

Preface

This text develops introductory topics in probability and statistics with particular emphasis on concepts that arise in computer science. It starts with the basic definitions of probability distributions and random variables and elaborates their properties and applications. Statistics is the major application domain for probability theory and consequently merits mention in the title. Unsurprisingly, then, the text treats the most common discrete and continuous distributions and shows how they find use in decision and estimation problems. It also constructs computer algorithms for generating observations from the various distributions.

However, the text has a major subtheme. It develops in a thorough and rigorous fashion all the necessary supporting mathematics. This approach contrasts with that adopted by most probability and statistics texts, which for economy of space or for fear of mixing presentations of different mathematical sophistication, simply cite supporting results that cannot be proved in the context of the moment. With careful organization, however, it is possible to develop all the needed mathematics beyond differential and integral calculus and introductory matrix algebra, and this text purports to do just that.

Of course, as the book lengthens to accommodate the supporting mathematics, some material from the typical introduction to probability theory must be omitted. I feel the omissions are minor and that all major introductory topics receive adequate attention. Moreover, engagement with the underlying mathematics provides an opportunity to understand probability and statistics at a much deeper level than that afforded by mechanical application of unproved theorems.

Although the presentation is as rigorous as a pure mathematics text, computer science students comprise the book's primary audience. Certain aspects of most computer science curriculums involve probabilistic reasoning, such as algorithm analysis and performance modeling, and frequently students are not sufficiently prepared for these courses. While it is true that most computer science curriculums do require a course in probability and statistics, these courses often fail to provide the necessary depth. This text certainly does not fail in presenting a thorough grounding in elementary probability and statistics. Moreover, it seizes the opportunity to extend the student's command of mathematical analysis. This approach is different than that taken by other probability and statistics texts currently aimed at computer science

curriculums. The more rigorous approach does require more work, both from the student and from the instructor, but the rewards are commensurate.

The engineering sciences, like computer science, also tend to use texts that place more emphasis on mechanical application of results than on the mathematical derivation of such results. Consequently, engineering science students will also benefit from the deeper presentation afforded by this text. Nevertheless, the primary audience remains computer science students because many of the illustrative examples are computer science applications. Therefore, from this point forward, I assume that I am addressing a computer science student or instructor.

Computer science students typically follow a traditional curriculum that includes one or two terms of probability and statistics, which follow prerequisite courses in differential and integral calculus and linear algebra. Although these prerequisite courses do introduce limit processes and matrix transformations, they typically emphasize formulas that isolate applications from the underlying theory. For example, if we drain a swimming pool with a sinusoidal cross-section, we can calculate how fast the water level falls without invoking limit operations. We simply set up a standard differential ratio and equate it to the drain flow rate. Why this works is buried in the theory and receives less and less emphasis once a satisfactory collection of calculation templates is available. This text provides an opportunity to reconnect with the theoretical concepts of these prerequisite courses. As it probes deeper into the properties of probability distributions, the text puts these concepts to fruitful use in constructing rigorous proofs.

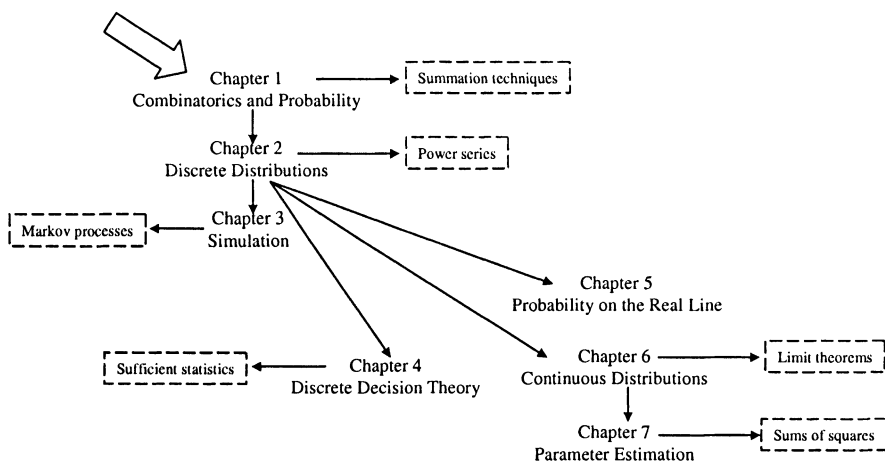
The book's ambient prose deals with the principal themes and applications of probability, and a sequence of mathematical support modules interrupts this prose at strategic junctures. With some exceptions, these modules appear as needed by the probability concepts under discussion. A reader can omit the modules and still obtain a good grounding in elementary probability and statistics, including philosophical interpretations of probability and ample exercise in the associated numerical techniques. Reading the support modules will, however, strengthen this understanding and will also arouse an appreciation for the mathematics itself.

