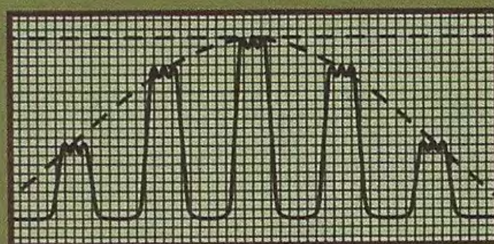


Probability, Random Variables AND Random Signal Principles

شكر خاص ل
أحمد العموش
لقيامه بعمل سكان للكتاب ..

PEYTON Z. PEEBLES, JR.

4TH EDITION
McGRAW-HILL INTERNATIONAL EDITION
Electrical Engineering Series
Mc Graw Hill

Probability, Random Variables, and Random Signal Principles

McGraw-Hill Series in Electrical and Computer Engineering

Circuits and Systems
Communications and Signal Processing
Computer Engineering
Control Theory and Robotics
Electromagnetics
Electronics and VLSI Circuits
Introductory
Power
Antennas, Microwaves, and Radar

Probability, Random Variables, and Random Signal Principles

FOURTH EDITION

Peyton Z. Peebles, Jr., Ph.D.

Professor Emeritus of Electrical and Computer Engineering
University of Florida

McGRAW-HILL, INC.

Boston Burr Ridge, IL Dubuque, IA Madison, WI New York
San Francisco St. Louis Bangkok Bogotá Caracas
Kuala Lumpur Lisbon London Madrid Mexico City
Milan Montreal New Delhi Santiago Seoul Singapore
Sydney Taipei Toronto

McGraw-Hill Higher Education
A Division of The McGraw-Hill Companies

PROBABILITY, RANDOM VARIABLES, AND RANDOM SIGNAL PRINCIPLES
International Edition 2001

Published by Irwin/McGraw-Hill, an imprint of The McGraw-Hill Companies, Inc. 1221 Avenue of the Americas, New York, NY, 10020.

20 19
20 15 14 13 12 11
CTP SLP

Library of Congress Cataloging-in-Publication Data

Peebles, Peyton Z.
Probability, random variables, and random signal principles / Peyton Z. Peebles, Jr— 4th ed.
p. cm. (McGraw-Hill series in electrical and computer engineering)
ISBN 0-07-366007-8
1. Probabilities. 2. Random variables. 3. Signal theory (Telecommunication) I. Title. II. Series.
TA340.P43 2001
519.2—dc21 00-034881

www.mhhe.com

When ordering this title, use ISBN 0-07-118181-4

Printed in Singapore

ABOUT THE AUTHOR

PEYTON Z. PEEBLES, Jr., is Professor Emeritus in the Department of Electrical and Computer Engineering at the University of Florida, Gainesville, Florida. He has twelve years of industrial experience and over thirty years of teaching experience including time spent at the University of Tennessee and the University of Hawaii, as well as the University of Florida. He earned his Ph.D. degree in 1967 from the University of Pennsylvania where he held a David Sarnoff Fellowship from RCA for two years.

Professor Peebles has authored or coauthored over 50 journal articles and conference papers as well as four other textbooks: *Radar Principles* (John Wiley & Sons, 1998); *Principles of Electrical Engineering* (McGraw-Hill, 1991, with T. A. Giuma); *Digital Communication Systems* (Prentice-Hall, 1987); and *Communication System Principles* (Addison-Wesley, 1976).

Dr. Peebles is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Pi Sigma, and Phi Beta Chi, and is a Life Fellow of the Institute of Electrical and Electronic Engineers (IEEE).

TO MY MOTHER
Maida Erlene Denton Dials

AND STEPFATHER
Ralph Phillip Dials

CONTENTS

PREFACE

In each of the book's earlier editions, all prior prefaces were reproduced to show the developments and purposes of the book over time. In this edition, they have been combined into one preface. It is, therefore, appropriate to first summarize the most important topics of prior editions that have either not changed or have evolved with time. They are:

- The *level* and *amount* of material remains that of a typical undergraduate program for courses of not more than one semester. The book is also applicable to courses at the first-year graduate level if students have had minimal or no prior exposure to the book's topics. Over the past 20 years the book has seen mostly small additions to aid teaching (more problems, more examples, and added topics that seemed to be needed in evolving programs).
- The more advanced material, and problems that require more than typical solution times, remain keyed by a star ($\star$), as before.
- The *need* for the book has not changed. Namely, a well-organized teaching book is needed for lower-level courses (juniors, seniors) as the material has migrated down from the upper graduate level over the years.
- The *background* needed to study the book remains typical of junior or senior undergraduate engineering students.

In preparation for this fourth edition, comments and suggestions were sought from many reviewers and instructors who have used the book. The result was the incorporation of a number of additions that range from relatively minor to important. In the former category, additions include: chapter-end summaries, more examples within the text, expanded discussions of probability as a relative frequency, material on permutations and combinations, more detailed material on random variable transformations, a bit more detail on ergodicity, the weak and strong laws of large numbers, sampling and estimation, various important inequalities (Chebychev, Schwarz, Markov, Hölder, Chernoff, etc.), some useful properties of impulse functions (Appendix A), a new appendix on various mathematical topics of interest (Appendix G), a few new problems, and many other small changes. Problems at each chapter end have also been combined into a single list with numbering that corresponds to the chapter's sections to which the problems mainly apply.

However, probably the most important new material relates to discrete-time (DT) random processes and sequences and other topics in the general area of digital signal processing (DSP) such as the DT linear system. There is now coverage of sampling theorems—baseband and bandpass—for random processes to establish a foundation for DT processes and sequences. For clear exposure, this material is placed mainly in Chapter 8, since it requires some understanding of the passing of random signals through networks.

