INTUITIVE PROBABILITY AND RANDOM PROCESSES USING MATLAB®

INTUITIVE PROBABILITY AND RANDOM PROCESSES USING MATLAB®

STEVEN M. KAY
University of Rhode Island

Springer

Steven M. Kay
University of Rhode Island
Dept. of Electrical & Computer
Engineering
Kingston, RI 02881

Kay, Steven M., 1951-
Intuitive probability and random processes using MATLAB / Steven M. Kay.
p. cm.
ISBN-13: 978-0-387-24157-9 (acid-free paper)
ISBN-10: 0-387-24157-4 (acid-free paper)
ISBN-10: 0-387-24158-2 (e-book)
1. Probabilities--Computer simulation--Textbooks. 2. Stochastic processes--Computer simulation--Textbooks. 3. MATLAB--Textbooks. I. Title.

QA273.K329 2005
519.2'01'13--dc22

2005051721

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 11054474

springeronline.com

To my wife
Cindy,
whose love and support
are without measure

and to my daughters
Lisa and Ashley,
who are a source of joy

NOTE TO INSTRUCTORS

As an aid to instructors interested in using this book for a course, the solutions to the exercises are available in electronic form. They may be obtained by contacting the author at kay@ele.uri.edu.

Preface

The subject of probability and random processes is an important one for a variety of disciplines. Yet, in the author's experience, a first exposure to this subject can cause difficulty in assimilating the material and even more so in applying it to practical problems of interest. The goal of this textbook is to lessen this difficulty. To do so we have chosen to present the material with an emphasis on conceptualization. As defined by Webster, a *concept* is "an abstract or generic idea generalized from particular instances." This embodies the notion that the "idea" is something we have formulated based on our past experience. This is in contrast to a *theorem*, which according to Webster is "an idea accepted or proposed as a demonstrable truth". A theorem then is the result of many *other* persons' past experiences, which may or may not coincide with our own. In presenting the material we prefer to first present "particular instances" or examples and then generalize using a definition/theorem. Many textbooks use the opposite sequence, which undeniably is cleaner and more compact, but omits the motivating examples that initially led to the definition/theorem. Furthermore, in using the definition/theorem-first approach, for the sake of mathematical correctness multiple concepts must be presented at once. This is in opposition to human learning for which "under most conditions, the greater the number of attributes to be bounded into a single concept, the more difficult the learning becomes"[1]. The philosophical approach of specific examples followed by generalizations is embodied in this textbook. It is hoped that it will provide an alternative to the more traditional approach for exploring the subject of probability and random processes.

To provide motivating examples we have chosen to use MATLAB[2], which is a very versatile scientific programming language. Our own engineering students at the University of Rhode Island are exposed to MATLAB as freshmen and continue to use it throughout their curriculum. Graduate students who have not been previously introduced to MATLAB easily master its use. The pedagogical utility of using MATLAB is that:

1. Specific computer generated examples can be constructed to provide motivation for the more general concepts to follow.

[1] Eli Saltz, *The Cognitive Basis of Human Learning*, Dorsey Press, Homewood, IL, 1971.
[2] Registered trademark of TheMathWorks, Inc.

2. Inclusion of computer code within the text allows the reader to interpret the mathematical equations more easily by seeing them in an alternative form.

3. Homework problems based on computer simulations can be assigned to illustrate and reinforce important concepts.

4. Computer experimentation by the reader is easily accomplished.

5. Typical results of probabilistic-based algorithms can be illustrated.

6. Real-world problems can be described and "solved" by implementing the solution in code.

Many MATLAB programs and code segments have been included in the book. In fact, most of the figures were generated using MATLAB. The programs and code segments listed within the book are available in the file `probbook_matlab_code.tex`, which can be found at http://www.ele.uri.edu/faculty/kay/New%20web/Books.htm. The use of MATLAB, along with a brief description of its syntax, is introduced early in the book in Chapter 2. It is then immediately applied to simulate outcomes of random variables and to estimate various quantities such as means, variances, probability mass functions, etc. *even though these concepts have not as yet been formally introduced.* This chapter sequencing is purposeful and is meant to expose the reader to some of the main concepts that will follow in more detail later. In addition, the reader will then immediately be able to simulate random phenomena to learn through doing, in accordance with our philosophy. In summary, we believe that the incorporation of MATLAB into the study of probability and random processes provides a "hands-on" approach to the subject and promotes better understanding.

Other pedagogical features of this textbook are the discussion of discrete random variables first to allow easier assimilation of the concepts followed by a parallel discussion for continuous random variables. Although this entails some redundancy, we have found less confusion on the part of the student using this approach. In a similar vein, we first discuss scalar random variables, then bivariate (or two-dimensional) random variables, and finally N-dimensional random variables, reserving separate chapters for each. All chapters, except for the introductory chapter, begin with a summary of the important concepts and point to the main formulas of the chapter, and end with a real-world example. The latter illustrates the utility of the material just studied and provides a powerful motivation for further study. It also will, hopefully, answer the ubiquitous question "Why do we have to study this?". We have tried to include real-world examples from many disciplines to indicate the wide applicability of the material studied. There are numerous problems in each chapter to enhance understanding with some answers listed in Appendix E. The problems consist of four types. There are "formula" problems, which are simple applications of the important formulas of the chapter; "word" problems, which require a problem-solving capability; and "theoretical" problems, which are more abstract and mathematically demanding; and finally, there are "computer" problems, which are either computer simulations or involve the application of computers to facilitate analytical solutions. A complete solutions manual for all the problems is available to instructors from the author upon request. Finally, we have provided warnings on how to avoid common errors as well as in-line explanations of equations within the derivations for clarification.

The book was written mainly to be used as a first-year graduate level course in probability and random processes. As such, we assume that the student has had some exposure to basic probability and therefore Chapters 3–11 can serve as a review and a summary of the notation. We then will cover Chapters 12–15 on probability and selected chapters from Chapters 16–22 on random processes. This book can also be used as a self-contained introduction to probability at the senior undergraduate or graduate level. It is then suggested that Chapters 1–7, 10, 11 be covered. Finally, this book is suitable for self-study and so should be useful to the practitioner as well as the student. The necessary background that has been assumed is a knowledge of calculus (a review is included in Appendix B); some linear/matrix algebra (a review is provided in Appendix C); and linear systems, which is necessary only for Chapters 18–20 (although Appendix D has been provided to summarize and illustrate the important concepts).

The author would like to acknowledge the contributions of the many people who over the years have provided stimulating discussions of teaching and research problems and opportunities to apply the results of that research. Thanks are due to my colleagues L. Jackson, R. Kumaresan, L. Pakula, and P. Swaszek of the University of Rhode Island. A debt of gratitude is owed to all my current and former graduate students. They have contributed to the final manuscript through many hours of pedagogical and research discussions as well as by their specific comments and questions. In particular, Lin Huang and Cuichun Xu proofread the entire manuscript and helped with the problem solutions, while Russ Costa provided feedback. Lin Huang also aided with the intricacies of LaTex while Lisa Kay and Jason Berry helped with the artwork and to demystify the workings of Adobe Illustrator 10.[3] The author is indebted to the many agencies and program managers who have sponsored his research, including the Naval Undersea Warfare Center, the Naval Air Warfare Center, the Air Force Office of Scientific Research, and the Office of Naval Research. As always, the author welcomes comments and corrections, which can be sent to kay@ele.uri.edu.

Steven M. Kay
University of Rhode Island
Kingston, RI 02881

[3]Registered trademark of Adobe Systems Inc.

Contents

Chapter 1

Introduction

1.1 What Is Probability?

Probability as defined by Webster's dictionary is "the chance that a given event will occur". Examples that we are familiar with are the probability that it will rain the next day or the probability that you will win the lottery. In the first example, there are many factors that affect the weather—so many, in fact, that we cannot be certain that it will or will not rain the following day. Hence, as a predictive tool we usually assign a number between 0 and 1 (or between 0% and 100%) indicating our degree of certainty that the event, rain, will occur. If we say that there is a 30% chance of rain, we believe that if identical conditions prevail, then 3 times out of 10, rain will occur the next day. Alternatively, we believe that the *relative frequency* of rain is 3/10. Note that if the science of meteorology had accurate enough models, then it is conceivable that we could determine exactly whether rain would or would not occur. Or we could say that the *probability* is either 0 or 1. Unfortunately, we have not progressed that far. In the second example, winning the lottery, our chance of success, assuming a fair drawing, is just one out of the number of possible lottery number sequences. In this case, we are uncertain of the outcome, not because of the inaccuracy of our model, but because the experiment has been designed to produce uncertain results.

The common thread of these two examples is the presence of a *random experiment*, a *set of outcomes*, and the *probabilities* assigned to these outcomes. We will see later that these attributes are common to all probabilistic descriptions. In the lottery example, the experiment is the drawing, the outcomes are the lottery number sequences, and the probabilities assigned are $1/N$, where $N =$ total number of lottery number sequences. Another common thread, which justifies the use of probabilistic methods, is the concept of *statistical regularity*. Although we may never be able to predict with certainty the outcome of an experiment, we are, nonetheless, able to predict "averages". For example, the average rainfall in the summer in Rhode Island is 9.76 inches, as shown in Figure 1.1, while in Arizona it is only 4.40

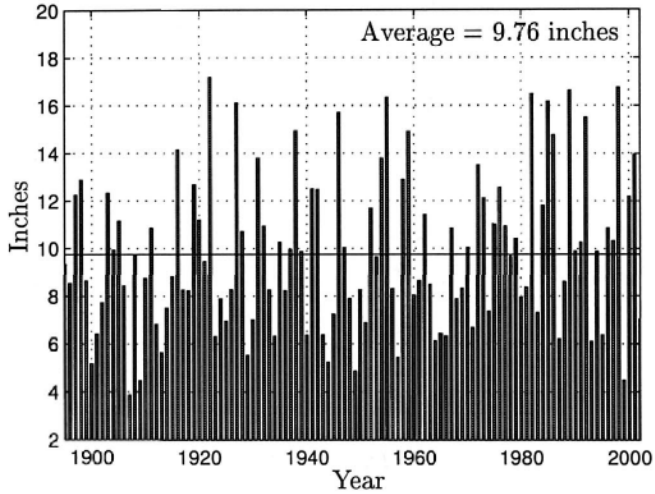


Figure 1.1: Annual summer rainfall in Rhode Island from 1895 to 2002 [NOAA/NCDC 2003].

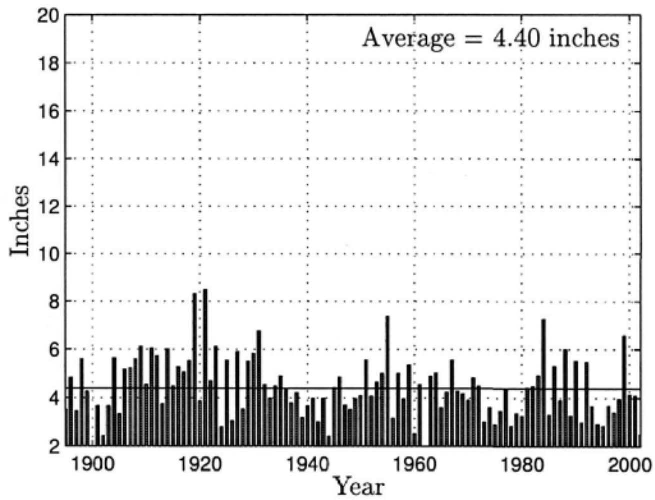


Figure 1.2: Annual summer rainfall in Arizona from 1895 to 2002 [NOAA/NCDC 2003].

inches, as shown in Figure 1.2. It is clear that the decision to plant certain types of crops could be made based on these averages. This is not to say, however, that we can predict the rainfall amounts for any given summer. For instance, in 1999 the summer rainfall in Rhode Island was only 4.5 inches while in 1984 the summer rainfall in Arizona was 7.3 inches. A somewhat more controlled experiment is the repeated tossing of a fair coin (one that is equally likely to come up heads or tails). We would expect about 50 heads out of 100 tosses, but of course, we could not predict the outcome of any one particular toss. An illustration of this is shown in Figure 1.3. Note that 53 heads were obtained in this particular experiment. This

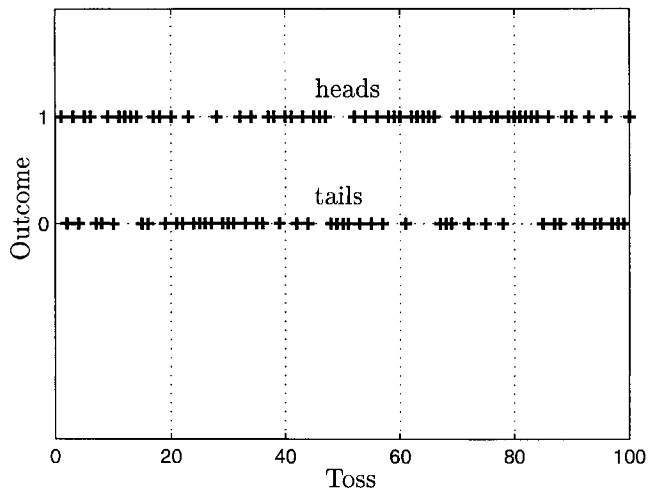


Figure 1.3: Outcomes for repeated fair coin tossings.

example, which is of seemingly little relevance to physical reality, actually serves as a good *model* for a variety of random phenomena. We will explore one example in the next section.

In summary, probability theory provides us with the ability to predict the behavior of random phenomena in the "long run." To the extent that this information is useful, probability can serve as a valuable tool for assessment and decision making. Its application is widespread, encountering use in all fields of scientific endeavor such as engineering, medicine, economics, physics, and others (see references at end of chapter).

1.2 Types of Probability Problems

Because of the mathematics required to determine probabilities, probabilistic methods are divided into two distinct types, *discrete* and *continuous*. A discrete approach is used when the number of experimental outcomes is finite (or infinite but countable as illustrated in Problem 1.7). For example, consider the number of persons at a business location that are talking on their respective phones anytime between 9:00 AM and 9:10 AM. Clearly, the possible outcomes are $0, 1, \ldots, N$, where N is the number of persons in the office. On the other hand, if we are interested in the length of time a particular caller is on the phone during that time period, then the outcomes may be anywhere from 0 to T minutes, where $T = 10$. Now the outcomes are infinite in number since they lie within the interval $[0, T]$. In the first case, since the outcomes are discrete (and finite), we can assign probabilities to the outcomes $\{0, 1, \ldots, N\}$. An equiprobable assignment would be to assign each outcome a probability of $1/(N+1)$. In the second case, the outcomes are continuous (and therefore infinite) and so it is not possible to assign a nonzero probability *to each outcome* (see Problem 1.6).

We will henceforth delineate between probabilities assigned to discrete outcomes and those assigned to continuous outcomes, with the discrete case always discussed first. The discrete case is easier to conceptualize and to describe mathematically. It will be important to keep in mind which case is under consideration since otherwise, certain paradoxes may result (as well as much confusion on the part of the student!).

1.3 Probabilistic Modeling

Probability models are simplified approximations to reality. They should be detailed enough to capture important characteristics of the random phenomenon so as to be useful as a prediction device, but not so detailed so as to produce an unwieldy model that is difficult to use in practice. The example of the number of telephone callers can be modeled by assigning a probability p to each person being on the phone anytime in the given 10-minute interval and *assuming* that whether one or more persons are on the phone does not affect the probability of others being on the phone. One can thus liken the event of being on the phone to a coin toss—if heads, a person is on the phone and if tails, a person is not on the phone. If there are $N = 4$ persons in the office, then the experimental outcome is likened to 4 coin tosses (either in succession or simultaneously—it makes no difference in the modeling). We can then ask for the probability that 3 persons are on the phone by determining the probability of 3 heads out of 4 coin tosses. The solution to this problem will be discussed in Chapter 3, where it is shown that the probability of k heads out of N coin tosses is given by

$$P[k] = \binom{N}{k} p^k (1-p)^{N-k} \tag{1.1}$$

where

$$\binom{N}{k} = \frac{N!}{(N-k)!k!}$$

for $k = 0, 1, \ldots, N$, and where $M! = 1 \cdot 2 \cdot 3 \cdots M$ for M a positive integer and by definition $0! = 1$. For our example, if $p = 0.75$ (we have a group of telemarketers) and $N = 4$ a compilation of the probabilities is shown in Figure 1.4. It is seen that the probability that three persons are on the phone is 0.42. Generally, the coin toss

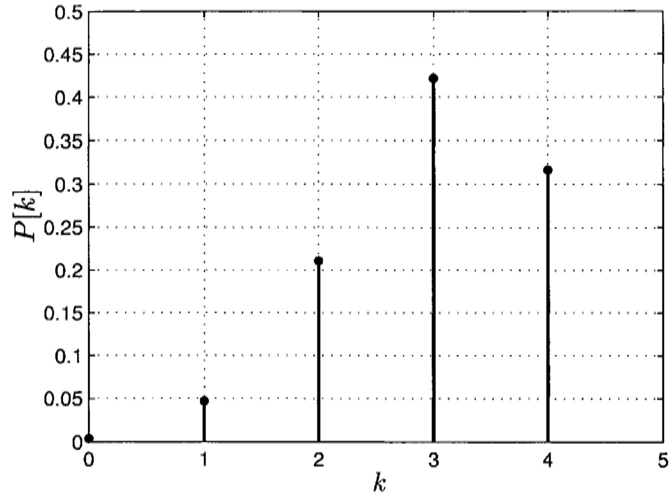


Figure 1.4: Probabilities for $N = 4$ coin tossings with $p = 0.75$.

model is a reasonable one for this type of situation. It will be poor, however, if the *assumptions are invalid.* Some practical objections to the model might be:

1. Different persons have different probabilities p (an eager telemarketer versus a not so eager one).

2. The probability of one person being on the phone is affected by whether his neighbor is on the phone (the two neighbors tend to talk about their planned weekends), i.e., the events are not "independent".

3. The probability p changes over time (later in the day there is less phone activity due to fatigue).

To accommodate these objections the model can be made more complex. In the end, however, the "more accurate" model may become a poorer predictor if the additional information used is not correct. It is generally accepted that a model should exhibit the property of "parsimony"—in other words, it should be as simple as possible.

The previous example had discrete outcomes. For continuous outcomes a frequently used probabilistic model is the *Gaussian* or "bell"-shaped curve. For the modeling of the length of time T a caller is on the phone it is not appropriate to ask for the probability that T will be *exactly*, for example, 5 minutes. This is because this probability will be zero (see Problem 1.6). Instead, we inquire as to the probability that T will be between 5 and 6 minutes. This question is answered by determining the area under the Gaussian curve shown in Figure 1.5. The form of the curve is given by

$$p_T(t) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(t-7)^2\right] \qquad -\infty < t < \infty \tag{1.2}$$

and although defined for all t, it is physically meaningful only for $0 \le t \le T_{\max}$,

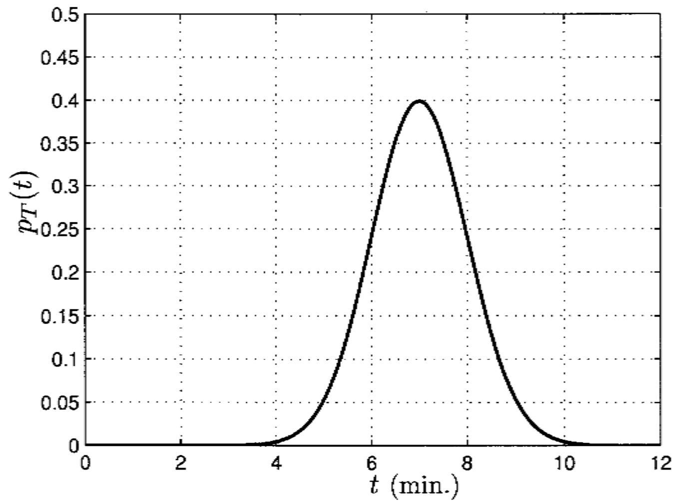


Figure 1.5: Gaussian or "bell"-shaped curve.

where $T_{\max} = 10$ for the current example. Since the area under the curve for times less than zero or greater than $T_{\max} = 10$ is nearly zero, this model is a reasonable approximation to physical reality. The curve has been chosen to be centered about $t = 7$ to relect an "average" time on the phone of 7 minutes for a given caller. Also, note that we let t denote the actual *value* of the *random* time T. Now, to determine the probability that the caller will be on the phone for between 5 and 6 minutes we integrate $p_T(t)$ over this interval to yield

$$P[5 \le T \le 6] = \int_5^6 p_T(t)dt = 0.1359. \tag{1.3}$$

The value of the integral must be numerically determined. Knowing the function $p_T(t)$ allows us to determine the probability for any interval. (It is called the probability density function (PDF) and is the probability per unit length. The PDF will be discussed in Chapter 10.) Also, it is apparent from Figure 1.5 that phone usage of duration less than 4 minutes or greater than 10 minutes is highly unlikely. Phone usage in the range of 7 minutes, on the other hand, is most probable. As before, some objections might be raised as to the accuracy of this model. A particularly lazy worker could be on the phone for only 3 minutes, as an example.