The encapsulation is as follows. An appendix gathers selected topics from set theory, limit processes, the structure of the real numbers, Riemann-Stieltjes integrals, matrix transformations, and determinants. The treatment first reviews the material at an introductory level. The prepared reader will be familiar with these concepts from previous courses, but the results are nevertheless proved in detail. The less prepared reader will certainly find frequent recourse to the appendix, and the text provides pointers to the appropriate sections. However, even the prepared reader will benefit from the introductory presentations, which serve both as a review of proof technique and as an introduction to the argument style pursued in the main text. Upon completing an introductory review, the appendix then extends the topics as necessary to support the arguments that appear in the main body of the text. Therefore,

all chapters depend on the appendix for completeness. Even a reader well grounded in the aforementioned prerequisites can expect to spend some time mastering the specialized tools developed in the appendix.

The appendix, with its eclectic collection of review topics and specialized extensions, provides general mathematical background. There is need, however, for more specific supporting mathematics in connection with particular probabilistic and statistical concepts. Until perhaps halfway through the text, this supporting mathematics appears in mathematical interludes, which occur in each chapter. These interludes introduce particular results that are needed for the first time in that chapter. The first interlude deals with summation techniques, which are useful tools for the combinatoric problems associated with probability over equally likely outcomes. Others treat convergence issues in power series, stability features of Markov matrices, and sufficient statistics. Before taking up continuous distributions, however, it is appropriate to devote a full chapter to the mathematical issues that arise when one attempts to generalize discrete probability to uncountable sets and to the real line in particular. This chapter is actually a brief introduction to measure theory, and its logical place is just prior to the discussion of the common distributions on the real line. Two further interludes follow in subsequent chapters. They deal with limit theorems for continuous random variables and with decompositions of the sample variance. In short, the text exploits opportunities to introduce the mathematical analysis necessary to establish the basic results of probability theory. Moreover, the presentation clearly considers the mathematical analysis and the probability theory to be of equal importance.

The following sketch shows the dependencies among the chapters, with the understanding that portions of the appendix are prerequisite for any given path. The dashed boxes note the mathematical interludes within the chapters.



The reader can study the appendix in detail to ensure familiarity with all the background mathematics needed in the text, or can start immediately

with the probability discussions of Chapter 1 and refer to the appendix as needed. Because of its breadth, the appendix is more difficult to master in its entirety than the mathematical interludes of the introductory chapters. A reader who prefers that the material increase monotonically in difficulty should start with the introductory chapters and digress into the appropriate appendix sections as needed. When using the text to support a course, an instructor should follow a similar path.

As noted earlier, the intended audience is computer science students. Once past colored balls in numbered urns, which constitute the traditional examples in combinatoric problems, the text uses examples that reflect this readership. Client-server performance evaluation, for instance, offers many opportunities for probabilistic analysis. These examples should provide no difficulty for other readers, such as students from the engineering sciences, because the examples make no profound references to advanced concepts, but rather use generally accessible quantities, such as terminal response time, server queue length, error count per 1000 programming statements, or operation count in an algorithm. These examples are no more difficult than those in a more general probability text that ventures beyond the traditional urns, colored balls, and dice.

The requirements of the Computer Science Accreditation Board and the Accreditation Board of Engineering Technology (CSAB/ABET) include a one-semester course in probability and statistics. This text satisfies that requirement. In truth, it is sufficient for a full-year course because it not only develops the traditional introductory probability concepts but also includes considerable material on mathematical reasoning. For a one-semester course, the following selection is appropriate. Note that the topics lie along an acceptable dependency chain in the earlier diagram.

Appendix. Sections as referenced in the items below

Chapter 1. Combinatorics and Probability

Chapter 2. Discrete Distributions

Chapter 3-4. Simulation, Sections 3.1 to 3.3, or Discrete Decision Theory, Sections 4.1 and 4.2

Chapter 6. Continuous distributions, Sections 6.1, 6.3, and 6.4

Chapter 7. Parameter Estimation, Sections 7.1, 7.2, and 7.4

The one-semester abbreviation is possible because Chapters 3 and 4 present major applications of discrete probability and, in the interest of time, need only be sampled. Chapter 5 is advanced material that elaborates the difficulties in extending discrete probability to uncountable sample spaces. It is present for logical completeness and to answer the nagging question that occurs to many students: Is the introduction of a sigma-algebra really necessary in the general definition of a probability space? Consequently, the proposed

one-semester course omits Chapter 5 with minimal impact on subsequent material. Finally, Chapter 7 undertakes major applications of continuous probability and also admits partial coverage.