Correlation functions and power spectrums for these processes and sequences are developed and connected through the sampling theorems and the discrete-time Fourier transforms (Chapters 7 and 8). The structure of linear DT digital systems is developed in both the sequence and transform domains. It is hoped this new material will better serve those readers with a strong interest in digital topics.

Other important additions are computer examples and problems scattered throughout the book at key places—mainly where new DT material is located. These examples and problems require no special toolboxes and assume the reader is familiar with the use of version 5.2 of MATLAB software. For the examples, the necessary coding is given, but the reader is expected to provide the coding for the problems. All of these examples and problems are keyed by a "computer" symbol M. With appreciation, I acknowledge the help of Mr. Kenneth Hild, a doctoral student, who worked, coded, and verified all these examples and problems.

Many persons have identified errors in the third edition and have offered suggestions for this new edition. Several of my students are included in this group, and I thank them for their contributions. I also appreciate the detailed comments and suggestions from the following professors who either have used the book, were reviewers, or both: Scott Acton, *Oklahoma State University*; Mahmood R. Azimi, *Colorado State University*; Ross Baldick, *University of Texas, Austin*; Charles Boncelet, *University of Delaware*; Oscar Norberto Bria, *Universidad Nacional de La Plata*; Kevin D. Donohue, *University of Kentucky*; Sammie Giles, Jr., *University of Toledo*; Subhash Kak, *Louisiana State University*; James Kang, *California State Polytechnic University, Pomona*; Venkatarama Krishnan, *University of Massachusetts, Lowell*; Jian Li, *University of Florida*; Rodney Roberts, *Florida A&M University* and *Florida State University*; Sumit Roy, *University of Washington*; Antal Sarkady, *U.S. Naval Academy*; Ness Shroff, *Purdue University*; Emmanouel Varvarigos, *University of California, Santa Barbara*; Donley Winger, *California State Polytechnic University, San Luis Obispo*. My thanks are also extended to Ms. Catherine Fields and Ms. Michelle Flomenhoft, editors at McGraw-Hill who were instrumental in the production of the fourth edition.

Finally, I thank those who have indicated to me that they were pleased that the third edition had relatively few errors. One even said he knew of none. Of course, there were some, but a great deal of effort did go into minimizing the number in print. Those same efforts have also been made in this new edition in hopes that it will prove as free of errors.

I especially thank Robert McDonough of the Johns Hopkins Applied Physics Laboratory for identifying a number of errors for correction.

Peyton Z. Peebles, Jr.
Gainesville, Florida
March 2000

CHAPTER 1

Probability

1.0 INTRODUCTION TO BOOK AND CHAPTER

The primary goals of this book are to introduce the reader to the principles of random signals and to provide tools whereby one can deal with systems involving such signals. Toward these goals, perhaps the first thing that should be done is define what is meant by random signal. A *random signal* is a time waveform† that can be characterized only in some probabilistic manner. In general, it can be either a desired or undesired waveform.

The reader has no doubt heard background hiss while listening to an ordinary broadcast radio receiver. The waveform causing the hiss, when observed on an oscilloscope, would appear as a randomly fluctuating voltage with time. It is undesirable, since it interferes with our ability to hear the radio program, and is called *noise*.

Undesired random waveforms (noise) also appear in the outputs of other types of systems. In a radio astronomer's receiver, noise interferes with the desired signal from outer space (which itself is a random, but desirable, signal). In a television system, noise shows up in the form of picture interference often called "snow." In a sonar system, randomly generated sea sounds give rise to a noise that interferes with the desired echoes.

The number of desirable random signals is almost limitless. For example, the bits in a computer bit stream appear to fluctuate randomly with time

†We shall usually assume random signals to be voltage-time waveforms. However, the theory to be developed throughout the book will apply, in most cases, to random functions other than voltage, of arguments other than time.

between the zero and one states, thereby creating a random signal. In another example, the output voltage of a wind-powered generator would be random because wind speed fluctuates randomly. Similarly, the voltage from a solar detector varies randomly due to the randomness of cloud and weather conditions. Still other examples are: the signal from an instrument designed to measure instantaneous ocean wave height; the space-originated signal at the output of the radio astronomer's antenna (the relative intensity of this signal from space allows the astronomer to form radio maps of the heavens); and the voltage from a vibration analyzer attached to an automobile driving over rough terrain.

In Chapters 8 and 9 we shall study methods of characterizing systems having random input signals. However, from the above examples, it is obvious that random signals only represent the behavior of more fundamental underlying random phenomena. Phenomena associated with the desired signals of the last paragraph are: information source for computer bit stream; wind speed; various weather conditions such as cloud density and size, cloud speed, etc.; ocean wave height; sources of outer space signals; and terrain roughness. All these phenomena must be described in some probabilistic way.

Thus, there are actually two things to be considered in characterizing random signals. One is how to describe any one of a variety of random phenomena; another is how to bring time into the problem so as to create the random signal of interest. To accomplish the first item, we shall introduce mathematical concepts in Chapters 2, 3, 4, and 5 (random variables) that are sufficiently general they can apply to any suitably defined random phenomena. To accomplish the second item, we shall introduce another mathematical concept, called a random process, in Chapters 6 and 7. All these concepts are based on probability theory.

The purpose of this chapter is to introduce the elementary aspects of probability theory on which all of our later work is based. Several approaches exist for the definition and discussion of probability. Only two of these are worthy of modern-day consideration, while all others are mainly of historical interest and are not commented on further here. Of the more modern approaches, one uses the relative frequency definition of probability. It gives a degree of physical insight which is popular with engineers, and is often used in texts having principal topics other than probability theory itself (for example, see Peebles, 1976).†

The second approach to probability uses the axiomatic definition. It is the most mathematically sound of all approaches and is most appropriate for a text having its topics based principally on probability theory. The axiomatic approach also serves as the best basis for readers wishing to proceed beyond the scope of this book to more advanced theory. Because of these facts, we mainly adopt the axiomatic approach in this book, but occasionally use the relative frequency method in some practical problems.