In this book we will henceforth assume that the models, which are mathematical in nature, are perfect and thus can be used to determine probabilities. In practice, the user must ultimately choose a model that is a reasonable one for the application of interest.

1.4 Analysis versus Computer Simulation

In the previous section we saw how to compute probabilities once we were given certain probability functions such as (1.1) for the discrete case and (1.2) for the continuous case. For many practical problems it is not possible to determine these functions. However, if we have a model for the random phenomenon, then we may carry out the experiment a large number of times to obtain an approximate probability. For example, to determine the probability of 3 heads in 4 tosses of a coin with probability of heads being $p = 0.75$, we toss the coin four times and count the number of heads, say $x_1 = 2$. Then, we repeat the experiment by tossing the coin four more times, yielding $x_2 = 1$ head. Continuing in this manner we execute a succession of 1000 experiments to produce the sequence of number of heads as $\{x_1, x_2, \ldots, x_{1000}\}$. Then, to determine the probability of 3 heads we use a *relative frequency* interpretation of probability to yield

$$P[3 \text{ heads}] = \frac{\text{Number of times 3 heads observed}}{1000}. \tag{1.4}$$

Indeed, early on probabilists did exactly this, although it was extremely tedious. *It is therefore of utmost importance to be able to simulate this procedure.* With the advent of the modern digital computer this is now possible. A digital computer has no problem performing a calculation once, 100 times, or 1,000,000 times. What is needed to implement this approach is a means to simulate the toss of a coin. Fortunately, this is quite easy as most scientific software packages have built-in *random number generators*. In MATLAB, for example, a number in the interval $(0,1)$ can be produced with the simple statement `x=rand(1,1)`. The number is chosen "at random" so that it is equally likely to be anywhere in the $(0,1)$ interval. As a result, a number in the interval $(0,1/2]$ will be observed with probability $1/2$ and a number in the remaining part of the interval $(1/2,1)$ also with probability $1/2$. Likewise, a number in the interval $(0,0.75]$ will be observed with probability $p = 0.75$. A computer simulation of the number of persons in the office on the telephone can thus be implemented with the MATLAB code (see Appendix 2A for a brief introduction to MATLAB):

```
number=0;
for i=1:4 % set up simulation for 4 coin tosses
  if rand(1,1)<0.75 % toss coin with p=0.75
    x(i,1)=1; % head
  else
```

```
    x(i,1)=0; % tail
  end
number=number+x(i,1); % count number of heads
end
```

Repeating this code segment 1000 times will result in a simulation of the previous experiment.

Similarly, for a continuous outcome experiment we require a means to generate a continuum of outcomes on a digital computer. Of course, strictly speaking this is not possible since digital computers can only provide a finite set of numbers, which is determined by the number of bits in each word. But if the number of bits is large enough, then the approximation is adequate. For example, with 64 bits we could represent 2^{64} numbers between 0 and 1, so that neighboring numbers would be $2^{-64} = 5 \times 10^{-20}$ apart. With this ability MATLAB can produce numbers that follow a Gaussian curve by invoking the statement `x=randn(1,1)`.

Throughout the text we will use MATLAB for examples and also exercises. However, any modern scientific software package can be used.

1.5 Some Notes to the Reader

The notation used in this text is summarized in Appendix A. Note that boldface type is reserved for vectors and matrices while regular face type will denote scalar quantities. All other symbolism is defined within the context of the discussion. Also, the reader will frequently be warned of potential "pitfalls". Common misconceptions leading to student errors will be described and noted. The pitfall or caution symbol shown below should be heeded.

The problems are of four types: computational or formula applications, word problems, computer exercises, and theoretical exercises. Computational or formula (denoted by **f**) problems are straightforward applications of the various formulas of the chapter, while word problems (denoted by **w**) require a more complete assimilation of the material to solve the problem. Computer exercises (denoted by **c**) will require the student to either use a computer to solve a problem or to simulate the analytical results. This will enhance understanding and can be based on MATLAB, although equivalent software may be used. Finally, theoretical exercises (denoted by **t**) will serve to test the student's analytical skills as well as to provide extensions to the material of the chapter. They are more challenging. Answers to selected problems are given in Appendix E. Those problems for which the answers are provided are noted in the problem section with the symbol (☺).

The version of MATLAB used in this book is 5.2, although newer versions should provide identical results. Many MATLAB outputs that are used for the text figures and for the problem solutions rely on random number generation. To match your results against those shown in the figures and the problem solutions, the same set of random numbers can be generated by using the MATLAB statements `rand('state',0)` and `randn('state',0)` at the beginning of each program. These statements will initialize the random number generators to produce the same set of random numbers. Finally, the MATLAB programs and code segments given in the book are indicated by the "typewriter" font, for example, `x=randn(1,1)`.

There are a number of other textbooks that the reader may wish to consult. They are listed in the following reference list, along with some comments on their contents.

Davenport, W.B., *Probability and Random Processes*, McGraw-Hill, New York, 1970. (Excellent introductory text.)

Feller, W., *An Introduction to Probability Theory and its Applications*, Vols. 1, 2, John Wiley, New York, 1950. (Definitive work on probability—requires mature mathematical knowledge.)

Hoel, P.G., S.C. Port, C.J. Stone, *Introduction to Probability Theory*, Houghton Mifflin Co., Boston, 1971. (Excellent introductory text but limited to probability.)

Leon-Garcia, A., *Probability and Random Processes for Electrical Engineering*, Addison-Wesley, Reading, MA, 1994. (Excellent introductory text.)

Parzen, E., *Modern Probability Theory and Its Applications*, John Wiley, New York, 1960. (Classic text in probability—useful for all disciplines).

Parzen, E., *Stochastic Processes*, Holden-Day, San Francisco, 1962. (Most useful for Markov process descriptions.)

Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965. (Classic but somewhat difficult text. Best used as a reference.)

Ross, S., *A First Course in Probability*, Prentice-Hall, Upper Saddle River, NJ, 2002. (Excellent introductory text covering only probability.)

Stark, H., J.W. Woods, *Probability and Random Processes with Applications to Signal Processing*, Third Ed., Prentice Hall, Upper Saddle River, NJ, 2002. (Excellent introductory text but at a somewhat more advanced level.)

References

Burdic, W.S., *Underwater Acoustic Systems Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

Ellenberg, S.S., D.M. Finklestein, D.A. Schoenfeld, "Statistical Issues Arising in AIDS Clinical Trials," *J. Am. Statist. Asssoc.*, Vol. 87, pp. 562–569, 1992.

Gustafson, D.E., A.S. Willsky, J.Y. Wang, M.C. Lancaster, J.H. Triebwasser, "ECG/VCG Rhythm Diagnosis Using Statistical Signal Analysis, Part II: Identification of Transient Rhythms," *IEEE Trans. Biomed. Eng.*, Vol. BME-25, pp. 353–361, 1978.

Ives, R.B., "The Applications of Advanced Pattern Recognition Techniques for the Discrimination Between Earthquakes and Nuclear Detonations," *Pattern Recognition*, Vol. 14, pp. 155–161, 1981.

Justice, J.H., "Array Processing in Exploration Seismology," in *Array Signal Processing*, S. Haykin, Ed., Prentice-Hall, Englewood Cliffs, NJ, 1985.

Knight, W.S., R.G. Pridham, S.M. Kay, "Digital Signal Processing for Sonar," *Proc. IEEE*, Vol. 69, pp. 1451–1506, Nov. 1981.

NOAA/NCDC, Lawrimore, J., "Climate at a Glance," National Oceanic and Atmospheric Administration, http://www.ncdc.noaa.gov/oa/climate/resarch/cag3/NA.html, 2003.

Proakis, J., *Digital Communications*, Second Ed., McGraw-Hill, New York, 1989.

Skolnik, M.I., *Introduction to Radar Systems*, McGraw-Hill, New York, 1980.

Taylor, S., *Modelling Financial Time Series*, John Wiley, New York, 1986.

Problems

1.1 (⌣̈) **(w)** A fair coin is tossed. Identify the random experiment, the set of outcomes, and the probabilities of each possible outcome.

1.2 (w) A card is chosen at random from a deck of 52 cards. Identify the random experiment, the set of outcomes, and the probabilities of each possible outcome.

1.3 (w) A fair die is tossed and the number of dots on the face noted. Identify the random experiment, the set of outcomes, and the probabilities of each possible outcome.

1.4 (w) It is desired to predict the annual summer rainfall in Rhode Island for 2010. If we use 9.76 inches as our prediction, how much in error might we be, based on the past data shown in Figure 1.1? Repeat the problem for Arizona by using 4.40 inches as the prediction.

1.5 (⌣̈) **(w)** Determine whether the following experiments have discrete or continuous outcomes:

a. Throw a dart with a point tip at a dartboard.

b. Toss a die.

c. Choose a lottery number.

d. Observe the outdoor temperature using an analog thermometer.

e. Determine the current time in hours, minutes, seconds, and AM or PM.

1.6 (w) An experiment has $N = 10$ outcomes that are equally probable. What is the probability of each outcome? Now let $N = 1000$ and also $N = 1,000,000$ and repeat. What happens as $N \to \infty$?

1.7 (⌣̈) **(f)** Consider an experiment with possible outcomes $\{1, 2, 3, \ldots\}$. If we assign probabilities

$$P[k] = \frac{1}{2^k} \qquad k = 1, 2, 3, \ldots$$

to the outcomes, will these probabilties sum to one? Can you have an infinite number of outcomes but still assign nonzero probabilities to each outcome? Reconcile these results with that of Problem 1.6.

1.8 (w) An experiment consists of tossing a fair coin four times in succession. What are the possible outcomes? Now count up the number of outcomes with three heads. If the outcomes are equally probable, what is the probability of three heads? Compare your results to that obtained using (1.1).

1.9 (w) Perform the following experiment by *actually tossing* a coin of your choice. Flip the coin four times and observe the number of heads. Then, repeat this experiment 10 times. Using (1.1) determine the probability for $k = 0, 1, 2, 3, 4$ heads. Next use (1.1) to determine the number of heads that is most probable for a single experiment? In your 10 experiments which number of heads appeared most often?

1.10 (⌣̈) **(w)** A coin is tossed 12 times. The sequence observed is the 12-tuple $(H, H, T, H, H, T, H, H, H, H, T, H)$. Is this a fair coin? Hint: Determine $P[k = 9]$ using (1.1) assuming a probability of heads of $p = 1/2$.

1.11 (t) Prove that $\sum_{k=0}^{N} P[k] = 1$, where $P[k]$ is given by (1.1). Hint: First prove the binomial theorem

$$(a + b)^N = \sum_{k=0}^{N} \binom{N}{k} a^k b^{N-k}$$

by induction (see Appendix B). Use Pascal's "triangle" rule

$$\binom{M}{k} = \binom{M-1}{k} + \binom{M-1}{k-1}$$

where

$$\binom{M}{k} = 0 \qquad k < 0 \text{ and } k > M.$$

1.12 (t) If $\int_a^b p_T(t)dt$ is the probability of observing T in the interval $[a, b]$, what is $\int_{-\infty}^{\infty} p_T(t)dt$?

1.13 (⌣) (f) Using (1.2) what is the probability of $T > 7$? Hint: Observe that $p_T(t)$ is symmetric about $t = 7$.

1.14 (⌣) (c) Evaluate the integral

$$\int_{-3}^{3} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}t^2\right] dt$$

by using the approximation

$$\sum_{n=-L}^{L} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(n\Delta)^2\right] \Delta$$

where L is the integer closest to $3/\Delta$ (the rounded value), for $\Delta = 0.1$, $\Delta = 0.01$, $\Delta = 0.001$.

1.15 (c) Simulate a fair coin tossing experiment by modifying the code given in Section 1.4. Using 1000 repetitions of the experiment, count the number of times three heads occur. What is the simulated probability of obtaining three heads in four coin tosses? Compare your result to that obtained using (1.1).

1.16 (c) Repeat Problem 1.15 but instead consider a biased coin with $p = 0.75$. Compare your result to Figure 1.4.

Chapter 2

Computer Simulation

2.1 Introduction

Computer simulation of random phenomena has become an indispensable tool in modern scientific investigations. So-called *Monte Carlo* computer approaches are now commonly used to promote understanding of probabilistic problems. In this chapter we continue our discussion of computer simulation, first introduced in Chapter 1, and set the stage for its use in later chapters. Along the way we will examine some well known properties of random events in the process of simulating their behavior. A more formal mathematical description will be introduced later but careful attention to the details now, will lead to a better intuitive understanding of the mathematical definitions and theorems to follow.

2.2 Summary

This chapter is an introduction to computer simulation of random experiments. In Section 2.3 there are examples to show how we can use computer simulation to provide counterexamples, build intuition, and lend evidence to a conjecture. However, it cannot be used to prove theorems. In Section 2.4 a simple MATLAB program is given to simulate the outcomes of a discrete random variable. Section 2.5 gives many examples of typical computer simulations used in probability, including probability density function estimation, probability of an interval, average value of a random variable, probability density function for a transformed random variable, and scatter diagrams for multiple random variables. Section 2.6 contains an application of probability to the "real-world" example of a digital communication system. A brief description of the MATLAB programming language is given in Appendix 2A.

2.3 Why Use Computer Simulation?

A computer simulation is valuable in many respects. It can be used

a. to provide counterexamples to proposed theorems

b. to build intuition by experimenting with random numbers

c. to lend evidence to a conjecture.

We now explore these uses by posing the following question: What is the effect of adding together the numerical outcomes of two or more experiments, i.e., what are the probabilities of the summed outcomes? Specifically, if U_1 represents the outcome of an experiment in which a number from 0 to 1 is chosen at random and U_2 is the outcome of an experiment in which another number is also chosen at random from 0 to 1, what are the probabilities of $X = U_1 + U_2$? The mathematical answer to this question is given in Chapter 12 (see Example 12.8), although at this point it is unknown to us. Let's say that someone asserts that there is a theorem that X is *equally likely* to be anywhere in the interval $[0, 2]$. To see if this is reasonable, we carry out a computer simulation by generating values of U_1 and U_2 and adding them together. Then we repeat this procedure M times. Next we plot a *histogram*, which gives the number of outcomes that fall in each subinterval within $[0, 2]$. As an example of a histogram consider the $M = 8$ possible outcomes for X of $\{1.7, 0.7, 1.2, 1.3, 1.8, 1.4, 0.6, 0.4\}$. Choosing the four subintervals (also called *bins*) $[0, 0.5]$, $(0.5, 1]$, $(1, 1.5]$, $(1.5, 2]$, the histogram appears in Figure 2.1. In this

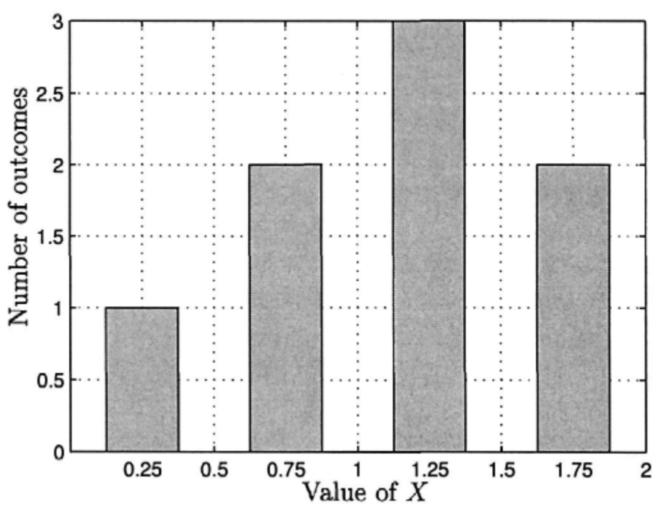


Figure 2.1: Example of a histogram for a set of 8 numbers in [0,2] interval.

example, 2 outcomes were between 0.5 and 1 and are therefore shown by the bar centered at 0.75. The other bars are similarly obtained. If we now increase the number of experiments to $M = 1000$, we obtain the histogram shown in Figure 2.2. Now it is clear that the values of X are *not equally likely*. Values near one appear

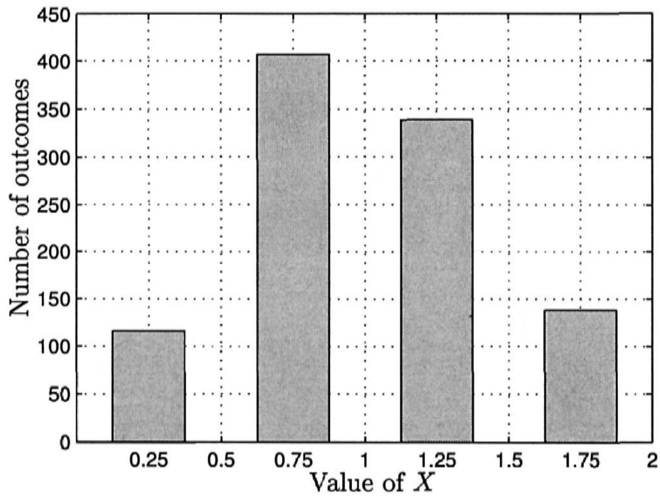


Figure 2.2: Histogram for sum of two equally likely numbers, both chosen in interval $[0, 1]$.

to be much more probable. Hence, we have generated a "counterexample" to the proposed theorem, or at least some evidence to the contrary.

We can build up our intuition by continuing with our experimentation. Attempting to justify the observed occurrences of X, we might suppose that the probabilities are higher near one because there are more ways to obtain these values. If we contrast the values of $X = 1$ versus $X = 2$, we note that $X = 2$ can only be obtained by choosing $U_1 = 1$ and $U_2 = 1$ but $X = 1$ can be obtained from $U_1 = U_2 = 1/2$ or $U_1 = 1/4, U_2 = 3/4$ or $U_1 = 3/4, U_2 = 1/4$, etc. We can lend credibility to this line of reasoning by supposing that U_1 and U_2 can only take on values in the set $\{0, 0.25, 0.5, 0.75, 1\}$ and finding all values of $U_1 + U_2$. In essence, we now look at a *simpler* problem in order to build up our intuition. An enumeration of the possible values is shown in Table 2.1 along with a "histogram" in Figure 2.3. It is clear now that the probability is highest at $X = 1$ because the number of combinations of U_1 and U_2 that will yield $X = 1$ is highest. Hence, we have learned about what happens when outcomes of experiments are added together by employing computer simulation.