At the time of this writing, many computer science curriculums include only a first probability course. However, there is a recognized need for further study, at least in the form of an elective second course, if not in a required sequel to the introductory course. Anticipating that this increased attention will also expose the need for a more complete mathematical treatment of the material, I have provided unusually detailed excursions into supporting topics, such as estimation arguments with limits, properties of power series, and Markov processes.

Buttressed by these mathematical excursions, the text provides a thorough introduction to probability and statistics—concepts, techniques, and applications. Consequently, it offers a continuing discussion of the real-world meaning of probabilities, particularly when the frequency-of-occurrence interpretation becomes somewhat strained. Any science that uses probability must face the interpretation challenge. How can you apply a result that holds only in a probabilistic sense to a particular data set? The text also discusses competing interpretations, such as the credibility-of-belief interpretation, which might appear more appropriate to history or psychology. The goal is, of course, to remain continually in touch with the real-world meaning of the concepts.

Probability as frequency of occurrence over many trials provides the most compelling interpretation of the phenomenon. It is intuitively plausible, for example, that a symmetric coin should have equal chances of landing heads or tails. The text attempts to carry this interpretation as far as possible. Indeed, the first chapter treats the combinatorics arising from symmetric situations, and this treatment serves as a prelude to the formal definitions of discrete probability. As the theory accumulates layer upon layer of reasoning, however, this viewpoint becomes difficult to sustain in certain cases. When testing a hypothesis, for example, we attempt to infer the prevailing state of nature from sampled data. What does it mean to assign *a priori* probabilities to the possible states? This practice allows statisticians to incorporate expert judgment into the decision rules, but the assigned probabilities do not admit a frequency-of-occurrence interpretation. Rather, they reflect relative strength-of-belief statements about the possible states. As necessary, the text interrupts the technical development to comment on the precise real-world interpretation of the model. Although beautiful as abstract theory, probability and statistics are also rightly praised for their ability to deliver meaningful statements about the real world. Interpreting the precise intent of these statements should be a primary goal of any text.

A trend in modern textbooks, particularly those not addressed specifically to a mathematics curriculum, is to avoid the theorem-proof presentation style. This style can be sterile and detached, in the sense that it provides sparse context for the motivation or application of the theorems. Without

the theorem-proof style, on the other hand, arguments lose some precision, and there is a blurring of the line between the general result and its specific applications. I have adopted what I consider a middle ground. I maintain a running prose commentary on the material, but I punctuate the dialog with frequent theorems. Often, the theorem's proof is a simple statement: "See discussion above." This serves to set off the general results, and it also provides reference points for later developments. The ambient prose remains connected with applications and with the questions that motivate the search for new general results.

Plentiful examples, displayed in a contrasting typographical style, play a major role in compensating for the perceived coldness of the theorems. Incidentally, I should say that I do not find the theorems cold, even in isolation. But I am responding to the spirit of the age, which suggests that a theorem wrapped in an example is more digestible than a naked theorem.

The theorems also further a second ambition, noted above, which is to involve the reader more extensively in precise mathematical argument. An aspect of proofs that attracts major criticism is the tendency to display, out of thin air, an expression that magically satisfies all the required constraints and invites the algebraic manipulations necessary to complete the proof. I have tried to avoid this practice by including some explanation of the mysterious expression's origin. I must admit, however, that I am not always successful in this ploy. Sometimes an explanation adds nothing to a careful contemplation of the expression. In such cases, I am tempted to suggest that the reader reflect on the beauty of the expression, note how one part attaches to the known information while another extends toward the desired result, and view the expression as a unifying link, growing naturally from a study of the context of the problem in question. Instead, however, I fall back on the age-old practice: "Consider the following expression . . ." The reader should take these words as an invitation to pause and ponder the situation.

In summary, the text develops a main theme of probability and statistics, together with the mathematical techniques needed to support it. Since it is not practical to start with the Peano axioms, there are, however, some prerequisites. Specifically, the text's mathematical level assumes that the reader has mastered differential and integral calculus and has some exposure to matrix algebra. Nevertheless, acknowledging the mechanical fashion in which these subjects are taught these days, the text provides considerable detail in all arguments. It does assume, nevertheless, that readers have some familiarity with limiting operations, even if they do not have significant experience with the concept. For example, readers should be comfortable with l'Hôpital's rule for evaluating limits that initially appear to produce indeterminate results. By contrast, the text does develop the theory of absolutely convergent series to the point of justifying the interchange of summation order in double summations.

