

- **Understand the "why":** In mathematics, understanding the underlying reasons and logic behind methods is more important than just knowing how to do computations. The "why" helps you grasp the broader implications and applications of the techniques you learn.
- **Gain practical experience:** Although this course emphasizes theoretical understanding and does not require programming, experimenting with statistical software such as R can be highly beneficial. Practical experience with data manipulation and analysis will enhance your comprehension and stimulate interest in the subject.
- **Engage with the material:** I will strive to make the course engaging and less boring. However, if this course is not your primary interest, focus on the aspects of the material that intrigue you. Try to have a general impression of the major concepts even though you do not remember any detail.
- **Exams are important, but more important is to enjoy the course.**

Textbooks and references

- DeGroot, M.H., and Schervish, M.J. (2011). *Probability and Statistics.*
- Hong, Y.M. (2017). *Probability Theory and Statistics for Economists.*

These textbooks are recommended references for this course, but you are not required to read them in their entirety. Focus on the lecture notes and exercises provided, and use these textbooks as supplementary resources as needed.

This content of this book is organized or follows. We start with probabilities based on counting, which should be familiar to high school graduates. Though rudimentary, they often yield surprising results, as rigorous calculations frequently challenge our intuitions about the likelihood of events. Special emphasis is placed on conditional probability, as conditional thinking is crucial both in academic studies as well as in daily life.

Next, we introduce the core concept of the random variable, which forms the foundation of all probability distributions and statistical theory. Random variables are essential tools that allow us to mathematically model uncertainty. We introduce two types of random variables: discrete and continuous. We begin with discrete random variables because they do not require calculus, offering a smoother learning curve for beginners. Key concepts such as expectations, variance, and covariance are introduced alongside well-known discrete distributions such as the Binomial, Geometric, and Poisson distributions. This arrangement ensures that students can grasp these important concepts without being overwhelmed by calculus. We also demonstrate how these fundamental distributions can be applied to solve real-world problems.

Following this, we move on to continuous distributions. We will see that the formulas from discrete distributions extend naturally to continuous distributions with the aid of calculus—essentially replacing summation with integration. We cover some of the most important continuous distributions, such as the Normal, Exponential, and Gamma distributions, and explore the interconnections between them. We also extend the concepts of expectations, variance, and joint distributions to their continuous forms.

The book concludes with a discussion on sampling and statistical inference. Since we cannot observe entire distribution, it becomes necessary to infer distribution properties from finite samples. We introduce two of the most important theorems in probability and statistical theory—the Law of Large Numbers and the Central Limit Theorem. The breadth and generality of these theorems are remarkable. But their most significant contribution to statistical applications is they allow us to gauge how close our sample estimates are to the true parameter values. The final chapter also includes a brief discussion on estimator accuracy, confidence intervals, and hypothesis testing. These topics are introduced briefly, as they serve primarily to prepare students for more advanced courses, such as econometrics.

The chapters are organized logically, with each chapter building on the knowledge presented in the previous ones. Therefore, it is recommended to follow the sequence of chapters rather than reading them independently. However, advanced readers who are already familiar with the topics may feel free to skip between chapters as needed. This manuscript is written tersely, serving as a skeleton to complement lecture materials. It is not intended as a substitute for lectures or comprehensive textbooks. Students who wish to learn the course material solely by reading are encouraged to consult a formal textbook.

This manuscript is a preliminary version, and while efforts have been made to ensure accuracy, errors may still be present. Your feedback on any mistakes or inaccuracies is greatly appreciated and will help improve the material.

Contents