We can now try to extend this result to the addition of three or more experimental outcomes via computer simulation. To do so define $X_3 = U_1 + U_2 + U_3$ and $X_4 = U_1 + U_2 + U_3 + U_4$ and repeat the simulation. A computer simulation with $M = 1000$ trials produces the histograms shown in Figure 2.4. It appears to

		U_2				
		0.00	0.25	0.50	0.75	1.00
	0.00	0.00	0.25	0.50	0.75	1.00
	0.25	0.25	0.50	0.75	1.00	1.25
U_1	0.50	0.50	0.75	1.00	1.25	1.50
	0.75	0.75	1.00	1.25	1.50	1.75
	1.00	1.00	1.25	1.50	1.75	2.00

Table 2.1: Possible values for $X = U_1 + U_2$ for intuition-building experiment.

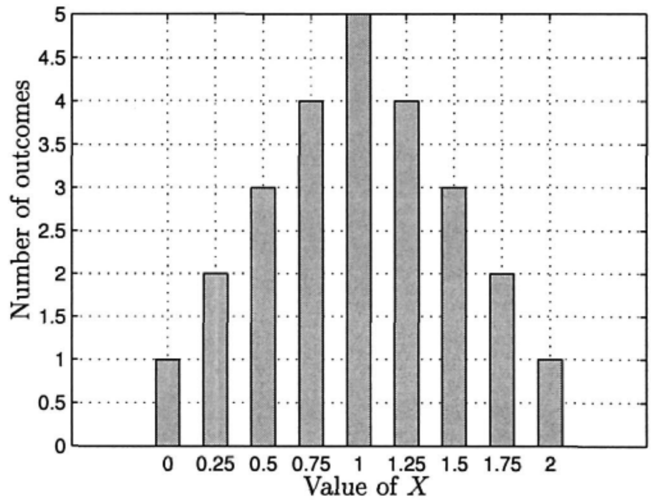
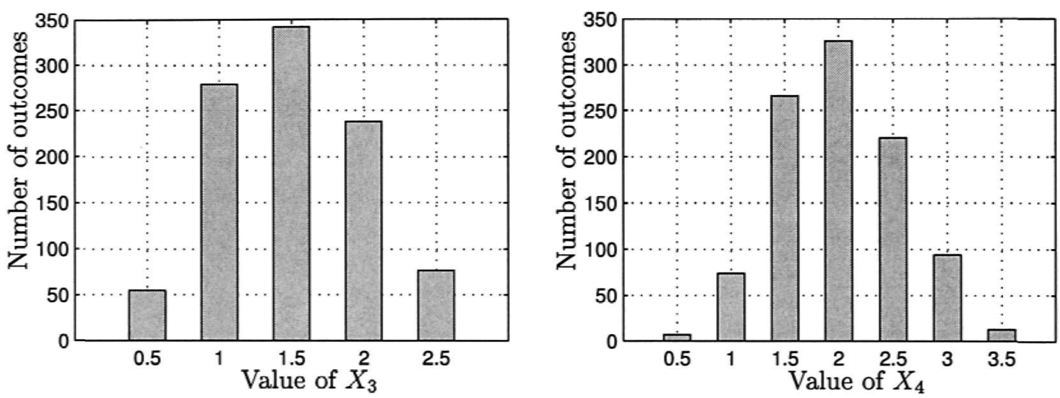


Figure 2.3: Histogram for X for intuition-building experiment.

bear out the conjecture that the most probable values are near the center of the $[0, 3]$ and $[0, 4]$ intervals, respectively. Additionally, the histograms appear more like a bell-shaped or Gaussian curve. Hence, we might now *conjecture*, based on these computer simulations, that as we add more and more experimental outcomes together, we will obtain a Gaussian-shaped histogram. This is in fact true, as will be proven later (see central limit theorem in Chapter 15). Note that we cannot *prove* this result using a computer simulation but only lend evidence to our theory. However, the use of computer simulations indicates *what* we need to prove, information that is invaluable in practice. In summary, computer simulation is a valuable tool for lending credibility to conjectures, building intuition, and uncovering new results.



(a) Sum of 3 U's

(b) Sum of 4 U's

Figure 2.4: Histograms for addition of outcomes.

Computer simulations cannot be used to prove theorems.

In Figure 2.2, which displayed the outcomes for 1000 trials, is it possible that the computer simulation could have produced 500 outcomes in [0,0.5], 500 outcomes in [1.5,2] and no outcomes in (0.5,1.5)? The answer is yes, although it is improbable. It can be shown that the probability of this occuring is

$$\binom{1000}{500}\left(\frac{1}{8}\right)^{1000} \approx 2.2 \times 10^{-604}$$

(see Problem 12.27).

2.4 Computer Simulation of Random Phenomena

In the previous chapter we briefly explained how to use a digital computer to simulate a random phenomenon. We now continue that discussion in more detail. Then, the following section applies the techniques to specific problems ecountered in probability. As before, we will distinguish between experiments that produce discrete outcomes from those that produce continuous outcomes.

We first define a *random variable* X as the *numerical outcome* of the random experiment. Typical examples are the number of dots on a die (discrete) or the distance of a dart from the center of a dartboard of radius one (continuous). The random variable X can take on the values in the set $\{1, 2, 3, 4, 5, 6\}$ for the first example and in the set $\{r : 0 \leq r \leq 1\}$ for the second example. We denote the random variable by a *capital letter*, say X, and its possible *values* by a small letter, say x_i for the discrete case and x for the continuous case. The distinction is analogous to that between a function *defined* as $g(x) = x^2$ and the *values* $y = g(x)$ that $g(x)$ can take on.

Now it is of interest to determine various properties of X. To do so we use a computer simulation, performing many experiments and observing the outcome for each experiment. The number of experiments, which is sometimes referred to as the number of *trials*, will be denoted by M. To simulate a discrete random variable we use `rand`, which generates a number at random within the $(0, 1)$ interval (see Appendix 2A for some MATLAB basics). Assume that in general the possible values of X are $\{x_1, x_2, \ldots, x_N\}$ with probabilities $\{p_1, p_2, \ldots, p_N\}$. As an example, if $N = 3$ we can generate M values of X by using the following code segment (which assumes `M,x1,x2,x3,p1,p2,p3` have been previously assigned):

```
for i=1:M
  u=rand(1,1);
  if u<=p1
    x(i,1)=x1;
  elseif u>p1 & u<=p1+p2
    x(i,1)=x2;
  elseif u>p1+p2
    x(i,1)=x3;
  end
end
```

After this code is executed, we will have generated M values of the random variable X. Note that the values of X so obtained are termed the *outcomes* or *realizations* of X. The extension to any number N of possible values is immediate. For a continuous random variable X that is Gaussian we can use the code segment:

```
for i=1:M
  x(i,1)=randn(1,1);
end
```

or equivalently `x=randn(M,1)`. Again at the conclusion of this code segment we will have generated M realizations of X. Later we will see how to generate realizations of random variables whose PDFs are not Gaussian (see Section 10.9).

2.5 Determining Characteristics of Random Variables

There are many ways to characterize a random variable. We have already alluded to the probability of the outcomes in the discrete case and the PDF in the continuous case. To be more precise consider a discrete random variable, such as that describing the outcome of a coin toss. If we toss a coin and let X be 1 if a head is observed and let X be 0 if a tail is observed, then the probabilities are defined to be p for $X = x_1 = 1$ and $1 - p$ for $X = x_2 = 0$. The probability p of $X = 1$ can be thought of as the relative frequency of the outcome of heads in a long succession of tosses. Hence, to determine the probability of heads we could toss a coin a large number of times and estimate p by the number of observed heads divided by the number of tosses. Using a computer to simulate this experiment, we might inquire as to the number of tosses that would be necessary to obtain an accurate estimate of the probability of heads. Unfortunately, this is not easily answered. A practical means, though, is to increase the number of tosses until the estimate so computed converges to a fixed number. A computer simulation is shown in Figure 2.5 where the estimate

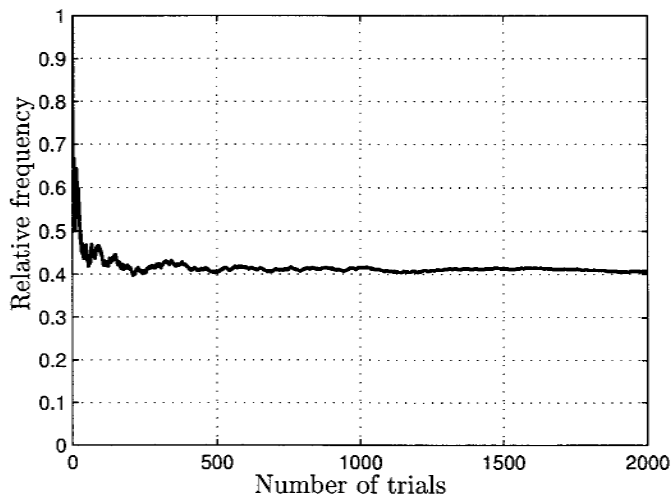


Figure 2.5: Estimate of probability of heads for various number of coin tosses.

appears to converge to about 0.4. Indeed, the true value (that value used in the simulation) was $p = 0.4$. It is also seen that the estimate of p is slightly higher than 0.4. This is due to the slight imperfections in the random number generator as well as computational errors. Increasing the number of trials will not improve the results. We next describe some typical simulations that will be useful to us. To illustrate the various simulations we will use a Gaussian random variable with realizations generated using `randn(1,1)`. Its PDF is shown in Figure 2.6.

Example 2.1 – Probability density function

A PDF may be estimated by first finding the histogram and then dividing the number of outcomes in each bin by M, the total number of realizations, to obtain the probability. Then to obtain the PDF $p_X(x)$ recall that the probability of X

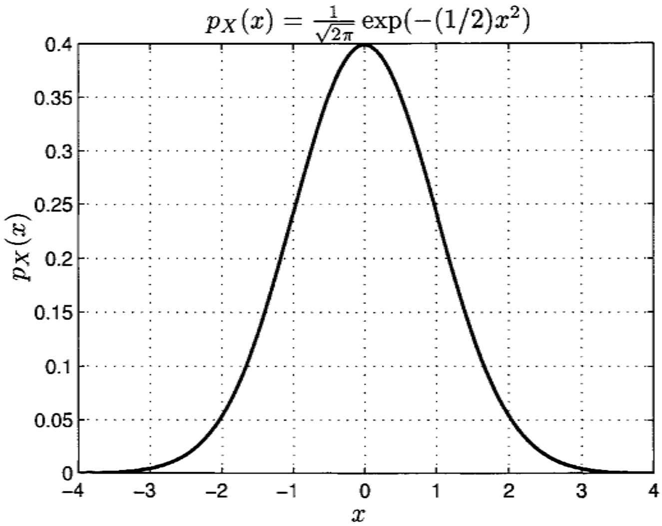


Figure 2.6: Gaussian probability density function.

taking on a value in an interval is found as the area under the PDF of that interval (see Section 1.3). Thus,

$$P[a \leq X \leq b] = \int_a^b p_X(x)dx \tag{2.1}$$

and if $a = x_0 - \Delta x/2$ and $b = x_0 + \Delta x/2$, where Δx is small, then (2.1) becomes

$$P[x_0 - \Delta x/2 \leq X \leq x_0 + \Delta x/2] \approx p_X(x_0)\Delta x$$

and therefore the PDF at $x = x_0$ is approximately

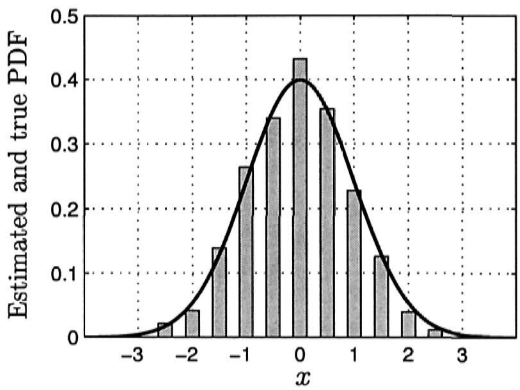
$$p_X(x_0) \approx \frac{P[x_0 - \Delta x/2 \leq X \leq x_0 + \Delta x/2]}{\Delta x}.$$

Hence, we need only divide the estimated probability by the bin width Δx. Also, note that as claimed in Chapter 1, $p_X(x)$ is seen to be the *probability per unit length.* In Figure 2.7 is shown the estimated PDF for a Gaussian random variable as well as the true PDF as given in Figure 2.6. The MATLAB code used to generate the figure is also shown.

◇

Example 2.2 – Probability of an interval

To determine $P[a \leq X \leq b]$ we need only generate M realizations of X, then count the number of outcomes that fall into the $[a, b]$ interval and divide by M. Of course




```
randn('state',0)
x=randn(1000,1);
bincenters=[-3.5:0.5:3.5]';
bins=length(bincenters);
h=zeros(bins,1);
for i=1:length(x)
  for k=1:bins
    if x(i)>bincenters(k)-0.5/2 ...
     & x(i)<=bincenters(k)+0.5/2
      h(k,1)=h(k,1)+1;
    end
  end
end
pxest=h/(1000*0.5);
xaxis=[-4:0.01:4]';
px=(1/sqrt(2*pi))*exp(-0.5*xaxis.^2);
```

Figure 2.7: Estimated and true probability density functions.

M should be large. In particular, if we let $a = 2$ and $b = \infty$, then we should obtain the value (which must be evaluated using numerical integration)

$$P[X > 2] = \int_2^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-(1/2)x^2\right) dx = 0.0228$$

and therefore very few realizations can be expected to fall in this interval. The results for an increasing number of realizations are shown in Figure 2.8. This illustrates the problem with the simulation of small probability events. It requires a large number of realizations to obtain accurate results. (See Problem 11.47 on how to reduce the number of realizations required.)

◇

Example 2.3 – Average value

It is frequently important to measure characteristics of X in addition to the PDF. For example, we might only be interested in the average or *mean* or *expected value* of X. If the random variable is Gaussian, then from Figure 2.6 we would expect X to be zero on the average. This conjecture is easily "verified" by using the *sample mean* estimate

$$\frac{1}{M}\sum_{i=1}^{M} x_i$$

M	Estimated $P[X > 2]$	True $P[X > 2]$
100	0.0100	0.0228
1000	0.0150	0.0228
10,000	0.0244	0.0288
100,000	0.0231	0.0288

```
randn('state',0)
M=100;count=0;
x=randn(M,1);
for i=1:M
  if x(i)>2
    count=count+1;
  end
end
probest=count/M
```

Figure 2.8: Estimated and true probabilities.

of the mean. The results are shown in Figure 2.9.

M	Estimated mean	True mean
100	0.0479	0
1000	−0.0431	0
10,000	0.0011	0
100,000	0.0032	0

```
randn('state',0)
M=100;
meanest=0;
x=randn(M,1);
for i=1:M
  meanest=meanest+(1/M)*x(i);
end
meanest
```

Figure 2.9: Estimated and true mean.

◊

Example 2.4 – A transformed random variable

One of the most important problems in probability is to determine the PDF for a transformed random variable, i.e., one that is a function of X, say X^2 as an example. This is easily accomplished by modifying the code in Figure 2.7 from `x=randn(1000,1)` to `x=randn(1000,1);x=x.^2;`. The results are shown in Figure 2.10. Note that the shape of the PDF is completely different than the original Gaussian shape (see Example 10.7 for the true PDF). Additionally, we can obtain the mean of X^2 by using

$$\frac{1}{M}\sum_{i=1}^{M} x_i^2$$

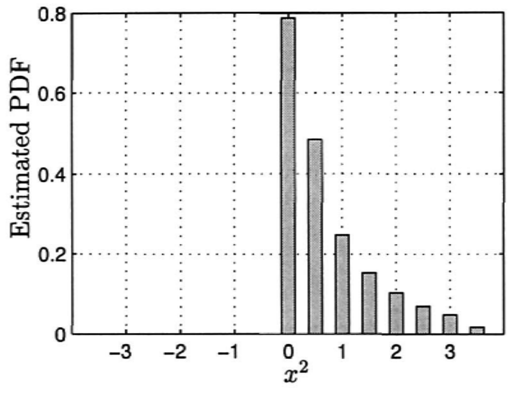


Figure 2.10: Estimated PDF of X^2 for X Gaussian.

as we did in Example 2.3. The results are shown in Figure 2.11.

M	Estimated mean	True mean
100	0.7491	1
1000	0.8911	1
10,000	1.0022	1
100,000	1.0073	1

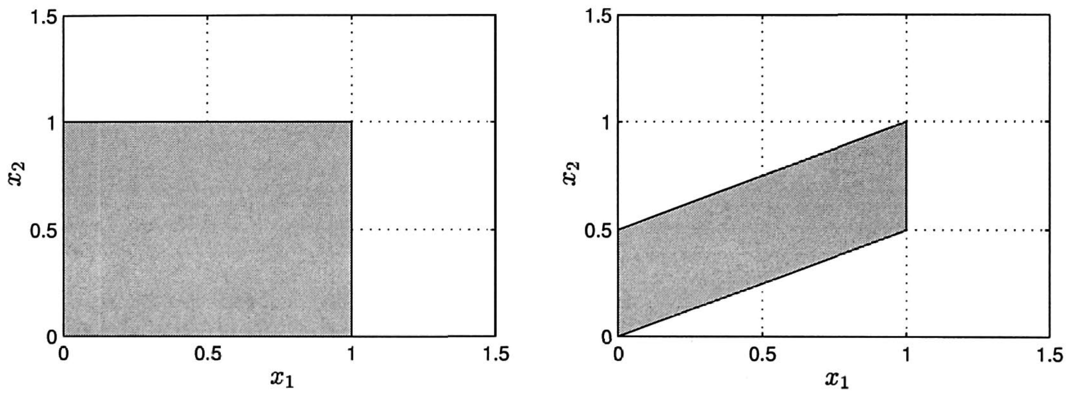
```
randn('state',0)
M=100;
meanest=0;
x=randn(M,1);
for i=1:M
  meanest=meanest+(1/M)*x(i)^2;
end
meanest
```

Figure 2.11: Estimated and true mean.

◇

Example 2.5 – Multiple random variables

Consider an experiment that yields two random variables or the *vector* random variable $[X_1\ X_2]^T$, where T denotes the transpose. An example might be the choice of a point in the square $\{(x, y) : 0 \le x \le 1, 0 \le y \le 1\}$ according to some procedure. This procedure may or may not cause the value of x_2 to depend on the value of x_1. For example, if the result of many repetitions of this experiment produced an even distribution of points indicated by the shaded region in Figure 2.12a, then we would say that there is no dependency between X_1 and X_2. On the other hand, if the points were evenly distributed within the shaded region shown in Figure 2.12b, then there is a strong dependency. This is because if, for example, $x_1 = 0.5$, then x_2 would have to lie in the interval $[0.25, 0.75]$. Consider next the random vector



(a) No dependency (b) Dependency

Figure 2.12: Relationships between random variables.

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}$$

where each U_i is generated using **rand**. The result of $M = 1000$ realizations is shown in Figure 2.13a. We say that the random variables X_1 and X_2 are *independent*. Of course, this is what we expect from a good random number generator. If instead, we defined the new random variables,

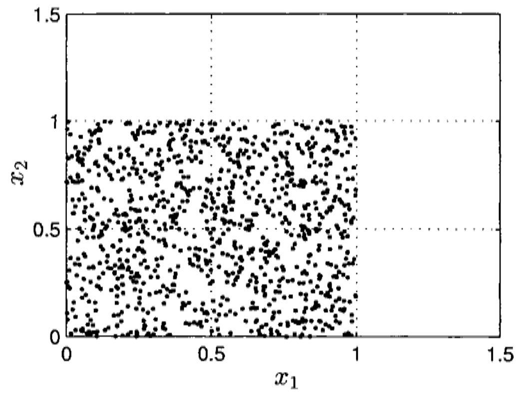
$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} U_1 \\ \frac{1}{2}U_1 + \frac{1}{2}U_2 \end{bmatrix}$$

then from the plot shown in Figure 2.13b, we would say that the random variables are dependent. Note that this type of plot is called a *scatter diagram*.