As another example, the readers should be generally conversant with power series, although the text develops this topic sufficiently to justify the term-by-term differentiation that is needed to recover parameters from a

random variable's moment generating function. The background appendix and the mathematical interludes should bridge the gap between prerequisite knowledge and that needed to establish all probability and statistical concepts encountered. They also serve to present a self-contained book, which is my preference when learning new material. A reader can always skip sections that are peripheral to the main point, but cannot as easily fill in omissions.

Some expositions consist of step-by-step procedures for solving a probabilistic or statistical problem. That is, they involve algorithms. For example, algorithms appear in sections concerned with computer simulations of probabilistic situations. In any case, algorithms in this text appear as mutilated C code, in the sense that I vary the standard syntax as necessary to describe the algorithm most clearly to a human reader. For instance, I use only two iterators, the while-loop and the for-loop, and in each case, the indentation serves to delineate the body of statements under repetition. The left fragment below, intentionally meaningless to focus attention on the code structure, must be reformulated as shown on the right to actually compile properly.

| | |
|---|---|
| <pre> while (X > Y) for (j = 1; j <=; j++) Y = Y - t[j]; t[j] = t[j] + 1.0; if (Y < 0.0) Y = 0.0; </pre> | <pre> while (X > Y) { for (j = 1; j <=; j++) { Y = Y - t[j]; t[j] = t[j] + 1.0; } if (Y < 0.0) Y = 0.0; } </pre> |
|---|---|

This practice amounts to omitting the opening and closing braces, and it is surprisingly efficient in reducing the code length. The price, of course, is that the code will not execute as written. I also omit variable declarations and simply use any variables that I need in an untyped manner. So a variable, or a function return, can be a scalar, a vector, a matrix, or a list. A particular usage is always clear from context, and these omissions, like those of scope enclosures, produce more compact code. In summary, the algorithms in this text are intended for human consumption. The emphasis is on the concept, not the detailed syntax of a programming language. Nevertheless, I have constructed all the algorithms in functional code and executed them to ensure that they perform as promised.

I emphasize that this text requires no C programming background. A C-based style finds frequent use in books and journals as a concise and precise algorithm presentation method for audiences that are not C literate. This widespread usage confirms the opinion that loose C structures are easily accessible to readers who do not have a background knowledge of C. Consequently, a reader with no previous C experience should not feel anxious about the algorithm descriptions in this text. Although readers must want to understand the algorithm and must expend the time to attend carefully the iterative processes in the C expression, they will find the algorithm comprehensible. Moreover, explanatory prose accompanies all C-structure descriptions.

Another feature that will be appreciated by today's students is the text's

early emphasis on discrete distributions. Most introductory probability books quickly proceed to the normal distribution and the central limit theorem because these tools enable certain computational approximations. I certainly do not wish to imply that the normal distribution or the central limit theorem is not important. These topics are developed in detail in the latter half of the book. However, much reasoning with discrete distributions can take place without replacing the discrete distribution with a more tractable normal distribution. This is especially true today when computers complete the mechanical, but sometimes tedious, computations. Working directly with discrete distributions develops a feeling for the subtleties of probability theory that can be overlooked in a rush to obtain approximate results. The many examples involving discrete distributions will sharpen a reader's capabilities for combinatorial reasoning and will exercise his or her skills in discrete mathematical argument. These skills are especially important in today's scientific world, which grows ever more digital, and therefore ever more discrete.

In keeping with this approach, the text develops simulation techniques and certain aspects of statistical inference directly after introducing a repertoire of discrete distributions but before developing continuous distributions. This means that certain proofs, which admit more elegant expositions with continuous tools, contain detailed arguments with limits. These arguments are conceptually similar to those associated with calculus (e.g., passing to an integral limit from a summation), but the text nevertheless presents more intermediate steps than one normally finds in a textbook

All definitions, theorems, and examples share a common numbering sequence within each chapter. To find Theorem 3.14, for example, you can use any definition, theorem, or example to direct your search forward or backward. I find it a considerable annoyance when these items are numbered separately. Of course, this means that there may be a Theorem 3.14 even though there is no Theorem 3.13. I try to make this dissonance more acceptable by placing the number before the theorem, definition, or example. Having located Theorem 3.14, for example, you find that it starts "3.14 Theorem. . . ." In this reading, Theorem 3.14 is an item, which happens to be a theorem. There is indeed an item 3.13, which may or may not be a theorem. In the traditional manner, tables and figures exhibit separate numbering sequences within each chapter.