†References are quoted by name and date of publication. They are listed at the end of the book.

Prior to the introduction of the axioms of probability, it is necessary that we first develop certain elements of set theory.†

1.1 SET DEFINITIONS

A *set* is a collection of objects. The objects are called *elements* of the set and may be anything whatsoever. We may have a set of voltages, a set of airplanes, a set of chairs, or even a set of sets, sometimes called a *class* of sets. A set is usually denoted by a capital letter while an element is represented by a lower-case letter. Thus, if a is an element of set A, then we write

$$a \in A \tag{1.1-1}$$

If a is not an element of A, we write

$$a \notin A \tag{1.1-2}$$

A set is specified by the content of two braces: $\{\cdot\}$. Two methods exist for specifying content, *the tabular method* and the rule method. In the *tabular* method the elements *are enumerated* explicitly. For example, the set of all integers between 5 and 10 would be $\{6, 7, 8, 9\}$. In the rule method, a set's content is determined by some rule, such as: {integers between 5 and 10}.‡ The rule method is usually more convenient to use when the set is large. For example, {integers from 1 to 1000 inclusive} would be cumbersome to write explicitly using the tabular method.

A set is said to be *countable* if its elements can be put in one-to-one correspondence with the natural numbers, which are the integers 1, 2, 3, etc. If a set is not countable, it is called *uncountable*. A set is said to be *empty* if it has no elements. The empty set is given the symbol $\varnothing$ and is often called the *null set*.

A *finite set* is one that is either empty or has elements that can be counted, with the counting process terminating. In other words, it has a finite number of elements. If a set is not finite, it is called *infinite*. An infinite set having countable elements is called *countably infinite*.

If every element of a set A is also an element in another set B, A is said to be contained in B. A is known as a *subset* of B and we write

$$A \subseteq B \tag{1.1-3}$$

If at least one element exists in B which is not in A, then A is a *proper subset* of B, denoted by (Thomas, 1969)

$$A \subset B \tag{1.1-4}$$

The null set is clearly a subset of all other sets.

---

†Our treatment is limited to the level required to introduce the desired probability concepts. For additional details the reader is referred to McFadden (1963), or Milton and Tsokos (1976).

‡Sometimes notations such as $\{I|5 < I < 10, I \text{ an integer}\}$ or $\{I : 5 < I < 10, I \text{ an integer}\}$ are seen in the literature.

Two sets, A and B, are called *disjoint* or *mutually exclusive* if they have no common elements.

EXAMPLE 1.1-1. To illustrate the topics discussed above, we identify the sets listed below.

$$A = \{1, 3, 5, 7\} \qquad D = \{0.0\}$$
$$B = \{1, 2, 3, \ldots\} \qquad E = \{2, 4, 6, 8, 10, 12, 14\}$$
$$C = \{0.5 < c \leq 8.5\} \qquad F = \{-5.0 < f \leq 12.0\}$$

The set A is tabularly specified, countable, and finite. B is also tabularly specified and countable, but is infinite. Set C is rule-specified, uncountable, and infinite, since it contains *all* numbers greater than 0.5 but not exceeding 8.5. Similarly, sets D and E are countably finite, while set F is uncountably infinite. It should be noted that D is *not* the null set; it has one element, the number zero.

Set A is contained in sets B, C, and F. Similarly, $C \subset F$, $D \subset F$, and $E \subset B$. Sets B and F are not subsets of any of the other sets or of each other. Sets A, D, and E are mutually exclusive of each other. The reader may wish to identify which of the remaining sets are also mutually exclusive.

The largest or all-encompassing set of objects under discussion in a given situation is called the *universal set*, denoted S. All sets (of the situation considered) are subsets of the universal set. An example will help clarify the concept of a universal set.

EXAMPLE 1.1-2. Suppose we consider the problem of rolling a die. We are interested in the numbers that show on the upper face. Here the universal set is $S = \{1, 2, 3, 4, 5, 6\}$. In a gambling game, suppose a person wins if the number comes up odd. This person wins for any number in the set $A = \{1, 3, 5\}$. Another person might win if the number shows four or less; that is, for any number in the set $B = \{1, 2, 3, 4\}$.

Observe that both A and B are subsets of S. For any universal set with N elements, there are 2^N possible subsets of S. (The reader should check this for a few values of N.) For the present example, $N = 6$ and $2^N = 64$, so that there are 64 ways one can define "winning" with one die.

It should be noted that winning or losing in the above gambling game is related to a set. The game itself is partially specified by its universal set (other games typically have a different universal set). These facts are not just coincidence, and we shall shortly find that sets form the basis on which our study of probability is constructed.

1.2 SET OPERATIONS

In working with sets, it is helpful to introduce a geometrical representation that enables us to associate a physical picture with sets.

Venn Diagram

Such a representation is the Venn diagram.† Here sets are represented by closed-plane figures. Elements of the sets are represented by the enclosed points (area). The universal set S is represented by a rectangle as illustrated in Figure 1.2-1a. Three sets A, B, and C are shown. Set C is disjoint from both A and B, while set B is a subset of A.

Equality and Difference

Two sets A and B are *equal* if all elements in A are present in B and all elements in B are present in A; that is, if $A \subseteq B$ and $B \subseteq A$. For equal sets we write $A = B$.