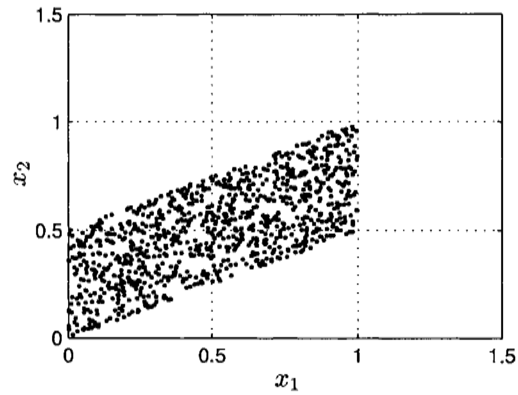
$\diamond$

2.6 Real-World Example – Digital Communications

In a phase-shift keyed (PSK) digital communication system a binary digit (also termed a *bit*), which is either a "0" or a "1", is communicated to a receiver by sending either $s_0(t) = A\cos(2\pi F_0 t + \pi)$ to represent a "0" or $s_1(t) = A\cos(2\pi F_0 t)$ to represent a "1", where $A > 0$ [Proakis 1989]. The receiver that is used to decode the transmission is shown in Figure 2.14. The input to the receiver is the noise



(a) No dependency



(b) Dependency

Figure 2.13: Relationships between random variables.

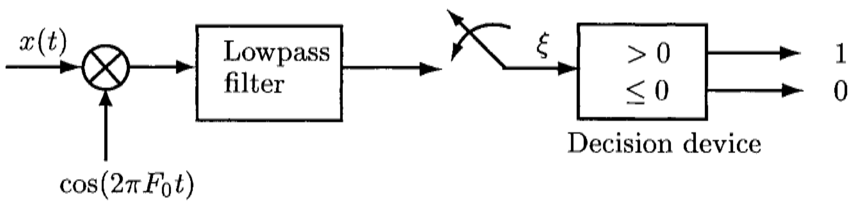


Figure 2.14: Receiver for a PSK digital communication system.

corrupted signal or $x(t) = s_i(t) + w(t)$, where $w(t)$ represents the channel noise. Ignoring the effect of noise for the moment, the output of the multiplier will be

$$\begin{aligned} s_0(t)\cos(2\pi F_0 t) &= A\cos(2\pi F_0 t + \pi)\cos(2\pi F_0 t) = -A\left(\frac{1}{2} + \frac{1}{2}\cos(4\pi F_0 t)\right) \\ s_1(t)\cos(2\pi F_0 t) &= A\cos(2\pi F_0 t)\cos(2\pi F_0 t) = A\left(\frac{1}{2} + \frac{1}{2}\cos(4\pi F_0 t)\right) \end{aligned}$$

for a 0 and 1 sent, respectively. After the lowpass filter, which filters out the $\cos(4\pi F_0 t)$ part of the signal, and sampler, we have

$$\xi = \begin{cases} -\frac{A}{2} & \text{for a } 0 \\ \frac{A}{2} & \text{for a } 1. \end{cases}$$

The receiver decides a 1 was transmitted if $\xi > 0$ and a 0 if $\xi \leq 0$. To model the channel noise we assume that the actual value of ξ observed is

$$\xi = \begin{cases} -\frac{A}{2} + W & \text{for a } 0 \\ \frac{A}{2} + W & \text{for a } 1 \end{cases}$$

where W is a Gaussian random variable. It is now of interest to determine how the error depends on the signal amplitude A. Consider the case of a 1 having been transmitted. Intuitively, if A is a large positive amplitude, then the chance that the noise will cause an error or equivalently, $\xi \leq 0$, should be small. This probability, termed the *probability of error* and denoted by P_e, is given by $P[A/2 + W \leq 0]$. Using a computer simulation we can plot P_e versus A with the result shown in Figure 2.15. Also, the true P_e is shown. (In Example 10.3 we will see how to analytically determine this probability.) As expected, the probability of error decreases as the

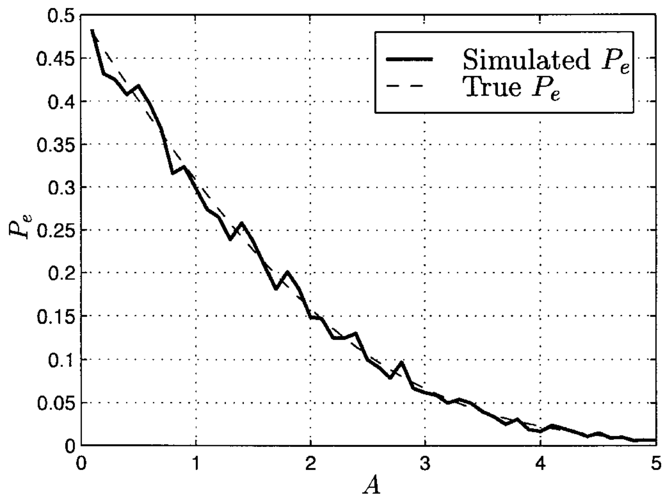


Figure 2.15: Probability of error for a PSK communication system.

signal amplitude increases. With this information we can design our system by choosing A to satisfy a given probability of error requirement. In actual systems this requirement is usually about $P_e = 10^{-7}$. Simulating this small probability would be exceedingly difficult due to the large number of trials required (but see also Problem 11.47). The MATLAB code used for the simulation is given in Figure 2.16.

References

Proakis, J., *Digitial Communications*, Second Ed., McGraw-Hill, New York, 1989.

Problems

Note: All the following problems require the use of a computer simulation. A realization of a *uniform* random variable is obtained by using `rand(1,1)` while a

```
A=[0.1:0.1:5]';
for k=1:length(A)
  error=0;
  for i=1:1000
    w=randn(1,1);
    if A(k)/2+w<=0
      error=error+1;
    end
  end
  Pe(k,1)=error/1000;
end
```

Figure 2.16: MATLAB code used to estimate the probability of error P_e in Figure 2.15.

realization of a *Gaussian* random variable is obtained by using **`randn(1,1)`**.

2.1 ($\ddot\smile$) **(c)** An experiment consists of tossing a fair coin twice. If a head occurs on the first toss, we let $x_1 = 1$ and if a tail occurs we let $x_1 = 0$. The same assignment is used for the outcome x_2 of the second toss. Defining the random variable as $Y = X_1X_2$, estimate the probabilities for the different possible values of Y. Explain your results.

2.2 (c) A pair of fair dice is tossed. Estimate the probability of "snake eyes" or a one for each die?

2.3 ($\ddot\smile$) **(c)** Estimate $P[-1 \leq X \leq 1]$ if X is a Gaussian random variable. Verify the results of your computer simulation by numerically evaluating the integral

$$\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx.$$

Hint: See Problem 1.14.

2.4 (c) Estimate the PDF of the random variable

$$X = \sum_{i=1}^{12} \left(U_i - \frac{1}{2}\right)$$

where U_i is a uniform random variable. Then, compare this PDF to the Gaussian PDF or

$$p_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right).$$

2.5 (c) Estimate the PDF of $X = U_1 - U_2$, where U_1 and U_2 are uniform random variables. What is the most probable range of values?

2.6 (☺) **(c)** Estimate the PDF of $X = U_1 U_2$, where U_1 and U_2 are uniform random variables. What is the most probable range of values?

2.7 (c) Generate realizations of a discrete random variable X, which takes on values 1, 2, and 3 with probabilities $p_1 = 0.1$, $p_2 = 0.2$ and $p_3 = 0.7$, respectively. Next based on the generated realizations estimate the probabilities of obtaining the various values of X.

2.8 (☺) **(c)** Estimate the mean of U, where U is a uniform random variable. What is the true value?

2.9 (c) Estimate the mean of $X + 1$, where X is a Gaussian random variable. What is the true value?

2.10 (c) Estimate the mean of X^2, where X is a Gaussian random variable.

2.11 (☺) **(c)** Estimate the mean of $2U$, where U is a uniform random variable. What is the true value?

2.12 (c) It is conjectured that if X_1 and X_2 are Gaussian random variables, then by subtracting them (let $Y = X_1 - X_2$), the probable range of values should be smaller. Is this true?

2.13 (☺) **(c)** A large circular dartboard is set up with a "bullseye" at the center of the circle, which is at the coordinate $(0, 0)$. A dart is thrown at the center but lands at (X, Y), where X and Y are two different Gaussian random variables. What is the average distance of the dart from the bullseye?

2.14 (☺) **(c)** It is conjectured that the mean of $\sqrt{U}$, where U is a uniform random variable, is $\sqrt{\text{mean of } U}$. Is this true?

2.15 (c) The Gaussian random variables X_1 and X_2 are linearly transformed to the new random variables

$$\begin{aligned} Y_1 &= X_1 + 0.1X_2 \\ Y_2 &= X_1 + 0.2X_2. \end{aligned}$$

Plot a scatter diagram for Y_1 and Y_2. Could you approximately determine the value of Y_2 if you knew that $Y_1 = 1$?

2.16 (c,w) Generate a scatter diagram for the linearly transformed random variables

$$\begin{aligned} X_1 &= U_1 \\ X_2 &= U_1 + U_2 \end{aligned}$$

where U_1 and U_2 are uniform random variables. Can you explain why the scatter diagram looks like a parallelogram? Hint: Define the vectors

$$\begin{aligned}
\mathbf{X} &= \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \\
\mathbf{e}_1 &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
\mathbf{e}_2 &= \begin{bmatrix} 0 \\ 1 \end{bmatrix}
\end{aligned}$$

and express $\mathbf{X}$ as a linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$.

Appendix 2A

Brief Introduction to MATLAB

A brief introduction to the scientific software package MATLAB is contained in this appendix. Further information is available at the Web site www.mathworks.com. MATLAB is a scientific computation and data presentation language.

Overview of MATLAB

The chief advantage of MATLAB is its use of high-level instructions for matrix algebra and built-in routines for data processing. In this appendix as well as throughout the text a MATLAB command is indicated with the typewriter font such as `end`. MATLAB treats matrices of any size (which includes vectors and scalars as special cases) as *elements* and hence matrix multiplication is as simple as `C=A*B`, where `A` and `B` are conformable matrices. In addition to the usual matrix operations of addition `C=A+B`, multiplication `C=A*B`, and scaling by a constant `c` as `B=c*A`, certain matrix operators are defined that allow convenient manipulation. For example, assume we first define the column vector $\mathbf{x} = [1\,2\,3\,4]^T$, where T denotes transpose, by using `x=[1:4]'`. The vector starts with the element 1 and ends with the element 4 and the colon indicates that the intervening elements are found by incrementing the start value by one, which is the default. For other increments, say 0.5, we use `x=[1:0.5:4]'`. To define the vector $\mathbf{y} = [1^2\,2^2\,3^2\,4^2]^T$, we can use the matrix *element by element* exponentiation operator `.^` to form `y=x.^2` if `x=[1:4]'`. Similarly, the operators `.*` and `./` perform element by element multiplication and division of the matrices, respectively. For example, if

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

Character	Meaning
+	addition (scalars, vectors, matrices)
-	subtraction (scalars, vectors, matrices)
*	multiplication (scalars, vectors, matrices)
/	division (scalars)
^	exponentiation (scalars, square matrices)
.*	element by element multiplication
./	element by element division
.^	element by element exponentiation
;	suppress printed output of operation
:	specify intervening values
'	conjugate transpose (transpose for real vectors, matrices)
...	line continuation (when command must be split)
%	remainder of line interpreted as comment
==	logical equals
|	logical or
&	logical and
~ =	logical not

Table 2A.1: Definition of common MATLAB characters.

then the statements `C=A.*B` and `D=A./B` produce the results

$$\begin{aligned}\mathbf{C} &= \begin{bmatrix} 1 & 4 \\ 9 & 16 \end{bmatrix} \\ \mathbf{D} &= \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\end{aligned}$$

respectively. A listing of some common characters is given in Table 2A.1. MATLAB has the usual built-in functions of `cos`, `sin`, etc. for the trigonometric functions, `sqrt` for a square root, `exp` for the exponential function, and `abs` for absolute value, as well as many others. When a function is applied to a matrix, the function is applied to each element of the matrix. Other built-in symbols and functions and their meanings are given in Table 2A.2.

Matrices and vectors are easily specified. For example, to define the column vector $\mathbf{c}_1 = [1\ 2]^T$, just use `c1=[1 2].'` or equivalently `c1=[1;2]`. To define the $\mathbf{C}$ matrix given previously, the construction `C=[1 4;9 16]` is used. Or we could first define $\mathbf{c}_2 = [4\ 16]^T$ by `c2=[4 16].'` and then use `C=[c1 c2]`. It is also possible to extract portions of matrices to yield smaller matrices or vectors. For example, to extract the first column from the matrix $\mathbf{C}$ use `c1=C(:,1)`. The colon indicates that all elements in the first column should be extracted. Many other convenient manipulations of matrices and vectors are possible.

Function	Meaning
`pi`	π
`i`	$\sqrt{-1}$
`j`	$\sqrt{-1}$
`round(x)`	rounds every element in $\mathbf{x}$ to the nearest integer
`floor(x)`	replaces every element in $\mathbf{x}$ by the nearest integer less than or equal to x
`inv(A)`	takes the inverse of the square matrix $\mathbf{A}$
`x=zeros(N,1)`	assigns an $N \times 1$ vector of all zeros to $\mathbf{x}$
`x=ones(N,1)`	assigns an $N \times 1$ vector of all ones to $\mathbf{x}$
`x=rand(N,1)`	generates an $N \times 1$ vector of all uniform random variables
`x=randn(N,1)`	generates an $N \times 1$ vector of all Gaussian random variables
`rand('state',0)`	initializes uniform random number generator
`randn('state',0)`	initializes Gaussian random number generator
`M=length(x)`	sets M equal to N if $\mathbf{x}$ is $N \times 1$
`sum(x)`	sums all elements in vector $\mathbf{x}$
`mean(x)`	computes the sample mean of the elements in $\mathbf{x}$
`flipud(x)`	flips the vector $\mathbf{x}$ upside down
`abs`	takes the absolute value (or complex magnitude) of every element of $\mathbf{x}$
`fft(x,N)`	computes the FFT of length N of $\mathbf{x}$ (zero pads if `N>length(x)`)
`ifft(x,N)`	computes the inverse FFT of length N of $\mathbf{x}$
`fftshift(x)`	interchanges the two halves of an FFT output
`pause`	pauses the execution of a program
`break`	terminates a loop when encountered
`whos`	lists all variables and their attributes in current workspace
`help`	provides help on commands, e.g., `help sqrt`

Table 2A.2: Definition of useful MATLAB symbols and functions.

Any vector that is generated whose dimensions are not explicitly specified is assumed to be a *row* vector. For example, if we say `x=ones(10)`, then it will be designated as the 1×10 row vector consisting of all ones. To yield a column vector use `x=ones(10,1)`.

Loops are implemented with the construction

```
for k=1:10
  x(k,1)=1;
end
```

which is equivalent to `x=ones(10,1)`. Logical flow can be accomplished with the construction

```
if x>0
  y=sqrt(x);
else
  y=0;
end
```

Finally, a good practice is to begin each program or script, which is called an "m" file (due to its syntax, for example, `pdf.m`), with a `clear all` command. This will clear all variables in the workspace, since otherwise the current program may inadvertently (on the part of the programmer) use previously stored variable data.

Plotting in MATLAB

Plotting in MATLAB is illustrated in the next section by example. Some useful functions are summarized in Table 2A.3.

Function	Meaning
`figure`	opens up a new figure window
`plot(x,y)`	plots the elements of **x** versus the elements of **y**
`plot(x1,y1,x2,y2)`	same as above except multiple plots are made
`plot(x,y,'.')`	same as `plot` except the points are not connected
`title('my plot')`	puts a title on the plot
`xlabel('x')`	labels the x axis
`ylabel('y')`	labels the y axis
`grid`	draws grid on the plot
`axis([0 1 2 4])`	plots only the points in range $0 \leq x \leq 1$ and $2 \leq y \leq 4$
`text(1,1,'curve 1')`	places the text "curve 1" at the point (1,1)
`hold on`	holds current plot
`hold off`	releases current plot

Table 2A.3: Definition of useful MATLAB plotting functions.

An Example Program

A complete MATLAB program is given below to illustrate how one might compute the samples of several sinusoids of different amplitudes. It also allows the sinusoids to be clipped. The sinusoid is $s(t) = A\cos(2\pi F_0 t + \pi/3)$, with $A = 1$, $A = 2$, and $A = 4$, $F_0 = 1$, and $t = 0, 0.01, 0.02, \ldots, 10$. The clipping level is set at ± 3, i.e., any sample above $+3$ is clipped to $+3$ and any sample less than -3 is clipped to -3.

```
%  matlabexample.m
%
%  This program computes and plots samples of a sinusoid
%  with amplitudes 1, 2, and 4.  If desired, the sinusoid can be
%  clipped to simulate the effect of a limiting device.
%  The frequency is 1 Hz and the time duration is 10 seconds.
%  The sample interval is 0.1 seconds. The code is not efficient but
%  is meant to illustrate MATLAB statements.
%
clear all % clear all variables from workspace
delt=0.01; % set sampling time interval
F0=1; % set frequency
t=[0:delt:10]'; % compute time samples 0,0.01,0.02,...,10
A=[1 2 4]'; % set amplitudes
clip='yes'; % set option to clip
for i=1:length(A) % begin computation of sinusoid samples
   s(:,i)=A(i)*cos(2*pi*F0*t+pi/3); % note that samples for sinusoid
                                    % are computed all at once and
                                    % stored as columns in a matrix
   if clip=='yes' % determine if clipping desired
      for k=1:length(s(:,i)) % note that number of samples given as
                             % dimension of column using length command
         if s(k,i)>3 % check to see if sinusoid sample exceeds 3
            s(k,i)=3; % if yes, then clip
         elseif s(k,i)<-3 % check to see if sinusoid sample is less
            s(k,i)=-3;    % than -3 if yes, then clip
         end
      end
   end
end
figure % open up a new figure window
plot(t,s(:,1),t,s(:,2),t,s(:,3)) % plot sinusoid samples versus time
                                 % samples for all three sinusoids
grid % add grid to plot
xlabel('time, t') % label x-axis
```

```
ylabel('s(t)') % label y-axis
axis([0 10 -4 4]) % set up axes using axis([xmin xmax ymin ymax])
legend('A=1','A=2','A=4') % display a legend to distinguish
                          % different sinusoids
```

The output of the program is shown in Figure 2A.1. Note that the different graphs will appear as different colors.

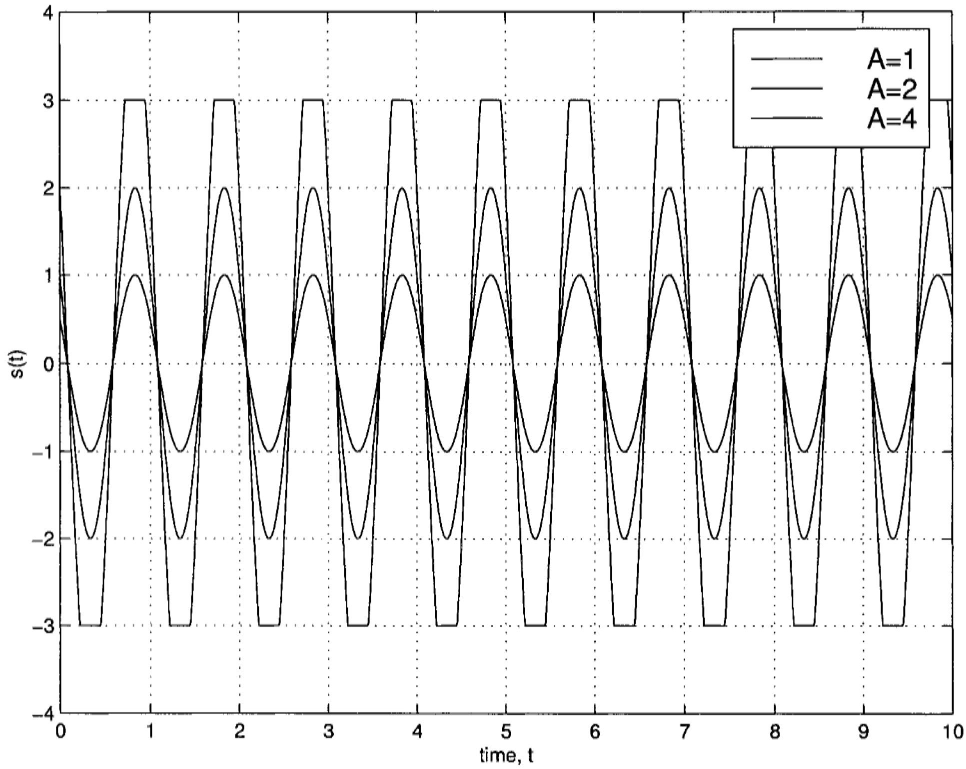


Figure 2A.1: Output of MATLAB program `matlabexample.m`.

