PREFACE

We believe that many foundational ideas of Probability and Statistics are best understood when their natural connection is emphasised. We feel that the interested student should learn the mathematical rigour of Probability, the motivating examples and techniques from Statistics, and an instructive technology to perform computations relating to both in an inclusive manner. These formed our main motivations for writing this book. We have chosen to use the R software environment to demonstrate an available computational tool.

The book is intended to be an undergraduate text for a course on Probability Theory. We had in mind courses such as the one year (two semester) Probability course at many universities in India such as the Indian Statistical Institute or Chennai Mathematical Institue, or a one semester (or two quarter) Probability course as is commonly offered as an upper division, post-calculus elective at many North American universities. The Statistics material and the package R are introduced so as to emphasise motivations and applications of the probabilistic material. We assume that our readers are well-versed in calculus, have a basic understanding of the theory of sets and functions, combinatorics, and proof techniques, and have at least a passing awareness of the distinction between countable and uncountable infinities. We do not assume any particular experience of Linear Algebra or Real Analysis.

In Chapter 1 of this book we begin with an introduction to Outcomes, Sample Space, Events and the axiomatic definition of Probability. Then we discuss the concepts of conditional probability, independence and Bayes' Theorem. We conclude this chapter with a basic introduction to R. R is a Free Open Source software environment that runs on all major software platforms, and instructions to download and install it are available at `https://www.r-project.org/`.

We begin Chapter 2 by applying the notion of independence to repeated trials (Bernoulli Trials) and discuss the Binomial and Geometric distributions. We introduce the Poisson distribution as a limiting approximation of the Binomial. We conclude this section with a discussion on Sampling with and without replacement. The Hypergeometric Distribution is thus introduced here and we prove its approximation to the Binomial. Throughout this chapter and later in the book we provide the R code for calculating the probabilities associated with common distributions.

In Chapters 3 and 4 we introduce discrete random variables (functions on a sample space whose range is countable) and related concepts. In Chapter 3, we define the probability mass function, distribution function, and independence for random variables. We introduce the Multinomial distribution and show the memoryless property of the Geometric random variable. The chapter concludes by providing a method to compute the distributions of functions of one and several random variables, defining the concept of joint distribution along the way. In Chapter 4, we define Expectation, Variance, Covariance, Conditional Expectation and Conditional Variance for discrete random variables. Results involving these quantities for standard distributions are presented (and proved) as well. We also state and prove the Markov and Chebyshev inequalities along with the notion of standardising random variables to mean zero and variance one.

Working with uncountable spaces and understanding the probability density function of an absolutely continuous random variable are challenging without assuming a background in Real Analysis but we make a modest attempt towards this in Chapter 5. We begin with a description of uncountable sample spaces. After having described events in a temporary manner in Chapter 1 we provide a precise definition here but comfort the reader that we shall avoid the most general events and at most consider countable union/intersection of intervals. This allows us to be fairly rigorous with random variables having piecewise continuous probability density functions using results from basic calculus. After this we imitate the program conducted in Chapter 3. Standard distributions such as Uniform, Exponential, and Normal are discussed. While computing densities of sums and ratios of independent random variables we introduce the Gamma distribution and use it to derive the Beta distribution as an example of ratio of dependent Gamma random variables.

In Chapter 6 we define Variance, Covariance, Conditional Expectation and Conditional Variance for continuous random variables and summarise their properties. Moment generating functions for random variables are defined. At this point, to respect the minimal background assumption on our reader we state a few important results without proof such as the fact that the moment generating functions characterise distribution of a random variable. The chapter ends with a section on Bivariate Normal random variables. Here we have done all computations in this section without using Linear Algebra but the notational efficiency of using Linear Algebra is explained via exercises for the interested reader.

With the foundational ideas of Probability laid out we proceed in Chapter 7 with Sampling and Descriptive Statistics. The empirical distribution, the sample mean, variance and proportion are defined along with their properties. Simulation is used to develop intuition regarding sampling variability, and plots such as Histograms, Hanging Rootograms, and Q-Q Plots are introduced and illustrated using R.

Limit Theorems for Sampling Distributions discussed in Chapter 7 are the objective of Chapter 8. We begin with a brief description of multivariate joint densities and Order statistics. The t-distribution and χ^2 (chi-square) distributions are introduced in this chapter. The sample mean and variance from a normal population are discussed in relation to t and χ^2. We prove the Weak Law of Large numbers and the Central Limit Theorem for random variables possessing a moment generating function. We do state a more general version of the Central Limit Theorem and also the Strong Law of Large numbers, providing a proof of the latter in the Appendix. Along with R code we discuss the continuity correction and applications of the Central Limit Theorem via examples.

We end the book with two chapters focused solely on results and techniques from statistics. In Chapter 9 we discuss Estimation, Confidence Intervals and Hypothesis Testing. We briefly describe Method of Moments estimators followed by Maximum Likelihood Estimation. We then introduce Confidence Intervals as a methodology to work with when a single estimate might not suffice for the mean. In Hypothesis Testing we focus on tests involving Normal, t, χ^2, and F distributions. We describe the use of traditional lookup tables to compute probabilities as well as the use of R as a more flexible alternative. We also discuss critical values and the concept of rejection region. In Chapter 10 we discuss elementary results from simple linear regression involving one independent variable and one dependent variable. We discuss the Least Squares method and also prove results that will be useful for predicting new data from the given model. In the final section we discuss applications of Hypothesis testing in Regression. As in earlier chapters, computations are illustrated using R.

R code for most of the computations done are given in the book itself, and the reader should be able to reproduce and extend them easily. Code for figures are not given in the book, but are available at a website accompanying the book, which also contains additional material for readers who are interested in learning more about R.

The Appendix includes proofs of results such as Strong Law of Large Numbers and other proofs that are of interest but beyond the scope of the text.

Siva Athreya, *Deepayan Sarkar*, and *Steve Tanner*
April 25, 2016