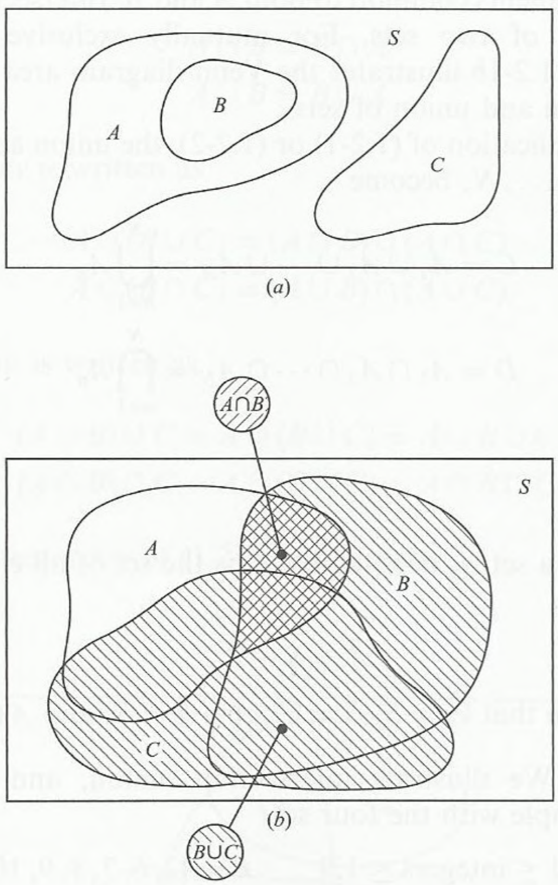


FIGURE 1.2-1
Venn diagrams. (a) Illustration of subsets and mutually exclusive sets, and (b) illustration of intersection and union of sets. [*Adapted from Peebles (1976), with permission of publishers Addison–Wesley, Advanced Book Program.*]

†After John Venn (1834–1923), an Englishman.

The *difference* of two sets A and B, denoted $A - B$, is the set containing all elements of A that are not present in B. For example, with $A = \{0.6 < a \leq 1.6\}$ and $B = \{1.0 \leq b \leq 2.5\}$, then $A - B = \{0.6 < c < 1.0\}$ or $B - A = \{1.6 < d \leq 2.5\}$. Note that $A - B \neq B - A$.

Union and Intersection

The *union* (call it C) of two sets A and B is written

$$C = A \cup B \tag{1.2-1}$$

It is the set of all elements of A or B or both. The union is sometimes called the *sum* of two sets.

The *intersection* (call it D) of two sets A and B is written

$$D = A \cap B \tag{1.2-2}$$

It is the set of all elements common to both A and B. Intersection is sometimes called the *product* of two sets. For mutually exclusive sets A and B, $A \cap B = \varnothing$. Figure 1.2-1b illustrates the Venn diagram area to be associated with the intersection and union of sets.

By repeated application of (1.2-1) or (1.2-2), the union and intersection of N sets A_n, $n = 1, 2, \ldots, N$, become

$$C = A_1 \cup A_2 \cup \cdots \cup A_N = \bigcup_{n=1}^{N} A_n \tag{1.2-3}$$

$$D = A_1 \cap A_2 \cap \cdots \cap A_N = \bigcap_{n=1}^{N} A_n \tag{1.2-4}$$

Complement

The *complement* of a set A, denoted by $\bar{A}$, is the set of all elements not in A. Thus,

$$\bar{A} = S - A \tag{1.2-5}$$

It is also easy to see that $\bar{\varnothing} = S$, $\bar{S} = \varnothing$, $A \cup \bar{A} = S$, and $A \cap \bar{A} = \varnothing$.

EXAMPLE 1.2-1. We illustrate intersection, union, and complement by taking an example with the four sets

$$S = \{1 \leq \text{integers} \leq 12\} \qquad B = \{2, 6, 7, 8, 9, 10, 11\}$$
$$A = \{1, 3, 5, 12\} \qquad C = \{1, 3, 4, 6, 7, 8\}$$

Applicable unions and intersections here are:

$$A \cup B = \{1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12\} \qquad A \cap B = \varnothing$$
$$A \cup C = \{1, 3, 4, 5, 6, 7, 8, 12\} \qquad A \cap C = \{1, 3\}$$
$$B \cup C = \{1, 2, 3, 4, 6, 7, 8, 9, 10, 11\} \qquad B \cap C = \{6, 7, 8\}$$

Complements are:

$$\bar{A} = \{2, 4, 6, 7, 8, 9, 10, 11\}$$
$$\bar{B} = \{1, 3, 4, 5, 12\}$$
$$\bar{C} = \{2, 5, 9, 10, 11, 12\}$$

The various sets are illustrated in Figure 1.2-2.

Algebra of Sets

All subsets of the universal set form an algebraic system for which a number of theorems may be stated (Thomas, 1969). Three of the most important of these relate to laws involving unions and intersections. The *commutative law* states that

$$A \cap B = B \cap A \tag{1.2-6}$$
$$A \cup B = B \cup A \tag{1.2-7}$$

The *distributive law* is written as

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \tag{1.2-8}$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \tag{1.2-9}$$

The *associative law* is written as

$$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C \tag{1.2-10}$$
$$(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C \tag{1.2-11}$$

These are just restatements of (1.2-3) and (1.2-4).

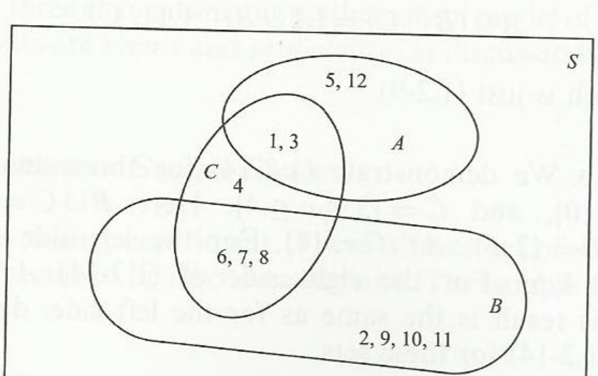


FIGURE 1.2-2
Venn diagram applicable to Example 1.2-1.