Chapter 3

Basic Probability

3.1 Introduction

We now begin the formal study of probability. We do so by utilizing the properties of sets in conjunction with the *axiomatic* approach to probability. In particular, we will see how to solve a class of probability problems via *counting methods*. These are problems such as determining the probability of obtaining a royal flush in poker or of obtaining a defective item from a batch of mostly good items, as examples. Furthermore, the axiomatic approach will provide the basis for all our further studies of probability. Only the methods of determining the probabilities will have to be modified in accordance with the problem at hand.

3.2 Summary

Section 3.3 reviews set theory, with Figure 3.1 illustrating the standard definitions. Manipulation of sets can be facilitated using De Morgan's laws of (3.6) and (3.7). The application of set theory to probability is summarized in Table 3.1. Using the three axioms described in Section 3.4 a theory of probability can be formulated and a means for computing probabilities constructed. Properties of the probability function are given in Section 3.5. In addition, the probability for a union of three events is given by (3.20). An equally likely probability assignment for a continuous sample space is given by (3.22) and is shown to satisfy the basic axioms. Section 3.7 introduces the determination of probabilities for discrete sample spaces with equally likely outcomes. The basic formula is given by (3.24). To implement this approach for more complicated problems in which brute-force counting of outcomes is not possible, the subject of combinatorics is described in Section 3.8. Permutations and combinations are defined and applied to several examples for computing probabilities. Based on these counting methods the hypergeometric probability law of (3.27) and the binomial probability law of (3.28) are derived in Section 3.9. Finally, an example of the application of the binomial law to a quality control problem is given in Section 3.10.

3.3 Review of Set Theory

The reader has undoubtedly been introduced to set theory at some point in his/her education. We now summarize only the salient definitions and properties that are germane to probability. A set is defined as a collection of objects, for example, the set of students in a probability class. The set A can be defined either by the *enumeration* method, i.e., a listing of the students as

$$A = \{\text{Jane, Bill, Jessica, Fred}\} \tag{3.1}$$

or by the *description* method

$$A = \{\text{students: each student is enrolled in the probability class}\}$$

where the ":" is read as "such that". Another example would be the set of natural numbers or

$$\begin{aligned} B &= \{1, 2, 3, \ldots\} && \text{(enumeration)} \\ B &= \{I : I \text{ is an integer and } I \geq 1\} && \text{(description).} \end{aligned} \tag{3.2}$$

Each object in the set is called an *element* and each element is *distinct*. For example, the sets $\{1, 2, 3\}$ and $\{1, 2, 1, 3\}$ are equivalent. There is no reason to list an element in a set more than once. Likewise, the *ordering* of the elements within the set is not important. The sets $\{1, 2, 3\}$ and $\{2, 1, 3\}$ are equivalent. Sets are said to be *equal* if they contain the same elements. For example, if $C_1 = \{\text{Bill, Fred}\}$ and $C_2 = \{\text{male members in the probability class}\}$, then $C_1 = C_2$. Although the description may change, it is ultimately the contents of the set that is of importance. An element x of a set A is denoted using the symbolism $x \in A$, and is read as "x is contained in A", as for example, $1 \in B$ for the set B defined in (3.2). Some sets have no elements. If the instructor in the probability class does not give out any grades of "A", then the set of students receiving an "A" is $D = \{\,\}$. This is called the *empty set* or the *null set*. It is denoted by $\emptyset$ so that $D = \emptyset$. On the other hand, the instructor may be an easy grader and give out all "A"s. Then, we say that $D = \mathcal{S}$, where $\mathcal{S}$ is called the *universal set* or the set of *all* students enrolled in the probability class. These concepts, in addition to some others, are further illustrated in the next example.

Example 3.1 – Set concepts

Consider the set of *all outcomes* of a tossed die. This is

$$A = \{1, 2, 3, 4, 5, 6\}. \tag{3.3}$$

The numbers $1, 2, 3, 4, 5, 6$ are its elements, which are distinct. The set of integer numbers from 1 to 6 or $B = \{I : 1 \leq I \leq 6\}$ is equal to A. The set A is also the universal set $\mathcal{S}$ since it contains all the outcomes. This is in contrast to the set

$C = \{2, 4, 6\}$, which contains only the even outcomes. The set C is called a *subset* of A. A *simple set* is a set containing a *single* element, as for example, $C = \{1\}$.

$\diamond$

Element vs. simple set

In the example of the probability class consider the set of instructors. Usually, there is only one instructor and so the set of instructors can be defined as the simple set $A = \{\text{Professor Laplace}\}$. However, this is not the same as the "element" given by Professor Laplace. A distinction is therefore made between the instructors teaching probability and an individual instructor. As another example, it is clear that sometimes elements in a set can be added, as, for example, $2 + 3 = 5$, but it makes no sense to add sets as in $\{2\} + \{3\} = \{5\}$.

More formally, a set B is defined as a subset of a set A if every element in B is also an element of A. We write this as $B \subset A$. This also includes the case of $B = A$. In fact, we can say that $A = B$ if $A \subset B$ *and* $B \subset A$.

Besides subsets, new sets may be derived from other sets in a number of ways. If $\mathcal{S} = \{x : -\infty < x < \infty\}$ (called the set of *real numbers*), then $A = \{x : 0 < x \leq 2\}$ is clearly a subset of $\mathcal{S}$. The *complement* of A, denoted by A^c, is the set of elements in $\mathcal{S}$ but not in A. This is $A^c = \{x : x \leq 0 \text{ or } x > 2\}$. Two sets can be combined together to form a new set. For example, if

$$\begin{aligned} A &= \{x : 0 \leq x \leq 2\} \\ B &= \{x : 1 \leq x \leq 3\} \end{aligned} \tag{3.4}$$

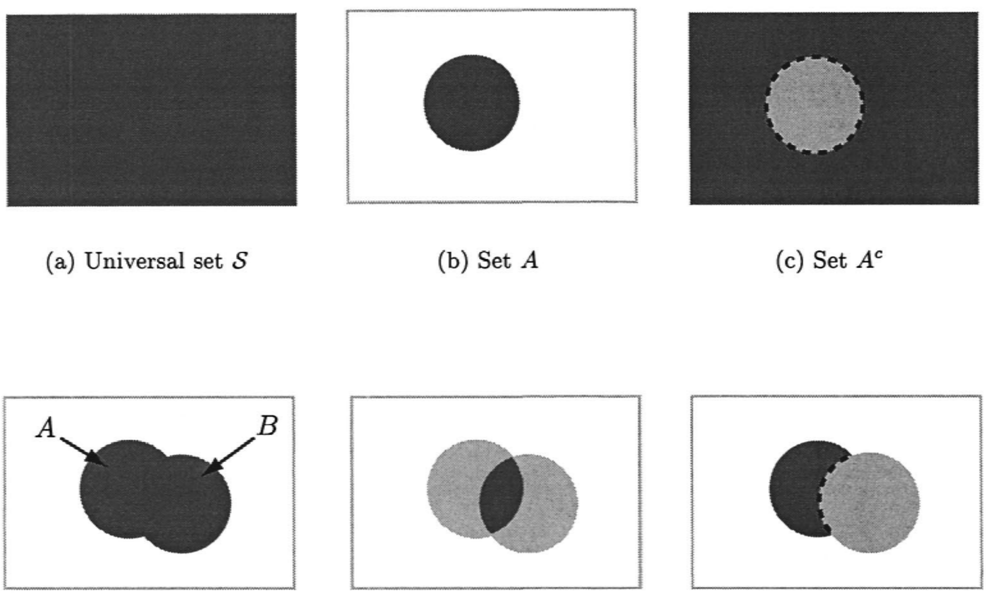
then the *union* of A and B, denoted by $A \cup B$, is the set of elements that belong to A *or* B or both A *and* B (so-called *inclusive or*). Hence, $A \cup B = \{x : 0 \leq x \leq 3\}$. This definition may be extended to multiple sets $A_1, A_2, \ldots, A_N$ so that the union is the set of elements for which each element belongs to at least one of these sets. It is denoted by

$$A_1 \cup A_2 \cup A_2 \cup \cdots \cup A_N = \bigcup_{i=1}^{N} A_i.$$

The *intersection* of sets A and B, denoted by $A \cap B$, is defined as the set of elements that belong to *both* A *and* B. Hence, $A \cap B = \{x : 1 \leq x \leq 2\}$ for the sets of (3.4). We will sometimes use the shortened symbolism AB to denote $A \cap B$. This definition may be extended to multiple sets $A_1, A_2, \ldots, A_N$ so that the intersection is the set of elements for which each element belongs to *all* of these sets. It is denoted by

$$A_1 \cap A_2 \cap A_2 \cap \cdots \cap A_N = \bigcap_{i=1}^{N} A_i.$$

The *difference* between sets, denoted by $A - B$, is the set of elements in A but *not* in B. Hence, for the sets of (3.4) $A - B = \{x : 0 \leq x < 1\}$. These concepts can be illustrated pictorially using a *Venn diagram* as shown in Figure 3.1. The darkly



(a) Universal set S (b) Set A (c) Set A^c

(d) Set $A \cup B$ (e) Set $A \cap B$ (f) Set $A - B$

Figure 3.1: Illustration of set definitions – darkly shaded region indicates the set.

shaded regions are the sets described. The dashed portions are *not* included in the sets. A Venn diagram is useful for visualizing set operations. As an example, one might inquire whether the sets $A - B$ and $A \cap B^c$ are equivalent or if

$$A - B = A \cap B^c. \tag{3.5}$$

From Figures 3.2 and 3.1f we see that they appear to be. However, to formally prove that this relationship is true requires one to let $C = A - B$, $D = A \cap B^c$ and prove that (a) $C \subset D$ and (b) $D \subset C$. To prove (a) assume that $x \in A - B$. Then, by definition of the difference set (see Figure 3.1f) $x \in A$ but x is *not* an element of B. Hence, $x \in A$ *and* x must also be an element of B^c. Since $D = A \cap B^c$, x must be an element of D. Hence, $x \in A \cap B^c$ and since this is true for every $x \in A - B$,

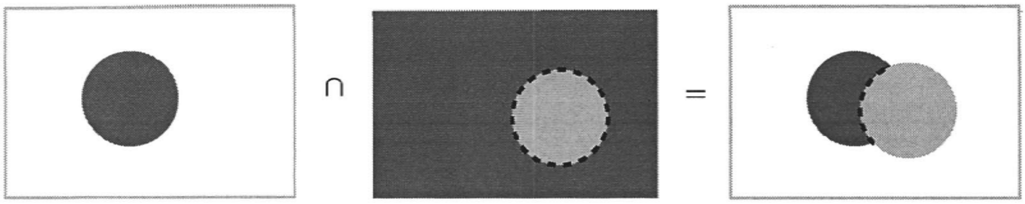


Figure 3.2: Using Venn diagrams to "validate" set relationships.

we have that $A - B \subset A \cap B^c$. The reader is asked to complete the proof of (b) in Problem 3.6.

With the foregoing set definitions a number of results follow. They will be useful in manipulating sets to allow easier calculation of probabilities. We now list these.

1. $(A^c)^c = A$

2. $A \cup A^c = \mathcal{S}$, $A \cap A^c = \emptyset$

3. $A \cup \emptyset = A$, $A \cap \emptyset = \emptyset$

4. $A \cup \mathcal{S} = \mathcal{S}$, $A \cap \mathcal{S} = A$

5. $\mathcal{S}^c = \emptyset$, $\emptyset^c = \mathcal{S}$.

If two sets A and B have no elements in common, they are said to be *disjoint*. The condition for being disjoint is therefore $A \cap B = \emptyset$. If, furthermore, the sets contain between them all the elements of $\mathcal{S}$, then the sets are said to *partition* the universe. This latter additional condition is that $A \cup B = \mathcal{S}$. An example of sets that partition the universe is given in Figure 3.3. Note also that the sets A and A^c

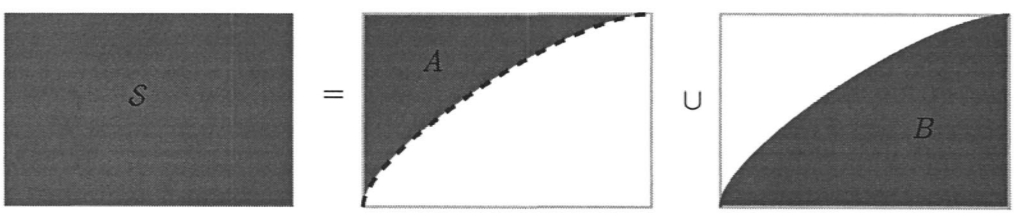


Figure 3.3: Sets that partition the universal set.

are always a partitioning of $\mathcal{S}$ (why?). More generally, *mutually* disjoint sets or sets $A_1, A_2, \ldots, A_N$ for which $A_i \cap A_j = \emptyset$ for all $i \neq j$ are said to partition the universe if $\mathcal{S} = \cup_{i=1}^N A_i$ (see also Problem 3.9 on how to construct these sets in general). For example, the set of students enrolled in the probability class, which is defined as the universe (although of course other universes may be defined such as the set of all students attending the given university), is partitioned by

$$\begin{aligned} A_1 &= \{\text{males}\} = \{\text{Bill, Fred}\} \\ A_2 &= \{\text{females}\} = \{\text{Jane, Jessica}\}. \end{aligned}$$

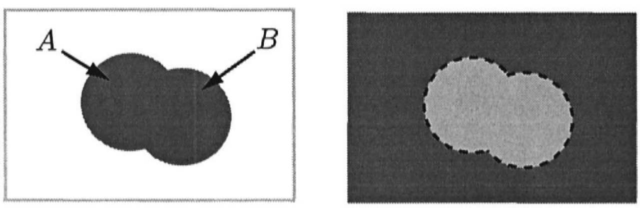
Algebraic rules for manipulating multiple sets, which will be useful, are

1. $A \cup B = B \cup A$
   $A \cap B = B \cap A$ commutative properties

2. $A \cup (B \cup C) = (A \cup B) \cup C$
   $A \cap (B \cap C) = (A \cap B) \cap C$ associative properties

3. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
   $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ distributive properties.

Another important relationship for manipulating sets is De Morgan's law. Referring



(a) Set $A \cup B$ (b) Set $A^c \cap B^c$

Figure 3.4: Illustration of De Morgan's law.

to Figure 3.4 it is obvious that

$$A \cup B = (A^c \cap B^c)^c \tag{3.6}$$

which allows one to convert from unions to intersections. To convert from intersections to unions we let $A = C^c$ and $B = D^c$ in (3.6) to obtain

$$C^c \cup D^c = (C \cap D)^c$$

and therefore

$$C \cap D = (C^c \cup D^c)^c. \tag{3.7}$$

In either case we can perform the conversion by the following set of rules:

1. Change the unions to intersections and the intersections to unions ($A \cup B \Rightarrow A \cap B$)

2. Complement each set ($A \cap B \Rightarrow A^c \cap B^c$)

3. Complement the overall expression ($A^c \cap B^c \Rightarrow (A^c \cap B^c)^c$).

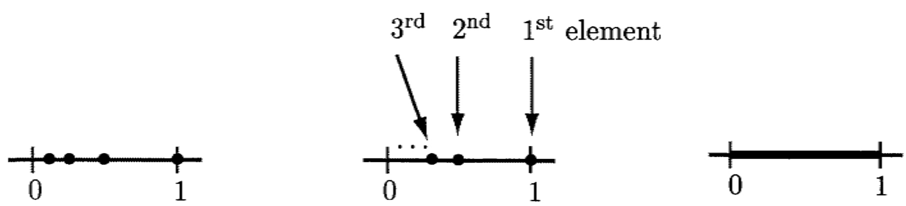
Finally, we discuss the *size* of a set. This will be of extreme importance in assigning probabilities. The set $\{2, 4, 6\}$ is a finite set, having a finite number of elements. The set $\{2, 4, 6, \ldots\}$ is an infinite set, having an infinite number of elements. In the latter case, although the set is infinite, it is said to be *countably infinite*. This means that "in theory" we can count the number of elements in the set. (We do so by pairing up each element in the set with an element in the set of natural numbers or $\{1, 2, 3, \ldots\}$). In either case, the set is said to be *discrete*. The set may be pictured as points on the real line. In contrast to these sets the set $\{x : 0 \le x \le 1\}$ is infinite and cannot be counted. This set is termed *continuous* and is pictured as a line segment on the real line. Another example follows.

Example 3.2 – Size of sets

The sets

$$\begin{aligned} A &= \left\{\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1\right\} && \text{finite set - discrete} \\ B &= \left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\right\} && \text{countably infinite set - discrete} \\ C &= \{x : 0 \le x \le 1\} && \text{infinite set - continuous} \end{aligned}$$

are pictured in Figure 3.5.



(a) Finite set, A (b) Countably infinite set, B (c) Infinite continuous set, C

Figure 3.5: Examples of sets of different sizes.

◊

3.4 Assigning and Determining Probabilities

In the previous section we reviewed various aspects of set theory. This is because the concept of sets and operations on sets provide an ideal description for a probabilistic model and the means for determining the probabilites associated with the model. Consider the tossing of a fair die. The possible outcomes comprise the elements of the set $\mathcal{S} = \{1,2,3,4,5,6\}$. Note that this set is composed of *all* the possible outcomes, and as such is the universal set. In probability theory $\mathcal{S}$ is termed the *sample space* and its elements s are the *outcomes* or *sample points.* At times we may be interested in a particular outcome of the die tossing experiment. Other times we might not be interested in a particular outcome, but whether or not the outcome was an even number, as an example. Hence, we would inquire as to whether the outcome was *included* in the set $E_{\text{even}} = \{2,4,6\}$. Clearly, E_{even} is a subset of $\mathcal{S}$ and is termed an *event.* The simplest type of events are the ones that contain only a single outcome such as $E_1 = \{1\}$, $E_2 = \{2\}$, or $E_6 = \{6\}$, as examples. These are called *simple events.* Other events are $\mathcal{S}$, the sample space itself, and $\emptyset = \{\}$, the set with no outcomes. These events are termed the *certain event* and the *impossible event,* respectively. This is because the outcome of the experiment must be an element of $\mathcal{S}$ so that $\mathcal{S}$ is certain to occur. Also, the event that does not contain any outcomes cannot occur so that this event is impossible. Note that we are saying that *an event occurs if the outcome is an element of the defining set of that event.* For example, the event that a tossed die produces an even number *occurs* if it comes up a 2 or a 4 or a 6. These numbers are just the elements of E_{even}. Disjoint sets such as $\{1,2\}$ and $\{3,4\}$ are said to be *mutually exclusive,* in that an outcome cannot be in both sets simultaneously and hence both events cannot occur. The events then are said to be mutually exclusive. It is seen that probabilistic questions can be formulated using set theory, albeit with its own terminology. A summary of the equivalent terms used is given in Table 3.1.

