1

Introduction

At present, almost all undergraduate curricula in engineering and applied sciences contain at least one basic course in probability and statistical inference. The recognition of this need for introducing the ideas of probability theory in a wide variety of scientific fields today reflects in part some of the profound changes in science and engineering education over the past 25 years.

One of the most significant is the greater emphasis that has been placed upon complexity and precision. A scientist now recognizes the importance of studying scientific phenomena having complex interrelations among their components; these components are often not only mechanical or electrical parts but also 'soft-science' in nature, such as those stemming from behavioral and social sciences. The design of a comprehensive transportation system, for example, requires a good understanding of technological aspects of the problem as well as of the behavior patterns of the user, land-use regulations, environmental requirements, pricing policies, and so on.

Moreover, precision is stressed – precision in describing interrelationships among factors involved in a scientific phenomenon and precision in predicting its behavior. This, coupled with increasing complexity in the problems we face, leads to the recognition that a great deal of uncertainty and variability are inevitably present in problem formulation, and one of the mathematical tools that is effective in dealing with them is probability and statistics.

Probabilistic ideas are used in a wide variety of scientific investigations involving randomness. Randomness is an empirical phenomenon characterized by the property that the quantities in which we are interested do not have a predictable outcome under a given set of circumstances, but instead there is a statistical regularity associated with different possible outcomes. Loosely speaking, statistical regularity means that, in observing outcomes of an experiment a large number of times (say n), the ratio m/n, where m is the number of observed occurrences of a specific outcome, tends to a unique limit as n becomes large. For example, the outcome of flipping a coin is not predictable but there is statistical regularity in that the ratio m/n approaches $\frac{1}{2}$ for either

Fundamentals of Probability and Statistics for Engineers T.T. Soong © 2004 John Wiley & Sons, Ltd
ISBNs: 0-470-86813-9 (HB) 0-470-86814-7 (PB)

heads or tails. Random phenomena in scientific areas abound: noise in radio signals, intensity of wind gusts, mechanical vibration due to atmospheric disturbances, Brownian motion of particles in a liquid, number of telephone calls made by a given population, length of queues at a ticket counter, choice of transportation modes by a group of individuals, and countless others. It is not inaccurate to say that randomness is present in any realistic conceptual model of a real-world phenomenon.

1.1 ORGANIZATION OF TEXT

This book is concerned with the development of basic principles in constructing probability models and the subsequent analysis of these models. As in other scientific modeling procedures, the basic cycle of this undertaking consists of a number of fundamental steps; these are schematically presented in Figure 1.1. A basic understanding of probability theory and random variables is central to the whole modeling process as they provide the required mathematical machinery with which the modeling process is carried out and consequences deduced. The step from B to C in Figure 1.1 is the induction step by which the structure of the model is formed from factual observations of the scientific phenomenon under study. Model verification and parameter estimation (E) on the basis of observed data (D) fall within the framework of statistical inference. A model

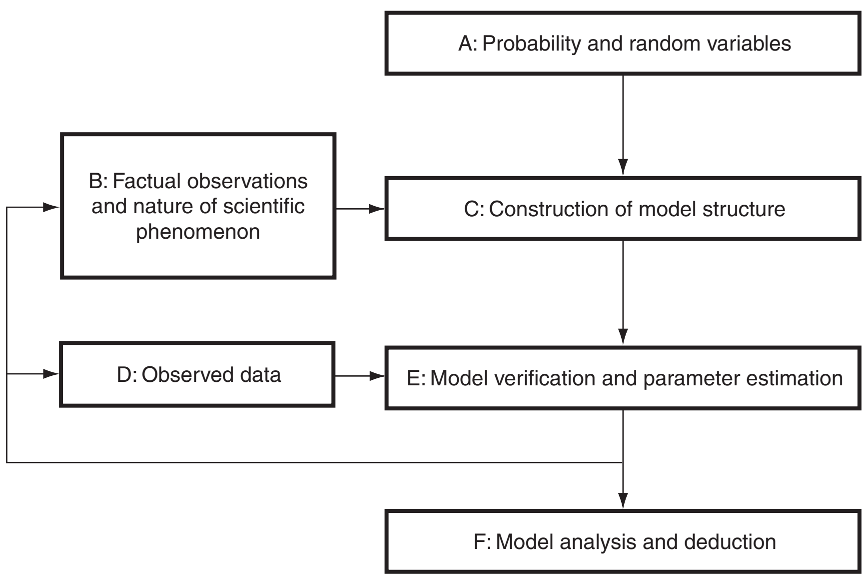

Figure 1.1 Basic cycle of probabilistic modeling and analysis

may be rejected at this stage as a result of inadequate inductive reasoning or insufficient or deficient data. A reexamination of factual observations or additional data may be required here. Finally, model analysis and deduction are made to yield desired answers upon model substantiation.

In line with this outline of the basic steps, the book is divided into two parts. Part A (Chapters 2–7) addresses probability fundamentals involved in steps $A \rightarrow C$, $B \rightarrow C$, and $E \rightarrow F$ (Figure 1.1). Chapters 2–5 provide these fundamentals, which constitute the foundation of all subsequent development. Some important probability distributions are introduced in Chapters 6 and 7. The nature and applications of these distributions are discussed. An understanding of the situations in which these distributions arise enables us to choose an appropriate distribution, or model, for a scientific phenomenon.

Part B (Chapters 8–11) is concerned principally with step $D \rightarrow E$ (Figure 1.1), the statistical inference portion of the text. Starting with data and data representation in Chapter 8, parameter estimation techniques are carefully developed in Chapter 9, followed by a detailed discussion in Chapter 10 of a number of selected statistical tests that are useful for the purpose of model verification. In Chapter 11, the tools developed in Chapters 9 and 10 for parameter estimation and model verification are applied to the study of linear regression models, a very useful class of models encountered in science and engineering.

The topics covered in Part B are somewhat selective, but much of the foundation in statistical inference is laid. This foundation should help the reader to pursue further studies in related and more advanced areas.

1.2 PROBABILITY TABLES AND COMPUTER SOFTWARE

The application of the materials in this book to practical problems will require calculations of various probabilities and statistical functions, which can be time consuming. To facilitate these calculations, some of the probability tables are provided in Appendix A. It should be pointed out, however, that a large number of computer software packages and spreadsheets are now available that provide this information as well as perform a host of other statistical calculations. As an example, some statistical functions available in Microsoft® Excel™ 2000 are listed in Appendix B.

1.3 PREREQUISITES

The material presented in this book is calculus-based. The mathematical prerequisite for a course using this book is a good understanding of differential and integral calculus, including partial differentiation and multidimensional integrals. Familiarity in linear algebra, vectors, and matrices is also required.