De Morgan's Laws

By use of a Venn diagram we may readily prove *De Morgan's laws*†, which state that the complement of a union (intersection) of two sets A and B equals the intersection (union) of the complements $\bar{A}$ and $\bar{B}$. Thus,

$$(\overline{A \cup B}) = \bar{A} \cap \bar{B} \tag{1.2-12}$$

$$(\overline{A \cap B}) = \bar{A} \cup \bar{B} \tag{1.2-13}$$

From the last two expressions one can show that if in an identity we replace unions by intersections, intersections by unions, and sets by their complements, then the identity is preserved (Papoulis, 1965, p. 23).

EXAMPLE 1.2-2. We verify De Morgan's law (1.2-13) by using the example sets $A = \{2 < a \leq 16\}$ and $B = \{5 < b \leq 22\}$ when $S = \{2 < s \leq 24\}$. First, if we define $C = A \cap B$, the reader can readily see from Venn diagrams that $C = A \cap B = \{5 < c \leq 16\}$ so $\bar{C} = \overline{A \cap B} = \{2 < c \leq 5, 16 < c \leq 24\}$. This result is the left side of (1.2-13).

Second, we compute $\bar{A} = S - A = \{16 < a \leq 24\}$ and $\bar{B} = S - B = \{2 < b \leq 5, 22 < b \leq 24\}$. Thus, $C = \bar{A} \cup \bar{B} = \{2 < c \leq 5, 16 < c \leq 24\}$. This result is the right side of (1.2-13) and De Morgan's law is verified.

Duality Principle

This principle (see Papoulis, 1965, for additional reading) states: if in an identity we replace unions by intersections, intersections by unions, S by $\varnothing$, and $\varnothing$ by S, then the identity is preserved. For example, since

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \tag{1.2–14}$$

is a valid identity from (1.2-8), it follows that

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \tag{1.2-15}$$

is also valid, which is just (1.2-9).

EXAMPLE 1.2-3. We demonstrate (1.2-14) for three sets $A = \{1, 2, 4, 6\}$, $B = \{2, 6, 8, 10\}$, and $C = \{3 \leq c \leq 4\}$. Here $B \cup C = \{2, 3 \leq c \leq 4, 6, 8, 10\}$, $A \cap B = \{2, 6\}$, $A \cap C = \{4\}$. For the left side of (1.2-14) $A \cap (B \cup C) = \{2, 4, 6\}$. For the right side of (1.2-14) $A \cap B) \cup (A \cap C) = \{2, 4, 6\}$. This result is the same as for the left side, demonstrating the validity of (1.2-14) for these sets.

†After Augustus De Morgan (1806–1871), an English mathematician.

1.3 PROBABILITY INTRODUCED THROUGH SETS AND RELATIVE FREQUENCY

In this section we define probability in two ways. The first is based on set theory and fundamental axioms; this approach is the more sound of the two. Also, it is perhaps a bit more difficult to interpret in a practical sense than the second, called *relative frequency*, which is based more on common sense and engineering or scientific observations. We begin with the introduction of probability using set concepts.

Basic to our study of probability is the idea of a physical *experiment*. In this section we develop a mathematical model of an experiment. Of course, we are interested only in experiments that are regulated in some probabilistic way. A single performance of the experiment is called a *trial* for which there is an *outcome*.

Experiments and Sample Spaces

Although there exists a precise mathematical procedure for defining an experiment, we shall rely on reason and examples. This simplified approach will ultimately lead us to a valid mathematical model for any real experiment.† To illustrate, one experiment might consist of rolling a single die and observing the number that shows up. There are six such numbers and they form all the possible outcomes in the experiment. If the die is "unbiased" our intuition tells us that each outcome is equally likely to occur and the *likelihood* of any one occurring is $\frac{1}{6}$ (later we call this number the *probability* of the outcome). This experiment is seen to be governed, in part, by two *sets*. One is the set of all possible outcomes, and the other is the set of the likelihoods of the outcomes. Each set has six elements. For the present, we consider only the set of outcomes.

The set of all possible outcomes in any given experiment is called the *sample space* and it is given the symbol S. In effect, the sample space is a universal set for the given experiment. S may be different for different experiments, but all experiments are governed by some sample space. The definition of sample space forms the first of three elements in our mathematical model of experiments. The remaining elements are *events* and *probability*, as discussed below.

Discrete and Continuous Sample Spaces

In the earlier die-tossing experiment, S was a finite set of six elements. Such sample spaces are said to be *discrete* and finite. The sample space can also be

†Most of our early definitions involving probability are rigorously established only through concepts beyond our scope. Although we adopt a simplified development of the theory, our final results are no less valid or useful than if we had used the advanced concepts.

discrete and *infinite* for some experiments. For example, S in the experiment "choose randomly a positive integer" is the countably infinite *set* $\{1, 2, 3, \ldots\}$.

Some experiments have an uncountably infinite sample space. An illustration would be the experiment "obtain a number by spinning the pointer on a wheel of chance numbered from 0 to 12." Here any number s from 0 to 12 can result and $S = \{0 < s \leq 12\}$. Such a sample space is called *continuous*.

Events

In most situations, we are interested in some *characteristic* of the outcomes of our experiment as opposed to the outcomes themselves. In the experiment "draw a card from a deck of 52 cards," we might be more interested in whether we draw a spade as opposed to having any interest in individual cards. To handle such situations we define the concept of an event.

An *event* is defined as a subset of the sample space. Because an event is a set, all the earlier definitions and operations applicable to sets will apply to events. For example, if two events have no common outcomes they are *mutually exclusive*.