Set theory	Probability theory	Probability symbol
universe	sample space (certain event)	$\mathcal{S}$
element	outcome (sample point)	s
subset	event	E
disjoint sets	mutually exclusive events	$E_1 \cap E_2 = \emptyset$
null set	impossible event	$\emptyset$
simple set	simple event	$E = \{s\}$

Table 3.1: Terminology for set and probability theory.

In order to develop a theory of probability we must next assign probabilities to events. For example, what is the probability that the tossed die will produce an even outcome? Denoting this probability by $P[E_{\text{even}}]$, we would intuitively say that it is 1/2 since there are 3 chances out of 6 to produce an even outcome. Note that P is a *probability function* or a function that assigns a number between 0 and 1 to sets. It is sometimes called a *set function.* The reader is familiar with ordinary functions such as $g(x) = \exp(x)$, in which a number y, where $y = g(x)$, is assigned to each x

for $-\infty < x < \infty$, and where each x is a *distinct number*. The probability function must assign a number to every event, or to *every set*. For a coin toss whose outcome is either a head H or a tail T, all the events are $E_1 = \{H\}$, $E_2 = \{T\}$, $E_3 = \mathcal{S}$, and $E_4 = \emptyset$. For a die toss all the events are $E_0 = \emptyset$, $E_1 = \{1\}, \ldots, E_6 = \{6\}$, $E_{12} = \{1,2\}, \ldots, E_{56} = \{5,6\}, \ldots, E_{12345} = \{1,2,3,4,5\}, \ldots, E_{23456} = \{2,3,4,5,6\}$, $E_{123456} = \{1,2,3,4,5,6\} = \mathcal{S}$. There are a total of 64 events. In general, if the sample space has N simple events, the total number of events is 2^N (see Problem 3.15). We must be able to assign probabilities to all of these. In accordance with our intuitive notion of probability we assign a number, either zero or positive, to each event. Hence, we require that

Axiom 1 $P[E] \geq 0$ for every event E.

Also, since the die toss will always produce an outcome that is included in $\mathcal{S} = \{1,2,3,4,5,6\}$ we should require that

Axiom 2 $P[\mathcal{S}] = 1$.

Next we might inquire as to the assignment of a probability to the event that the die comes up *either* less than or equal to 2 *or* equal to 3. Intuitively, we would say that it is 3/6 since

$$\begin{aligned} P[\{1,2\} \cup \{3\}] &= P[\{1,2\}] + P[\{3\}] \\ &= \frac{2}{6} + \frac{1}{6} = \frac{1}{2}. \end{aligned}$$

However, we would *not* assert that the probability of the die coming up *either* less than or equal to 3 *or* equal to 3 is

$$\begin{aligned} P[\{1,2,3\} \cup \{3\}] &= P[\{1,2,3\}] + P[\{3\}] \\ &= \frac{3}{6} + \frac{1}{6} = \frac{4}{6}. \end{aligned}$$

This is because the event $\{1,2,3\} \cup \{3\}$ is just $\{1,2,3\}$ (we should not count the 3 twice) and so the probability should be 1/2. In the first example, the events are *mutually exclusive* (the sets are disjoint) while in the second example they are not. Hence, the probability of an event that is the union of two mutually exclusive events should be the sum of the probabilities. Combining this axiom with the previous ones produces the full set of axioms, which we summarize next for convenience.

Axiom 1 $P[E] \geq 0$ for every event E

Axiom 2 $P[\mathcal{S}] = 1$

Axiom 3 $P[E \cup F] = P[E] + P[F]$ for E and F mutually exclusive.

Using induction (see Problem 3.17) the third axiom may be extended to

Axiom 3′ $P[\bigcup_{i=1}^{N} E_i] = \sum_{i=1}^{N} P[E_i]$ for all E_i's mutually exclusive.

The acceptance of these axioms as the basis for probability is called the *axiomatic approach to probability*. It is remarkable that these three axioms, along with a fourth axiom to be introduced later, are adequate to formulate the entire theory. We now illustrate the application of these axioms to probability calculations.

Example 3.3 – Die toss

Determine the probability that the outcome of a fair die toss is even. The event is $E_{\text{even}} = \{2, 4, 6\}$. The assumption that the die is fair means that each outcome must be *equally likely*. Defining E_i as the simple event $\{i\}$ we note that

$$\mathcal{S} = \bigcup_{i=1}^{6} E_i$$

and from Axiom 2 we must have

$$P\left[\bigcup_{i=1}^{6} E_i\right] = P[\mathcal{S}] = 1. \tag{3.8}$$

But since each E_i is a simple event and by definition the simple events are mutually exclusive (only one outcome or simple event can occur), we have from Axiom 3′ that

$$P\left[\bigcup_{i=1}^{6} E_i\right] = \sum_{i=1}^{6} P[E_i]. \tag{3.9}$$

Next we note that the outcomes are assumed to be equally likely which means that $P[E_1] = P[E_2] = \cdots = P[E_6] = p$. Hence, we must have from (3.8) and (3.9) that

$$\sum_{i=1}^{6} P[E_i] = 6p = 1$$

or $P[E_i] = 1/6$ for all i. We can now finally determine $P[E_{\text{even}}]$ since $E_{\text{even}} = E_2 \cup E_4 \cup E_6$. By applying Axiom 3′ once again we have

$$P[E_{\text{even}}] = P[E_2 \cup E_4 \cup E_6] = P[E_2] + P[E_4] + P[E_6] = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

◇

In general, the probabilities assigned to each simple event need not be the same, i.e., the outcomes of a die toss may not have equal probabilities. One might have weighted the die so that the number 6 comes up twice as often as all the others. The numbers $1, 2, 3, 4, 5$ could still be equally likely. In such a case, since the probabilities of the all the simple events must sum to one, we would have the assignment $P[\{i\}] =$

$1/7$ for $i = 1,2,3,4,5$ and $P[\{6\}] = 2/7$. In either case, to compute the probability of *any event* it is only necessary to sum the probabilities of the simple events that make up that event. Letting $P[\{s_i\}]$ be the probability of the ith simple event we have that

$$P[E] = \sum_{\{i: s_i \in E\}} P[\{s_i\}]. \tag{3.10}$$

We now simplify the notation by omitting the $\{\,\}$ when referring to events. Instead of $P[\{1\}]$ we will use $P[1]$. Another example follows.

Example 3.4 – Defective die toss

A defective die is tossed whose sides have been mistakenly manufactured with the number of dots being $1,1,2,2,3,4$. The simple events are $s_1 = 1$, $s_2 = 1$, $s_3 = 2$, $s_4 = 2$, $s_5 = 3$, $s_6 = 4$. Even though some of the outcomes have the same number of dots, they are actually different in that a *different side* is being observed. Each side is equally likely to appear. What is the probability that the outcome is less than 3? Noting that the event of interest is $\{s_1, s_2, s_3, s_4\}$, we use (3.10) to obtain

$$P[E] = P[\text{outcome} < 3] = \sum_{i=1}^{4} P[s_i] = \frac{4}{6}.$$

$\diamondsuit$

The formula given by (3.10) also applies to probability problems for which the sample space is countably infinite. Therefore, it applies to all *discrete* sample spaces (see also Example 3.2).

Example 3.5 – Countably infinite sample space

A habitually tardy person arrives at the theater late by s_i minutes, where

$$s_i = i \quad i = 1,2,3\ldots.$$

If $P[s_i] = (1/2)^i$, what is the probability that he will be more than 1 minute late? The event is $E = \{2,3,4,\ldots\}$. Using (3.10) we have

$$P[E] = \sum_{i=2}^{\infty} \left(\frac{1}{2}\right)^i.$$

Using the formula for the sum of a geometric progression (see Appendix B)

$$\sum_{i=k}^{\infty} a^i = \frac{a^k}{1-a} \quad \text{for } |a| < 1$$

we have that

$$P[E] = \frac{\left(\frac{1}{2}\right)^2}{1-\frac{1}{2}} = \frac{1}{2}.$$

$\diamondsuit$

In the above example we have implicitly used the relationship

$$P\left[\bigcup_{i=1}^{\infty} E_i\right] = \sum_{i=1}^{\infty} P[E_i] \tag{3.11}$$

where $E_i = \{s_i\}$ and hence the E_i's are mutually exclusive. This does not automatically follow from Axiom 3′ since N is now infinite. However, we will assume for our problems of interest that it does. Adding (3.11) to our list of axioms we have

Axiom 4 $P[\bigcup_{i=1}^{\infty} E_i] = \sum_{i=1}^{\infty} P[E_i]$ for all E_i's mutually exclusive.

See [Billingsley 1986] for further details.

3.5 Properties of the Probability Function

From the four axioms we may derive many useful properties for evaluating probabilities. We now summarize these properties.

Property 3.1 – Probability of complement event

$$P[E^c] = 1 - P[E]. \tag{3.12}$$

Proof: By definition $E \cup E^c = \mathcal{S}$. Also, by definition E and E^c are mutually exclusive. Hence,

$$\begin{aligned} 1 &= P[\mathcal{S}] && \text{(Axiom 2)} \\ &= P[E \cup E^c] && \text{(definition of complement set)} \\ &= P[E] + P[E^c] && \text{(Axiom 3)} \end{aligned}$$

from which (3.12) follows.

□

We could have determined the probability in Example 3.5 without the use of the geometric progression formula by using $P[E] = 1 - P[E^c] = 1 - P[1] = 1/2$.

Property 3.2 – Probability of impossible event

$$P[\emptyset] = 0. \tag{3.13}$$

Proof: Since $\emptyset = \mathcal{S}^c$ we have

$$\begin{aligned} P[\emptyset] &= P[\mathcal{S}^c] \\ &= 1 - P[\mathcal{S}] && \text{(from Property 3.1)} \\ &= 1 - 1 && \text{(from Axiom 2)} \\ &= 0. \end{aligned}$$

□

We will see later that there are other events for which the probability can be zero. Thus, the converse is *not true.*

Property 3.3 – All probabilities are between 0 and 1.

Proof:

$$\begin{aligned} \mathcal{S} &= E \cup E^c && \text{(definition of complement set)} \\ P[\mathcal{S}] &= P[E] + P[E^c] && \text{(Axiom 3)} \\ 1 &= P[E] + P[E^c] && \text{(Axiom 2)} \end{aligned}$$

But from Axiom 1 $P[E^c] \geq 0$ and therefore

$$P[E] = 1 - P[E^c] \leq 1. \tag{3.14}$$

Combining this result with Axiom 1 proves Property 3.3.

□

Property 3.4 – Formula for $P[E \cup F]$ where E and F are not mutually exclusive

$$P[E \cup F] = P[E] + P[F] - P[EF]. \tag{3.15}$$

(We have shortened $E \cap F$ to EF.)
Proof: By the definition of $E - F$ we have that $E \cup F = (E - F) \cup F$ (see Figure 3.1d,f). Also, the events $E - F$ and F are by definition mutually exclusive. It follows that

$$P[E \cup F] = P[E - F] + P[F] \quad \text{(Axiom 3)}. \tag{3.16}$$

But by definition $E = (E - F) \cup EF$ (draw a Venn diagram) and $E - F$ and EF are mutually exclusive. Thus,

$$P[E] = P[E - F] + P[EF] \qquad \text{(Axiom 3)}. \tag{3.17}$$

Combining (3.16) and (3.17) produces Property 3.4.

□

The effect of this formula is to make sure that the intersection EF is not counted twice in the probability calculation. This would be the case if Axiom 3 were mistakenly applied to sets that were not mutually exclusive. In the die example, if we wanted the probability of the die coming up either *less than or equal to 3* or *equal to 3*, then we would first define

$$\begin{aligned} E &= \{1, 2, 3\} \\ F &= \{3\} \end{aligned}$$

so that $EF = \{3\}$. Using Property 3.4, we have that

$$P[E \cup F] = P[E] + P[F] - P[EF] = \frac{3}{6} + \frac{1}{6} - \frac{1}{6} = \frac{3}{6}.$$

Of course, we could just as easily have noted that $E \cup F = \{1,2,3\} = E$ and then applied (3.10). Another example follows.

Example 3.6 – Switches in parallel

A switching circuit shown in Figure 3.6 consists of two potentially faulty switches in parallel. In order for the circuit to operate properly at least one of the switches must

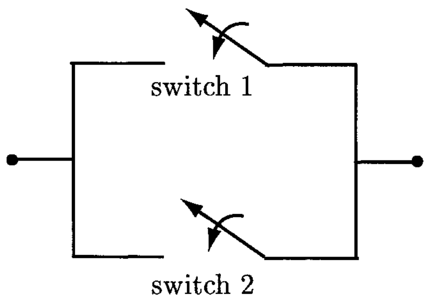


Figure 3.6: Parallel switching circuit.

close to allow the overall circuit to be closed. Each switch has a probability of 1/2 of closing. The probability that both switches close simultaneously is 1/4. What is the probability that the switching circuit will operate correctly? To solve this problem we first define the events $E_1 = \{\text{switch 1 closes}\}$ and $E_2 = \{\text{switch 2 closes}\}$. The event that *at least one* switch closes is $E_1 \cup E_2$. This includes the possibility that both switches close. Then using Property 3.4 we have

$$\begin{aligned} P[E_1 \cup E_2] &= P[E_1] + P[E_2] - P[E_1E_2] \\ &= \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}. \end{aligned}$$

Note that by using two switches in parallel as opposed to only one switch, the probability that the circuit will operate correctly has been increased. What do you think would happen if we had used three switches in parallel? Or if we had used N switches? Could you ever be assured that the circuit would operate flawlessly? (See Problem 3.26.)

◊

Property 3.5 – Monotonicity of probability function

Monotonicity asserts that the larger the set, the larger the probability of that set. Mathematically, this translates into the statement that if $E \subset F$, then $P[E] \leq P[F]$.

Proof: If $E \subset F$, then by definition $F = E \cup (F - E)$, where E and $F - E$ are mutually exclusive by definition. Hence,

$$\begin{aligned} P[F] &= P[E] + P[F-E] \quad \text{(Axiom 3)} \\ &\geq P[E] \quad \text{(Axiom 1).} \end{aligned}$$

□

Note that since $EF \subset F$ and $EF \subset E$, we have that $P[EF] \leq P[E]$ and also that $P[EF] \leq P[F]$. The probability of an intersection is always less than or equal to the probability of the set with the smallest probability.

Example 3.7 – Switches in series

A switching circuit shown in Figure 3.7 consists of two potentially faulty switches in series. In order for the circuit to operate properly *both* switches must close. For the

Figure 3.7: Series switching circuit.

same switches as described in Example 3.6 what is the probability that the circuit will operate properly? Now we need to find $P[E_1E_2]$. This was given as $1/4$ so that

$$\frac{1}{4} = P[E_1E_2] \leq P[E_1] = \frac{1}{2}$$

Could the series circuit ever outperform the parallel circuit? (See Problem 3.27.)

◊

One last property that is often useful is the probability of a union of more than two events. This extends Property 3.4. Consider first three events so that we wish to derive a formula for $P[E_1 \cup E_2 \cup E_3]$, which is equivalent to $P[(E_1 \cup E_2) \cup E_3]$ or $P[E_1 \cup (E_2 \cup E_3)]$ by the associative property. Writing this as $P[E_1 \cup (E_2 \cup E_3)]$, we have

$$\begin{aligned} P[E_1 \cup E_2 \cup E_3] &= P[E_1 \cup (E_2 \cup E_3)] \\ &= P[E_1] + P[E_2 \cup E_3] - P[E_1(E_2 \cup E_3)] \quad \text{(Property 3.4)} \\ &= P[E_1] + (P[E_2] + P[E_3] - P[E_2E_3]) \\ &\quad - P[E_1(E_2 \cup E_3)] \quad \text{(Property 3.4)} \end{aligned} \tag{3.18}$$

But $E_1(E_2 \cup E_3) = E_1E_2 \cup E_1E_3$ by the distributive property (draw a Venn diagram) so that

$$\begin{aligned} P[E_1(E_2 \cup E_3)] &= P[E_1E_2 \cup E_1E_3] \\ &= P[E_1E_2] + P[E_1E_3] - P[E_1E_2E_3] \quad \text{(Property 3.4).} \end{aligned} \tag{3.19}$$

Substituting (3.19) into (3.18) produces

$$P[E_1 \cup E_2 \cup E_3] = P[E_1]+P[E_2]+P[E_3]-P[E_2E_3]-P[E_1E_2]-P[E_1E_3]+P[E_1E_2E_3] \tag{3.20}$$

which is the desired result. It can further be shown that (see Problem 3.29)

$$P[E_1E_2] + P[E_1E_3] + P[E_2E_3] \geq P[E_1E_2E_3]$$

so that

$$P[E_1 \cup E_2 \cup E_3] \leq P[E_1] + P[E_2] + P[E_3] \tag{3.21}$$

which is known as *Boole's inequality* or the *union bound.* Clearly, equality holds if and only if the E_i's are mutually exclusive. Both (3.20) and (3.21) can be extended to any finite number of unions [Ross 2002].

3.6 Probabilities for Continuous Sample Spaces

We have introduced the axiomatic approach to probability and illustrated the approach with examples from a discrete sample space. The axiomatic approach is completely general and applies to continuous sample spaces as well. However, (3.10) cannot be used to determine probabilities of events. This is because the simple events of the continuous sample space are not countable. For example, suppose one throws a dart at a "linear" dartboard as shown in Figure 3.8 and measures the horizontal distance from the "bullseye" or center at $x = 0$. We will then have a sample space

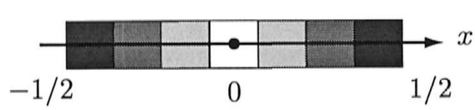


Figure 3.8: "Linear" dartboard.

$\mathcal{S} = \{x : -1/2 \leq x \leq 1/2\}$, which is not countable. A possible approach is to assign probabilities to *intervals* as opposed to sample points. If the dart is equally likely to land anywhere, then we could assign the interval $[a, b]$ a probability equal to the length of the interval or

$$P[a \leq x \leq b] = b - a \qquad -1/2 \leq a \leq b \leq 1/2. \tag{3.22}$$

Also, we will *assume* that the probability of disjoint intervals is the sum of the probabilities for each interval. This assignment is entirely consistent with our axioms since

$$\begin{aligned}
P[E] &= P[a \leq x \leq b] = b - a \geq 0. && \text{(Axiom 1)} \\
P[\mathcal{S}] &= P[-1/2 \leq x \leq 1/2] = 1/2 - (-1/2) = 1. && \text{(Axiom 2)} \\
P[E \cup F] &= P[a \leq x \leq b \cup c \leq x \leq d] \\
&= (b - a) + (d - c) && \text{(assumption)} \\
&= P[a \leq x \leq b] + P[c \leq x \leq d] \\
&= P[E] + P[F] && \text{(Axiom 3)}
\end{aligned}$$

for $a \leq b < c \leq d$ so that E and F are mutually exclusive. Hence, an equally likely type probability assignment for a continuous sample space is a valid one and produces a probability equal to the length of the interval. If the sample space does not have unity length, as for example, a dartboard with a length L, then we should use

$$P[E] = \frac{\text{Length of interval}}{\text{Length of dartboard}} = \frac{\text{Length of interval}}{L}. \tag{3.23}$$

Probability of a bullseye

It is an inescapable fact that the probability of the dart landing at say $x = 0$ is zero since the length of this interval is zero. For that matter the probability of the dart landing at any one particular point x_0 is zero as follows from (3.22) with $a = b = x_0$. The first-time reader of probability will find this particularly disturbing and argue that "How can the probability of landing at every point be zero if indeed the dart had to land at *some* point?" From a pragmatic viewpoint we will seldom be interested in probabilities of points in a continuous sample space but only in those of intervals. How many darts are there whose tips have width zero and so can be said to land at a point? It is more realistic in practice then to ask for the probability that the dart lands in the bullseye, which is a small interval with some nonzero length. That probability is found by using (3.22). From a mathematical viewpoint it is not possible to "sum" up an infinite number of positive numbers of *equal value* and not obtain infinity, as opposed to one, as assumed in Axiom 2. The latter is true for continuous sample spaces, in which we have an uncountably infinite set, and also for discrete sample spaces, which is composed of a infinite but countable set. (Note that in Example 3.5 we had a countably infinite sample space *but* the probabilities were not equal.)

Since the probability of a point event occurring is zero, the probability of any interval is the same whether or not the endpoints are included. Thus, for our example

$$P[a \leq x \leq b] = P[a < x \leq b] = P[a \leq x < b] = P[a < x < b].$$

3.7 Probabilities for Finite Sample Spaces – Equally Likely Outcomes

We now consider in more detail a discrete sample space with a finite number of outcomes. Some examples that we are already familiar with are a coin toss, a die toss, or the students in a class. Furthermore, we assume that the simple events or outcomes are *equally likely.* Many problems have this structure and can be approached using *counting methods* or *combinatorics.* For example, if two dice are tossed, then the sample space is

$$\mathcal{S} = \{(i,j) : i = 1, \ldots, 6; j = 1, \ldots, 6\}$$

which consists of 36 outcomes with each outcome or simple event denoted by an ordered pair of numbers. If we wish to assign probabilities to events, then we need only assign probabilities to the simple events and then use (3.10). But if all the simple events, denoted by s_{ij}, are equally likely, then

$$P[s_{ij}] = \frac{1}{N_{\mathcal{S}}} = \frac{1}{36}$$

where $N_{\mathcal{S}}$ is the number of outcomes in $\mathcal{S}$. Now using (3.10) we have for any event that

$$\begin{aligned} P[E] &= \sum\sum_{\{(i,j):\ s_{ij} \in E\}} P[s_{ij}] \\ &= \sum\sum_{\{(i,j):\ s_{ij} \in E\}} \frac{1}{N_{\mathcal{S}}} \\ &= \frac{N_E}{N_{\mathcal{S}}} \\ &= \frac{\text{Number of outcomes in } E}{\text{Number of outcomes in } \mathcal{S}}. \end{aligned} \tag{3.24}$$

We will use combinatorics to determine N_E and $N_{\mathcal{S}}$ and hence $P[E]$.

Example 3.8 – Probability of equal values for two-dice toss

Each outcome with equal values is of the form (i,i) so that

$$P[E] = \frac{\text{Number of outcomes with } (i,i)}{\text{Total number of outcomes}}.$$

There are 6 outcomes with equal values or (i,i) for $i = 1, 2, \ldots, 6$. Thus,

$$P[E] = \frac{6}{36} = \frac{1}{6}.$$

◇

Example 3.9 – A more challenging problem - urns
An urn contains 3 red balls and 2 black balls. Two balls are chosen in succession. The first ball is returned to the urn before the second ball is chosen. Each ball is chosen *at random*, which means that we are equally likely to choose any ball. What is the probability of choosing first a red ball and then a black ball? To solve this problem we first need to define the sample space. To do so we assign numbers to the balls as follows. The red balls are numbered $1, 2, 3$ and the black balls are numbered $4, 5$. The sample space is then $\mathcal{S} = \{(i,j) : i = 1,2,3,4,5; j = 1,2,3,4,5\}$. The event of interest is $E = \{(i,j) : i = 1,2,3; j = 4,5\}$. We assume that all the simple events are equally likely. An enumeration of the outcomes is shown in Table 3.2. The outcomes with the asterisks comprise E. Hence, the probability is $P[E] = 6/25$. This problem could also have been solved using combinatorics as follows. Since there

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
$i=1$	$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)^*$	$(1,5)^*$
$i=2$	$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)^*$	$(2,5)^*$
$i=3$	$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)^*$	$(3,5)^*$
$i=4$	$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$	$(4,5)$
$i=5$	$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$

Table 3.2: Enumeration of outcomes for urn problem of Example 3.9.

are 5 possible choices for each ball, there are a total of $5^2 = 25$ outcomes in the sample space. There are 3 possible ways to choose a red ball on the first draw and 2 possible ways to choose a black ball on the second draw, yielding a total of $3 \cdot 2 = 6$ possible ways of choosing a red ball *followed* by a black ball. We thus arrive at the same probability.

◇

3.8 Combinatorics

Combinatorics is the study of counting. As illustrated in Example 3.9, we often have an outcome that can be represented as a 2-tuple or (z_1, z_2), where z_1 can take on one of N_1 values and z_2 can take on one of N_2 values. For that example, the total number of 2-tuples in $\mathcal{S}$ is $N_1 N_2 = 5 \cdot 5 = 25$, while that in E is $N_1 N_2 = 3 \cdot 2 = 6$, as can be verified by referring to Table 3.2. It is important to note that *order matters* in the description of a 2-tuple. For example, the 2-tuple $(1,2)$ is not the same as the 2-tuple $(2,1)$ since each one describes a different outcome of the experiment. We will frequently be using 2-tuples and more generally r-tuples denoted by $(z_1, z_2, \ldots, z_r)$ to describe the outcomes of urn experiments.

In drawing balls from an urn there are two possible strategies. One method is to draw a ball, note which one it is, return it to the urn, and then draw a second ball. This is called *sampling with replacement* and was used in Example 3.9. However, it is also possible that the first ball is not returned to the urn before the second one is chosen. This method is called *sampling without replacement*. The contrast between the two strategies is illustrated next.

Example 3.10 – Computing probabilities of drawing balls from urns - with and without replacement

An urn has k red balls and $N-k$ black balls. If two balls are chosen in succession and at random *with replacement*, what is the probability of a red ball *followed* by a black ball? We solve this problem by first labeling the k red balls with $1, 2, \ldots, k$ and the black balls with $k+1, k+2, \ldots, N$. In doing so the possible outcomes of the experiment can be represented by a 2-tuple (z_1, z_2), where $z_1 \in \{1, 2, \ldots, N\}$ and $z_2 \in \{1, 2, \ldots, N\}$. A successful outcome is a red ball followed by a black one so that the successful event is $E = \{(z_1, z_2) : z_1 = 1, \ldots, k; z_2 = k+1, \ldots, N\}$. The total number of 2-tuples in the sample space is $N_S = N^2$, while the total number of 2-tuples in E is $N_E = k(N-k)$ so that

$$\begin{aligned} P[E] &= \frac{N_E}{N_S} \\ &= \frac{k(N-k)}{N^2} \\ &= \frac{k}{N}\left(1 - \frac{k}{N}\right). \end{aligned}$$

Note that if we let $p = k/N$ be the proportion of red balls, then $P[E] = p(1-p)$. Next consider the case of sampling *without replacement*. Now since the same ball cannot be chosen twice in succession, and therefore, $z_1 \neq z_2$, we have one fewer choice for the second ball. Therefore, $N_S = N(N-1)$. As before, the number of successful 2-tuples is $N_E = k(N-k)$, resulting in

$$\begin{aligned} P[E] &= \frac{k(N-k)}{N(N-1)} = \frac{k}{N}\frac{N-k}{N}\frac{N}{N-1} \\ &= p(1-p)\frac{N}{N-1}. \end{aligned}$$

The probability is seen to be higher. Can you explain this? (It may be helpful to think about the effect of a successful first draw on the probability of a success on the second draw.) Of course, for large N the probabilities for sampling with and without replacement are seen to be approximately the same, as expected.

$\diamond$

If we now choose r balls *without replacement* from an urn containing N balls, then all the possible outcomes are of the form $(z_1, z_2, \ldots, z_r)$, where the z_i's must be different. On the first draw we have N possible balls, on the second draw we have $N-1$ possible balls, etc. Hence, the total number of possible outcomes or number of r-tuples is $N(N-1)\cdots(N-r+1)$. We denote this by $(N)_r$. If all the balls are selected, forming an N-tuple, then the number of outcomes is

$$(N)_N = N(N-1)\cdots 1$$

which is defined as $N!$ and is termed *N factorial*. As an example, if there are 3 balls labeled A,B,C, then the number of 3-tuples is $3! = 3\cdot 2\cdot 1 = 6$. To verify this we have by enumeration that the possible 3-tuples are (A,B,C), (A,C,B), (B,A,C), (B,C,A), (C,A,B), (C,B,A). Note that 3! is the number of ways that 3 objects can be arranged. These arrangements are termed the *permutations* of the letters A, B, and C. Note that with the definition of a factorial we have that $(N)_r = N!/(N-r)!$. Another example follows.

Example 3.11 – More urns - using permutations

Five balls numbered $1, 2, 3, 4, 5$ are drawn from an urn *without replacement*. What is the probability that they will be drawn in the same order as their number? Each outcome is represented by the 5-tuple $(z_1, z_2, z_3, z_4, z_5)$. The only outcome in E is $(1, 2, 3, 4, 5)$ so that $N_E = 1$. To find N_S we require the number of ways that the numbers $1, 2, 3, 4, 5$ can be arranged or the number of permutations. This is $5! = 120$. Hence, the desired probability is $P[E] = 1/120$.

$\diamond$

Before continuing, we give one more example to explain our fixation with drawing balls out of urns.

Example 3.12 – The birthday problem

A probability class has N students enrolled. What is the probability that at least two of the students will have the same birthday? We first assume that each student in the class is equally likely to be born on any day of the year. To solve this problem consider a "birthday urn" that contains 365 balls. Each ball is labeled with a different day of the year. Now allow each student to select a ball at random, note its date, and return it to the urn. The day of the year on the ball becomes his/her birthday. The probability desired is of the event that two or more students choose the same ball. It is more convenient to determine the probability of the complement event or that no two students have the same birthday. Then, using Property 3.1

$$P[\text{at least 2 students have same birthday}] = 1 - P[\text{no students have same birthday}].$$

The sample space is composed of $N_S = 365^N$ N-tuples (sampling with replacement). The number of N-tuples for which all the outcomes are different is $N_E = (365)_N$. This is because the event that no two students have the same birthday occurs if the first student chooses any of the 365 balls, the second student chooses any of the remaining 364 balls, etc., which is the same as if sampling without replacement were used. The probability is then

$$P[\text{at least 2 students have same birthday}] = 1 - \frac{(365)_N}{365^N}.$$

This probability is shown in Figure 3.9 as a function of the number of students. It is seen that if the class has 23 or more students, there is a probability of 0.5 or greater that two students will have the same birthday.

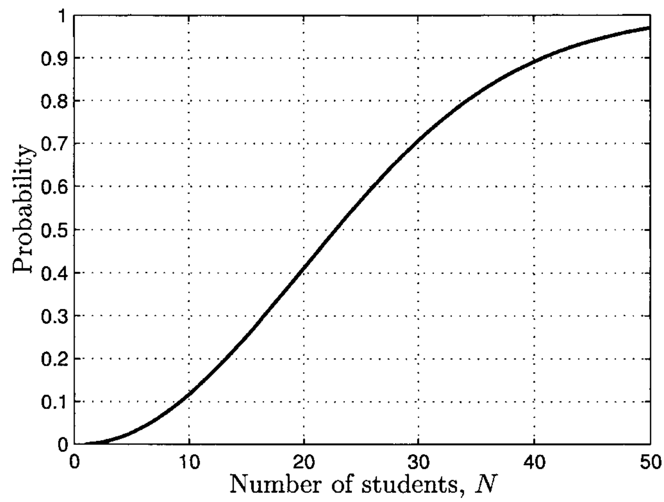


Figure 3.9: Probability of at least two students having the same birthday.

◇

Why this doesn't appear to make sense.

This result may seem counterintuitive at first, but this is only because the reader is misinterpreting the question. Most persons would say that you need about 180 people for a 50% chance of two identical birthdays. In contrast, if the question was posed as to the probability that at least two persons were born on January 1, then the event would be at least two persons choose the ball labeled "January 1" from the birthday urn. For 23 people this probability is considerably smaller (see Problem 3.38). It is the possibility that the two identical birthdays *can occur on any day of the year* (365 possibilities) that leads to the unexpected large probability. To verify this result the MATLAB program given below can be used. When run, the estimated probability for 10,000 repeated experiments was 0.5072. The reader may wish to reread Section 2.4 at this point.

```
% birthday.m
%
clear all
rand('state',0)
BD=[0:365]';
event=zeros(10000,1); % initialize to no successful events
for ntrial=1:10000
for i=1:23
   x(i,1)=ceil(365*rand(1,1)); % chooses birthdays at random
                               % (ceil rounds up to nearest integer)
end
y=sort(x); % arranges birthdays in ascending order
z=y(2:23)-y(1:22); % compares successive birthdays to each other
w=find(z==0); % flags same birthdays
if length(w)>0
   event(ntrial)=1; % event occurs if one or more birthdays the same
end
end
prob=sum(event)/10000
```

We summarize our counting formulas so far. Each outcome of an experiment produces an r-tuple, which can be written as $(z_1, z_2, \ldots, z_r)$. If we are choosing balls in succession from an urn containing N balls, then with replacement each z_i can take on one of N possible values. The number of possible r-tuples is then N^r. If we sample without replacement, then the number of r-tuples is only $(N)_r = N(N-1)\cdots(N-r+1)$. If we sample without replacement and $r = N$ or all the balls are chosen, then the number of r-tuples is $N!$. In arriving at these formulas we have used the r-tuple representation in which the *ordering* is used in the counting. For example, the 3-tuple (A,B,C) is different than (C,A,B), which is different than (C,B,A), etc. In fact, there are 3! possible orderings or permutations of the letters A, B, and C. We are frequently not interested in the ordering but only in the number of *distinct* elements. An example might be to determine the number of possible sum-values that can be made from one penny (p), one nickel (n), and one dime (d) if two coins are chosen. To determine this we use a *tree diagram* as shown in Figure 3.10. Note that since this is essentially sampling without replacement, we cannot have the outcomes pp, nn, or dd (shown in Figure 3.10 as dashed). The number of possible outcomes are 3 for the first coin and 2 for the second so that as usual there are $(3)_2 = 3 \cdot 2 = 6$ outcomes. However, *only 3 of these are distinct* or produce different sum-values for the two coins. The outcome (p,n) is counted the same as (n,p) for example. Hence, the ordering of the outcome does not matter. Both orderings are treated as the *same outcome*. To remind us that

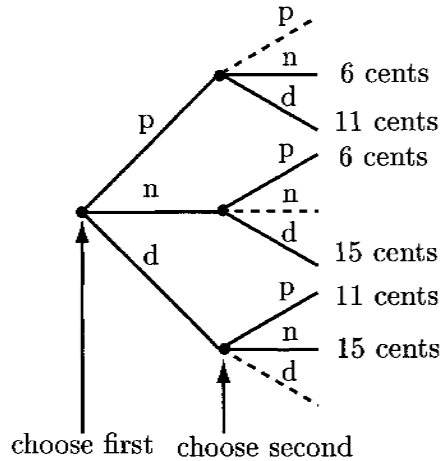


Figure 3.10: Tree diagram enumerating possible outcomes.

ordering is immaterial we will replace the 2-tuple description by the set description (recall that the elements of a set may be arranged in any order to yield the same set). The outcomes of this experiment are therefore {p,n}, {p,d}, {n,d}. In effect, all permutations are considered as a *single combination*. Thus, to find the number of combinations:

$$\text{Number of combinations} \times \text{Number of permutations} = \begin{array}{l}\text{Total number of}\\ r\text{-tuple outcomes}\end{array}$$

or for this example,

$$\text{Number of combinations} \times 2! = (3)_2$$

which yields

$$\text{Number of combinations} = \frac{(3)_2}{2!} = \frac{3!}{1!2!} = 3.$$

The number of combinations is given by the symbol $\binom{3}{2}$ and is said to be "3 things taken 2 at a time". Also, $\binom{3}{2}$ is termed the binomial coefficient due to its appearance in the binomial expansion (see Problem 3.43). In general the number of combinations of N things taken k at a time, i.e., order does not matter, is

$$\binom{N}{k} = \frac{(N)_k}{k!} = \frac{N!}{(N-k)!k!}.$$

Example 3.13 – Correct change

If a person has a penny, nickel, and dime in his pocket and selects two coins at random, what is the probability that the sum-value will be 6 cents? The sample space is now $\mathcal{S} = \{\{\text{p}, \text{n}\}, \{\text{p}, \text{d}\}, \{\text{n}, \text{d}\}\}$ and $E = \{\{\text{p}, \text{n}\}\}$. Thus,

$$\begin{aligned} P[6 \text{ cents}] &= P[\{\text{p}, \text{n}\}] = \frac{N_E}{N_{\mathcal{S}}} \\ &= \frac{1}{3}. \end{aligned}$$

Note that each simple event is of the form $\{\cdot, \cdot\}$. Also, $N_{\mathcal{S}}$ can be found from the original problem statement as $\binom{3}{2} = 3$.

◇

Example 3.14 – How probable is a royal flush?

A person draws 5 cards from a deck of 52 freshly shuffled cards. What is the probability that he obtains a royal flush? To obtain a royal flush he must draw an ace, king, queen, jack, and ten of the same suit *in any order*. There are 4 possible suits that will be produce the flush. The total number of combinations of cards or "hands" that can be drawn is $\binom{52}{5}$ and a royal flush will result from 4 of these combinations. Hence,

$$P[\text{royal flush}] = \frac{4}{\binom{52}{5}} \approx 0.00000154.$$

◇

Ordered vs. unordered

It is sometimes confusing that $\binom{52}{5}$ is used for $N_{\mathcal{S}}$. It might be argued that the first card can be chosen in 52 ways, the second card in 51 ways, etc. for a total of $(52)_5$ possible outcomes. Likewise, for a royal flush in hearts we can choose any of 5 cards, followed by any of 4 cards, etc. for a total of 5! possible outcomes. Hence, the probability of a royal flush in hearts should be

$$P[\text{royal flush in hearts}] = \frac{5!}{(52)_5}.$$

But this is just the same as $1/\binom{52}{5}$ which is the same as obtained by counting combinations. In essence, we have reduced the sample space by a factor of 5! but additionally each event is commensurately reduced by 5!, yielding the same probability. Equivalently, we have grouped together each set of 5! permutations to yield a single combination.

3.9 Binomial Probability Law

In Chapter 1 we cited the binomial probability law for the number of heads obtained for N tosses of a coin. The same law also applies to the problem of drawing balls from an urn. First, however, we look at a related problem that is of considerable practical interest. Specifically, consider an urn consisting of a *proportion* p of red balls and the remaining proportion $1-p$ of black balls. What is the probability of drawing k red balls in M drawings *without replacement*? Note that we can associate the drawing of a red ball as a "success" and the drawing of a black ball as a "failure". Hence, we are equivalently asking for the probability of k successes out of a maximum of M successes. To determine this probability we first assume that the urn contains N balls, of which N_R are red and N_B are black. We sample the urn by drawing M balls without replacement. To make the balls distinguishable we label the red balls as $1, 2, \ldots, N_R$ and the black ones as $N_R+1, N_R+2, \ldots, N$. The sample space is

$$\mathcal{S} = \{(z_1, z_2, \ldots, z_M) : z_i = 1, \ldots, N \text{ and no two } z_i\text{'s are the same}\}.$$

We assume that the balls are selected at random so that the outcomes are equally likely. The total number of outcomes is $N_S = (N)_M$. Hence, the probability of obtaining k red balls is

$$P[k] = \frac{N_E}{(N)_M}. \tag{3.25}$$

N_E is the number of M-tuples that contain k distinct integers in the range from 1 to N_R and $M-k$ distinct integers in the range N_R+1 to N. For example, if $N_R = 3$, $N_B = 4$ (and hence $N = 7$), $M = 4$, and $k = 2$, the red balls are contained in $\{1, 2, 3\}$, the black balls are contained in $\{4, 5, 6, 7\}$ and we choose 4 balls without replacement. A successful outcome has two red balls and two black balls. Some successful outcomes are $(1,4,2,5)$, $(1,4,5,2)$, $(1,2,4,5)$, etc. or $(2,3,4,6)$, $(2,4,3,6)$, $(2,6,3,4)$, etc. Hence, N_E is the total number of outcomes for which two of the z_i's are elements of $\{1,2,3\}$ and two of the z_i's are elements of $\{4,5,6,7\}$. To determine this number of successful M-tuples we

1. Choose the k positions of the M-tuple to place the red balls. (The remaining positions will be occupied by the black balls.)