In the above card experiment, 13 of the 52 possible outcomes are spades. Since any one of the spade outcomes satisfies the event "draw a spade," this event is a set with 13 elements. We have earlier stated that a set with N elements can have as many as 2^N subsets (events defined on a sample space having N possible outcomes). In the present example, $2^N = 2^{52} \approx 4.5(10^{15})$ events.

As with the sample space, events may be either discrete or continuous. The card event "draw a spade" is a discrete, finite event. An example of a discrete, countably infinite event would be "select an odd integer" in the experiment "randomly select a positive integer." The event has a countably infinite number of elements: $\{1, 3, 5, 7, \ldots\}$. However, events defined on a countably infinite sample space do not *have* to be countably infinite. The event $\{1, 3, 5, 7\}$ is clearly not infinite but applies to the integer selection experiment.

Events defined on continuous sample spaces are usually continuous. In the experiment "choose randomly a number a from 6 to 13," the sample space is $S = \{6 \leq s \leq 13\}$. An event of interest might correspond to the chosen number falling between 7.4 and 7.6; that is, the event (call it A) is $A = \{7.4 < a < 7.6\}$.

Discrete events may also be defined on continuous sample spaces. An example of such an event is $A = \{6.13692\}$ for the sample space $S = \{6 \leq s \leq 13\}$ of the previous paragraph. We comment later on this type of event (following Example 1.3-1).

The above definition of an event as a subset of the sample space forms the second of three elements in our mathematical model of experiments. The third element involves defining probability.

Probability Definition and Axioms

To each event defined on a sample space S, we shall assign a nonnegative number called *probability*. Probability is therefore a function; it is a function of the events defined. We adopt the notation $P(A)$† for "the probability of event A." When an event is stated explicitly as a set by using braces, we employ the notation $P\{\cdot\}$ instead of $P(\{\cdot\})$.

The assigned probabilities are chosen so as to satisfy three *axioms*. Let A be any event defined on a sample space S. Then the first two axioms are

$$\text{axiom 1:} \qquad P(A) \geq 0 \tag{1.3-1a}$$

$$\text{axiom 2:} \qquad P(S) = 1 \tag{1.3-1b}$$

The first only represents our desire to work with nonnegative numbers. The second axiom recognizes that the sample space itself is an event, and, since it is the all-encompassing event, it should have the highest possible probability, which is selected as unity. For this reason, S is known as the *certain event*. Alternatively, the null set $\varnothing$ is an event with no elements; it is known as the *impossible event* and its probability is 0.

The third axiom applies to N events A_n, $n = 1, 2, \ldots, N$, where N may possibly be infinite, defined on a sample space S, and having the property $A_m \cap A_n = \varnothing$ for all $m \neq n$. It is

$$\text{axiom 3:} \qquad P\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} P(A_n) \qquad \text{if} \qquad A_m \cap A_n = \varnothing \tag{1.3-1c}$$

for all $m \neq n = 1, 2, \ldots, N$, with N possibly infinite. The axiom states that the probability of the event equal to the union of any number of mutually exclusive events is equal to the sum of the individual event probabilities.

An example should help give a physical picture of the meaning of the above axioms.

EXAMPLE 1.3-1. Let an experiment consist of obtaining a number x by spinning the pointer on a "fair" wheel of chance that is labeled from 0 to 100 points. The sample space is $S = \{0 < x \leq 100\}$. We reason that probability of the pointer falling between any two numbers $x_2 \geq x_1$ should be $(x_2 - x_1)/100$ since the wheel is fair. As a check on this assignment, we see that the event $A = \{x_1 < x \leq x_2\}$ satisfies axiom 1 for all x_1 and x_2, and axiom 2 when $x_2 = 100$ and $x_1 = 0$.

Now suppose we break the wheel's periphery into N contiguous segments $A_n = \{x_{n-1} < x \leq x_n\}$, $x_n = (n)100/N$, $n = 1, 2, \ldots, N$, with $x_0 = 0$. Then $P(A_n) = 1/N$, and, for any N,

$$P\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} P(A_n) = \sum_{n=1}^{N} \frac{1}{N} = 1 = P(S)$$

from axiom 3.

†Occasionally it will be convenient to use brackets, such as $P[A]$ when A is itself an event such as $C - (B \cap D)$.

Example 1.3-1 allows us to return to our earlier discussion of discrete events defined on continuous sample spaces. If the interval $x_n - x_{n-1}$ is allowed to approach zero ($\rightarrow 0$), the probability $P(A_n) \rightarrow P(x_n)$; that is, $P(A_n)$ becomes the probability of the pointer falling exactly on the point x_n. Since $N \rightarrow \infty$ in this situation, $P(A_n) \rightarrow 0$. Thus, the probability of a discrete event defined on a continuous sample space is 0. This fact is true in general.

A consequence of the above statement is that events can occur even if their probability is 0. Intuitively, any number can be obtained from the wheel of chance, but that precise number may never occur again. The infinite sample space has only one outcome satisfying such a discrete event, so its probability is 0. Such events are *not* the same as the impossible event which has *no* elements and *cannot* occur. The converse situation can also happen where events with probability 1 may *not* occur. An example for the wheel of chance experiment would be the event $A = \{\text{all numbers except the number } x_n\}$. Events with probability 1 (that may not occur) are not the same as the certain event which *must* occur.

Mathematical Model of Experiments

The axioms of probability, introduced above, complete our mathematical model of an experiment. We pause to summarize. Given some real physical experiment having a set of particular outcomes possible, we first defined a *sample space* to mathematically represent the physical outcomes. Second, it was recognized that certain characteristics of the outcomes in the real experiment were of interest, as opposed to the outcomes themselves; *events* were defined to mathematically represent these characteristics. Finally, *probabilities* were assigned to the defined events to mathematically account for the random nature of the experiment.

Thus, a real experiment is defined mathematically by three things: (1) assignment of a sample space; (2) definition of events of interest; and (3) making probability assignments to the events such that the axioms are satisfied. Establishing the correct model for an experiment is probably the single most difficult step in solving probability problems.

EXAMPLE 1.3-2. An experiment consists of observing the sum of the numbers showing up when two dice are thrown. We develop a model for this experiment.

The sample space consists of $6^2 = 36$ points as shown in Figure 1.3-1. Each possible outcome corresponds to a sum having values from 2 to 12.

Suppose we are mainly interested in three events defined by $A = \{\text{sum} = 7\}$, $B = \{8 < \text{sum} \leq 11\}$, and $C = \{10 < \text{sum}\}$. In assigning probabilities to these events, it is first convenient to define 36 *elementary events* $A_{ij} = \{\text{sum for outcome } (i, j) = i + j\}$, where i represents the row and j represents the column locating a particular possible outcome in Figure 1.3-1. An elementary event has only one element.

For probability assignments, intuition indicates that each possible outcome has the same likelihood of occurrence if the dice are fair, so

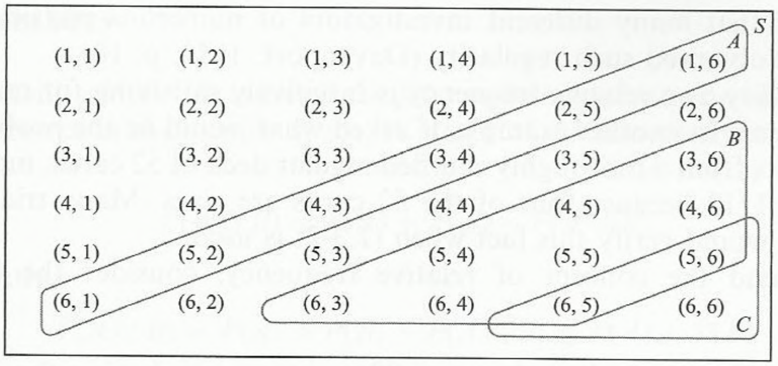


FIGURE 1.3-1
Sample space applicable to Example 1.3-2.

$P(A_{ij}) = \frac{1}{36}$. Now because the events A_{ij}, i and $j = 1, 2, \ldots, N = 6$, are mutually exclusive, they must satisfy axiom 3. But since the events A, B, and C are simply the unions of appropriate elementary events, their probabilities are derived from axiom 3. From Figure 1.3-1 we easily find

$$P(A) = P\left(\bigcup_{i=1}^{6} A_{i,7-i}\right) = \sum_{i=1}^{6} P(A_{i,7-i}) = 6\left(\frac{1}{36}\right) = \frac{1}{6}$$

$$P(B) = 9\left(\frac{1}{36}\right) = \frac{1}{4}$$

$$P(C) = 3\left(\frac{1}{36}\right) = \frac{1}{12}$$

As a matter of interest, we also observe the probabilities of the events $B \cap C$ and $B \cup C$ to be $P(B \cap C) = 2(\frac{1}{36}) = \frac{1}{18}$ and $P(B \cup C) = 10(\frac{1}{36}) = \frac{5}{18}$.

Probability as a Relative Frequency

The use of common sense and engineering and scientific observations leads to a definition of probability as a *relative frequency* of occurrence of some event. For example, everyone can surmise that if a fair coin is flipped several times, the side that shows up will be "heads" about half the time with good "regularity." What reason is saying here is that, if the coin is flipped many times (say n) and heads shows up n_H times out of the n flips, then

$$\lim_{n\to\infty} (n_H/n) = P(H) \qquad (1.3\text{-}2)$$

where $P(H)$ is interpreted as the probability of the event "heads." The ratio n_H/n is the *relative frequency* (or average number of successes) for this event. The idea of *statistical regularity* is used to account for the fact that relative frequencies approach a fixed value (a probability) as n becomes large (Cooper and McGillem, 1986, p. 9). That such regularity is reasonable is based purely on the fact that many different investigators of numerous physical experiments have observed such regularity (Davenport, 1970, p. 10).

Probability as a relative frequency is intuitively satisfying for many practical problems. In another example, if asked what would be the probability of drawing a six from a thoroughly shuffled regular deck of 52 cards, most would say $4/52 = 1/13$ because four of the 52 cards are sixes. Many trials of this experiment would verify this fact when (1.3-2) is used.

To extend the concept of relative frequency, consider the following example.

EXAMPLE 1.3-3. In a box there are 80 resistors each having the same size and shape. Of the 80 resistors 18 are 10 Ω, 12 are 22 Ω, 33 are 27 Ω, and 17 are 47 Ω. If the experiment is to randomly draw out one resistor from the box with each one being "equally likely" to be drawn, then relative frequency suggests that the following probabilities may reasonably be assumed:

$$P(\text{draw } 10\,\Omega) = 18/80 \qquad P(\text{draw } 22\,\Omega) = 12/80$$
$$P(\text{draw } 27\,\Omega) = 33/80 \qquad P(\text{draw } 47\,\Omega) = 17/80 \tag{1}$$

Since all resistors are distinct, mutually exclusive, both individually and by type, have only four types, and some resistor *must* be chosen, the sum of the probabilities of all four events must equal 1.

Next, suppose a 22-Ω resistor is drawn from the box and *not replaced.* A second resistor is then drawn from the box. We ask, what are now the probabilities of drawing a resistor of any one of the four values? Since the "population" in the box is now 79 resistors, we have, for the second drawing,

$$P(\text{draw } 10\,\Omega|22\,\Omega) = 18/79 \qquad P(\text{draw } 22\,\Omega|22\,\Omega) = 11/79$$
$$P(\text{draw } 27\,\Omega|22\,\Omega) = 33/79 \qquad P(\text{draw } 47\,\Omega|22\,\Omega) = 17/79 \tag{2}$$

We use the notation $P(\cdot|22\,\Omega)$ to note that probabilities on the second drawing are now *conditional* on the outcome of the first drawing. This reasoning is readily extended to other drawings in sequence without replacement.