2. Place the N_R red balls in the k positions obtained from step 1.

3. Place the N_B black balls in the remaining $M-k$ positions.

Step 1 is accomplished in $\binom{M}{k}$ ways since any permutation of the chosen positions produces the same set of positions. Step 2 is accomplished in $(N_R)_k$ ways and step

3 is accomplished in $(N_B)_{M-k}$ ways. Thus, we have that

$$\begin{aligned} N_E &= \binom{M}{k}(N_R)_k(N_B)_{M-k} \\ &= \frac{M!}{(M-k)!k!}(N_R)_k(N_B)_{M-k} \\ &= M!\binom{N_R}{k}\binom{N_B}{M-k} \end{aligned} \tag{3.26}$$

so that finally we have from (3.25)

$$\begin{aligned} P[k] &= \frac{M!\binom{N_R}{k}\binom{N_B}{M-k}}{(N)_M} \\ &= \frac{\binom{N_R}{k}\binom{N_B}{M-k}}{\binom{N}{M}}. \end{aligned} \tag{3.27}$$

This law is called the *hypergeometric law* and describes the probability of k successes when sampling *without replacement* is used. If sampling *with replacement* is used, then the binomial law results. However, instead of repeating the entire derivation for sampling with replacement, we need only assume that N is large. Then, whether the balls are replaced or not will not affect the probability. To show that this is indeed the case, we start with the expression given by (3.26) and note that for N large and $M \ll N$, then $(N)_M \approx N^M$. Similarly, we assume that $M \ll N_R$ and $M \ll N_B$ and make similar approximations. As a result we have from (3.25) and (3.26)

$$\begin{aligned} P[k] &\approx \binom{M}{k}\frac{N_R^k N_B^{M-k}}{N^M} \\ &= \binom{M}{k}\left(\frac{N_R}{N}\right)^k\left(\frac{N_B}{N}\right)^{M-k}. \end{aligned}$$

Letting $N_R/N = p$ and $N_B/N = (N - N_R)/N = 1 - p$, we have at last the *binomial law*

$$P[k] = \binom{M}{k}p^k(1-p)^{M-k}. \tag{3.28}$$

To summarize, the binomial law not only applies to the drawing of balls from urns *with replacement* but also applies to the drawing of balls *without replacement if the number of balls in the urn is large.* We next use our results in a quality control application.

3.10 Real-World Example – Quality Control

A manufacturer of electronic memory chips produces batches of 1000 chips for shipment to computer companies. To determine if the chips meet specifications the manufacturer initially tests all 1000 chips in each batch. As demand for the chips grows, however, he realizes that it is impossible to test all the chips and so proposes that only a subset or sample of the batch be tested. The criterion for acceptance of the batch is that at least 95% of the sample chips tested meet specifications. If the criterion is met, then the batch is accepted and shipped. This criterion is based on past experience of what the computer companies will find acceptable, i.e., if the batch "yield" is less than 95% the computer companies will not be happy. The production manager proposes that a sample of 100 chips from the batch be tested and if 95 or more are deemed to meet specifications, then the batch is judged to be acceptable. However, a quality control supervisor argues that even if only 5 of the sample chips are defective, then it is still quite probable that the batch will not have a 95% yield and thus be defective.

The quality control supervisor wishes to convince the production manager that a defective batch can frequently produce 5 or fewer defective chips in a chip sample of size 100. He does so by determining the probability that a *defective batch* will have a chip sample with 5 or fewer defective chips as follows. He first needs to assume the proportion of chips in the defective batch that will be good. Since a good batch has a proportion of good chips of 95%, a defective batch will have a proportion of good chips of less than 95%. Since he is quite conservative, he chooses this proportion as exactly $p = 0.94$, although it may actually be less. Then, according to the production manager a batch is judged to be acceptable if the sample produces $95, 96, 97, 98, 99$, or 100 good chips. The quality control supervisor likens this problem to the drawing of 100 balls from an "chip urn" containing 1000 balls. In the urn there are $1000p$ good balls and $1000(1-p)$ bad ones. The probability of drawing 95 or more good balls from the urn is given *approximately* by the binomial probability law. We have assumed that the true law, which is hypergeometric due to the use of sampling without replacement, can be approximated by the binomial law, which assumes sampling with replacement. See Problem 3.48 for the accuracy of this approximation.

Now the defective batch will be judged as acceptable if there are 95 or more successes out of a possible 100 draws. The probability of this occurring is

$$P[k \geq 95] = \sum_{k=95}^{100} \binom{100}{k} p^k (1-p)^{100-k}$$

where $p = 0.94$. The probability $P[k \geq 95]$ versus p is plotted in Figure 3.11. For $p = 0.94$ we see that the defective batch will be accepted with a probability of about 0.45 or almost half of the defective batches will be shipped. The quality control supervisor is indeed correct. The production manager does not believe the

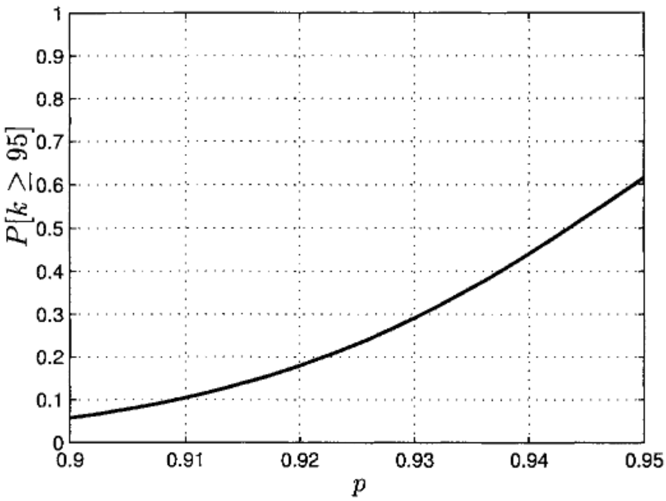


Figure 3.11: Probability of accepting a defective batch versus proportion of good chips in the defective batch – accept if 5 or fewer bad chips in a sample of 100.

result since it appears to be too high. Using sampling with replacement, which will produce results in accordance with the binomial law, he performs a computer simulation (see Problem 3.49). Based on the simulated results he reluctantly accepts the supervisor's conclusions. In order to reduce this probability the quality control supervisor suggests changing the acceptance strategy to one in which the batch is accepted only if 98 or more of the samples meet the specifications. Now the probability that the defective batch will be judged as acceptable is

$$P[k \geq 98] = \sum_{k=98}^{100} \binom{100}{k} p^k (1-p)^{100-k}$$

where $p = 0.94$, the assumed proportion of good chips in the defective batch. This produces the results shown in Figure 3.12. The acceptance probability for a defective batch is now reduced to only about 0.05.

There is a price to be paid, however, for only accepting a batch if 98 or more of the samples are good. Many more good batches will be rejected than if the previous strategy were used (see Problem 3.50). This is deemed to be a reasonable tradeoff. Note that the supervisor may well be advised to examine his initial assumption about p for the defective batch. If, for instance, he assumed that a defective batch could be characterized by $p = 0.9$, then according to Figure 3.11, the production manager's original strategy would produce a probability of less than 0.1 of accepting a defective batch.

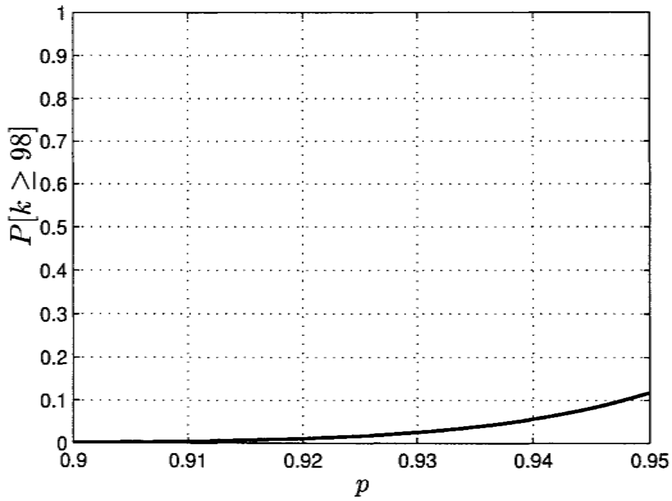


Figure 3.12: Probability of accepting a defective batch versus proportion of good chips in the defective batch – accept if 2 or fewer bad chips in a sample of 100.

References

Billingsley, P., *Probability and Measure*, John Wiley & Sons, New York, 1986.

Ross, S., *A First Course in Probability*, Prentice-Hall, Upper Saddle River, NJ, 2002.

Problems

3.1 (⌣) **(w)** The universal set is given by $\mathcal{S} = \{x : -\infty < x < \infty\}$ (the real line). If $A = \{x : x > 1\}$ and $B = \{x : x \leq 2\}$, find the following:

a. A^c and B^c

b. $A \cup B$ and $A \cap B$

c. $A - B$ and $B - A$

3.2 (w) Repeat Problem 3.1 if $\mathcal{S} = \{x : x \geq 0\}$.

3.3 (w) A group of voters go to the polling place. Their names and ages are Lisa, 21, John, 42, Ashley, 18, Susan, 64, Phillip, 58, Fred, 48, and Brad, 26. Find the following sets:

a. Voters older than 30
b. Voters younger than 30
c. Male voters older than 30
d. Female voters younger than 30
e. Voters that are male or younger than 30
f. Voters that are female and older than 30

Next find any two sets that partition the universe.

3.4 (w) Given the sets $A_i = \{x : 0 \leq x \leq i\}$ for $i = 1, 2, \ldots, N$, find $\cup_{i=1}^{N} A_i$ and $\cap_{i=1}^{N} A_i$. Are the A_i's disjoint?

3.5 (w) Prove that the sets $A = \{x : x \geq -1\}$ and $B = \{x : 2x + 2 \geq 0\}$ are equal.

3.6 (t) Prove that if $x \in A \cap B^c$, then $x \in A - B$.

3.7 (⌣) (w) If $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$, find sets A and B that are disjoint. Next find sets C and D that partition the universe.

3.8 (w) If $\mathcal{S} = \{(x, y) : 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1\}$, find sets A and B that are disjoint. Next find sets C and D that partition the universe.

3.9 (t) In this problem we see how to construct disjoint sets from ones that are not disjoint so that their unions will be the same. We consider only three sets and ask the reader to generalize the result. Calling the nondisjoint sets A, B, C and the union $D = A \cup B \cup C$, we wish to find three disjoint sets E_1, E_2, and E_3 so that $D = E_1 \cup E_2 \cup E_3$. To do so let

$$\begin{aligned} E_1 &= A \\ E_2 &= B - E_1 \\ E_3 &= C - (E_1 \cup E_2). \end{aligned}$$

Using a Venn diagram explain this procedure. If we now have sets $A_1, A_2, \ldots, A_N$, explain how to construct N disjoint sets with the same union.

3.10 (⌣) (f) Replace the set expression $A \cup B \cup C$ with one using intersections and complements. Replace the set expression $A \cap B \cap C$ with one using unions and complements.

3.11 (w) The sets A, B, C are subsets of $\mathcal{S} = \{(x, y) : 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1\}$. They are defined as

$$\begin{aligned} A &= \{(x, y) : x \leq 1/2, 0 \leq y \leq 1\} \\ B &= \{(x, y) : x \geq 1/2, 0 \leq y \leq 1\} \\ C &= \{(x, y) : 0 \leq x \leq 1, y \leq 1/2\}. \end{aligned}$$

Explicitly determine the set $A \cup (B \cap C)^c$ by drawing a picture of it as well as pictures of all the individual sets. For simplicity you can ignore the edges of the sets in drawing any diagrams. Can you represent the resultant set using only unions and complements?

3.12 (⌣̈) **(w)** Give the size of each set and also whether it is discrete or continuous. If the set is infinite, determine if it is countably infinite or not.

a. $A = \{\text{seven-digit numbers}\}$

b. $B = \{x : 2x = 1\}$

c. $C = \{x : 0 \leq x \leq 1 \text{ and } 1/2 \leq x \leq 2\}$

d. $D = \{(x, y) : x^2 + y^2 = 1\}$

e. $E = \{x : x^2 + 3x + 2 = 0\}$

f. $F = \{\text{positive even integers}\}$

3.13 (w) Two dice are tossed and the number of dots on each side that come up are added together. Determine the sample space, outcomes, impossible event, three different events including a simple event, and two mutually exclusive events. Use appropriate set notation.

3.14 (⌣̈) **(w)** The temperature in Rhode Island on a given day in August is found to always be in the range from 30° F to 100° F. Determine the sample space, outcomes, impossible event, three different events including a simple event, and two mutually exclusive events. Use appropriate set notation.

3.15 (t) Prove that if the sample space has size N, then the total number of events (including the impossible event and the certain event) is 2^N. Hint: There are $\binom{N}{k}$ ways to choose an event with k outcomes from a total of N outcomes. Also, use the binomial formula

$$(a+b)^N = \sum_{k=0}^{N} \binom{N}{k} a^k b^{N-k}$$

which was proven in Problem 1.11.

3.16 (w) An urn contains 2 red balls and 3 black balls. The red balls are labeled with the numbers 1 and 2 and the black balls are labeled as 3, 4, and 5. Three balls are drawn without replacement. Consider the events that

$$\begin{aligned} A &= \{\text{a majority of the balls drawn are black}\} \\ B &= \{\text{the sum of the numbers of the balls drawn} \geq 10\}. \end{aligned}$$

Are these events mutually exclusive? Explain your answer.

3.17 (t) Prove Axiom 3′ by using mathematical induction (see Appendix B) and Axiom 3.

3.18 (☺) **(w)** A roulette wheel has numbers 1 to 36 equally spaced around its perimeter. The odd numbers are colored red while the even numbers are colored black. If a spun ball is equally likely to yield any of the 36 numbers, what is the probability of a black number, of a red number? What is the probability of a black number that is greater than 24? What is the probability of a black number or a number greater than 24?

3.19 (☺) **(c)** Use a computer simulation to simulate the tossing of a fair die. Based on the simulation what is the probability of obtaining an even number? Does it agree with the theoretical result? Hint: See Section 2.4.

3.20 (w) A fair die is tossed. What is the probability of obtaining an even number, an odd number, a number that is even or odd, a number that is even and odd?

3.21 (☺) **(w)** A die is tossed that yields an even number with twice the probability of yielding an odd number. What is the probability of obtaining an even number, an odd number, a number that is even or odd, a number that is even and odd?

3.22 (w) If a single letter is selected at random from $\{A, B, C\}$, find the probability of all events. Recall that the total number of events is 2^N, where N is the number of simple events. Do these probabilities sum to one? If not, why not? Hint: See Problem 3.15.

3.23 (☺) **(w)** A number is chosen from $\{1, 2, 3, \ldots\}$ with probability

$$P[i] = \begin{cases} \frac{4}{7} & i = 1 \\ \frac{2}{7} & i = 2 \\ \left(\frac{1}{8}\right)^{i-2} & i \geq 3 \end{cases}$$

Find $P[i \geq 4]$.

3.24 (f) For a sample space $\mathcal{S} = \{0, 1, 2, \ldots\}$ the probability assignment

$$P[i] = \exp(-2)\frac{2^i}{i!}$$

is proposed. Is this a valid assignment?

3.25 (☺) **(w)** Two fair dice are tossed. Find the probability that only one die comes up a 6.

3.26 (w) A circuit consists of N switches in parallel (see Example 3.6 for $N = 2$). The sample space can be summarized as $\mathcal{S} = \{(z_1, z_2, \ldots, z_N) : z_i = \text{s or f}\}$, where s indicates a success or the switch closes and f indicates a failure or the switch fails to close. Assuming that all the simple events are equally likely, what is the probability that a circuit is closed when all the switches are activated to close? Hint: Consider the complement event.

3.27 (⌣̈) (w) Can the series circuit of Figure 3.7 ever outperform the parallel circuit of Figure 3.6 in terms of having a higher probability of closing when both switches are activated to close? Assume that switch 1 closes with probability p, switch 2 closes with probability p, and both switches close with probability p^2.

3.28 (w) Verify the formula (3.20) for $P[E_1 \cup E_2 \cup E_3]$ if E_1, E_2, E_3 are events that are not necessarily mutually exclusive. To do so use a Venn diagram.

3.29 (t) Prove that

$$P[E_1E_2] + P[E_1E_3] + P[E_2E_3] \geq P[E_1E_2E_3].$$

3.30 (w) A person always arrives at his job between 8:00 AM and 8:20 AM. He is equally likely to arrive anytime within that period. What is the probability that he will arrive at 8:10 AM? What is the probability that he will arrive between 8:05 and 8:10 AM?

3.31 (w) A random number generator produces a number that is equally likely to be anywhere in the interval $(0, 1)$. What are the simple events? Can you use (3.10) to find the probability that a generated number will be less than $1/2$? Explain.

3.32 (w) If two fair dice are tossed, find the probability that the same number will be observed on each one. Next, find the probability that different numbers will be observed.

3.33 (⌣̈) (w) Three fair dice are tossed. Find the probability that 2 of the numbers will be the same and the third will be different.

3.34 (w,c) An urn contains 4 red balls and 2 black balls. Two balls are chosen at random and without replacement. What is the probability of obtaining one red ball and one black ball in any order? Verify your results by enumerating all possibilities using a computer evaluation.

3.35 (⌣̈) (f) Rhode Island license plate numbers are of the form GR315 (2 letters followed by 3 digits). How many different license plates can be issued?

3.36 (f) A baby is to be named using four letters of the alphabet. The letters can be used as often as desired. How many different names are there? (Of course, some of the names may not be pronounceable).

3.37 (c) It is difficult to compute $N!$ when N is large. As an approximation, we can use Stirling's formula, which says that for large N

$$N! \approx \sqrt{2\pi} N^{N+1/2} \exp(-N).$$

Compare Stirling's approximation to the true value of $N!$ for $N = 1, 2, \ldots, 100$ using a digital computer. Next try calculating the exact value of $N!$ for $N = 200$ using a computer. Hint: Try printing out the logarithm of $N!$ and compare it to the logarithm of its approximation.

3.38 ($\ddot\smile$) (t) Determine the probability that in a class of 23 students two or more students have birthdays on January 1.

3.39 (c) Use a computer simulation to verify your result in Problem 3.38.

3.40 ($\ddot\smile$) (w) A pizza can be ordered with up to four different toppings. Find the total number of different pizzas (including no toppings) that can be ordered. Next, if a person wishes to pay for only two toppings, how many two-topping pizzas can he order?

3.41 (f) How many subsets of size three can be made from $\{A, B, C, D, E\}$?

3.42 (w) List all the combinations of two coins that can be chosen from the following coins: one penny (p), one nickel (n), one dime (d), one quarter (q). What are the possible sum-values?

3.43 (f) The binomial theorem states that

$$(a+b)^N = \sum_{k=0}^{N} \binom{N}{k} a^k b^{N-k}.$$

Expand $(a+b)^3$ and $(a+b)^4$ into powers of a and b and compare your results to the formula.

3.44 ($\ddot\smile$) (w) A deck of poker cards contains an ace, king, queen, jack, 10, 9, 8, 7, 6, 5, 4, 3, 2 in each of the four suits, hearts (h), clubs (c), diamonds (d), and spades (s), for a total of 52 cards. If 5 cards are chosen at random from a deck, find the probability of obtaining 4 of a kind, as for example, 8-h, 8-c, 8-d, 8-s, 9-c. Next find the probability of a flush, which occurs when all five cards have the same suit, as for example, 8-s, queen-s, 2-s, ace-s, 5-s.