In Example 1.3-3 it was seen that the probability of some event may depend (be conditional) on the occurrence of another event. We consider such probabilities in more detail in the next section.

1.4
JOINT AND CONDITIONAL PROBABILITY

In some experiments, such as in Example 1.3-2 above, it may be that some events are not mutually exclusive because of common elements in the sample space. These elements correspond to the simultaneous or *joint* occurrence of the nonexclusive events. For two events A and B, the common elements form the event $A \cap B$.

Joint Probability

The probability $P(A \cap B)$ is called the *joint probability* for two events A and B which intersect in the sample space. A study of a Venn diagram will readily show that

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) \tag{1.4-1}$$

Equivalently,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B) \tag{1.4-2}$$

In other words, the probability of the union of two events never exceeds the sum of the event probabilities. The equality holds only for mutually exclusive events because $A \cap B = \varnothing$, and therefore, $P(A \cap B) = P(\varnothing) = 0$.

Conditional Probability

Given some event B with nonzero probability

$$P(B) > 0 \tag{1.4-3}$$

we define the *conditional probability* of an event A, given B, by

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1.4-4}$$

The probability $P(A|B)$ simply reflects the fact that the probability of an event A may depend on a second event B. If A and B are mutually exclusive, $A \cap B = \varnothing$, and $P(A|B) = 0$.

Conditional probability is a defined quantity and cannot be proven. However, as a probability it must satisfy the three axioms given in (1.3-1). $P(A|B)$ obviously satisfies axiom 1 by its definition because $P(A \cap B)$ and $P(B)$ are nonnegative numbers. The second axiom is shown to be satisfied by letting $S = A$:

$$P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1 \tag{1.4-5}$$

The third axiom may be shown to hold by considering the union of A with an event C, where A and C are mutually exclusive. If $P(A \cup C|B) = P(A|B) + P(C|B)$ is true, then axiom 3 holds. Since $A \cap C = \varnothing$ then events $A \cap B$ and $B \cap C$ are mutually exclusive (use a Venn diagram to verify this fact) and

$$P[(A \cup C) \cap B] = P[(A \cap B) \cup (C \cap B)] = P(A \cap B) + P(C \cap B) \tag{1.4-6}$$

Thus, on substitution into (1.4-4)

$$\begin{aligned} P[(A \cup C)|B] &= \frac{P[(A \cup C) \cap B]}{P(B)} = \frac{P(A \cap B)}{P(B)} + \frac{P(C \cap B)}{P(B)} \\ &= P(A|B) + P(C|B) \end{aligned} \tag{1.4-7}$$

and axiom 3 holds.

Joint Probability

The probability $P(A \cap B)$ is called the *joint probability* for two events A and B which intersect in the sample space. A study of a Venn diagram will readily show that

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) \tag{1.4-1}$$

Equivalently,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B) \tag{1.4-2}$$

In other words, the probability of the union of two events never exceeds the sum of the event probabilities. The equality holds only for mutually exclusive events because $A \cap B = \varnothing$, and therefore, $P(A \cap B) = P(\varnothing) = 0$.

Conditional Probability

Given some event B with nonzero probability

$$P(B) > 0 \tag{1.4-3}$$

we define the *conditional probability* of an event A, given B, by

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1.4-4}$$

The probability $P(A|B)$ simply reflects the fact that the probability of an event A may depend on a second event B. If A and B are mutually exclusive, $A \cap B = \varnothing$, and $P(A|B) = 0$.

Conditional probability is a defined quantity and cannot be proven. However, as a probability it must satisfy the three axioms given in (1.3-1). $P(A|B)$ obviously satisfies axiom 1 by its definition because $P(A \cap B)$ and $P(B)$ are nonnegative numbers. The second axiom is shown to be satisfied by letting $S = A$:

$$P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1 \tag{1.4-5}$$

The third axiom may be shown to hold by considering the union of A with an event C, where A and C are mutually exclusive. If $P(A \cup C|B) = P(A|B) + P(C|B)$ is true, then axiom 3 holds. Since $A \cap C = \varnothing$ then events $A \cap B$ and $B \cap C$ are mutually exclusive (use a Venn diagram to verify this fact) and

$$P[(A \cup C) \cap B] = P[(A \cap B) \cup (C \cap B)] = P(A \cap B) + P(C \cap B) \tag{1.4-6}$$

Thus, on substitution into (1.4-4)

$$\begin{aligned} P[(A \cup C)|B] &= \frac{P[(A \cup C) \cap B]}{P(B)} = \frac{P(A \cap B)}{P(B)} + \frac{P(C \cap B)}{P(B)} \\ &= P(A|B) + P(C|B) \end{aligned} \tag{1.4-7}$$

and axiom 3 holds.

EXAMPLE 1.4-1. In a box there are 100 resistors having resistance and tolerance as shown in Table 1.4-1. Let a resistor be selected from the box and assume each resistor has the same likelihood of being chosen. Define three events: A as "draw a 47-Ω resistor," B as "draw a resistor with 5% tolerance," and C as "draw a 100-Ω resistor." From the table, the applicable probabilities are†

$$P(A) = P(47\ \Omega) = \frac{44}{100}$$

$$P(B) = P(5\%) = \frac{62}{100}$$

$$P(C) = P(100\ \Omega) = \frac{32}{100}$$

The joint probabilities are

$$P(A \cap B) = P(47\ \Omega \cap 5\%) = \frac{28}{100}$$

$$P(A \cap C) = P(47\ \Omega \cap 100\ \Omega) = 0$$

$$P(B \cap C) = P(5\% \cap 100\ \Omega) = \frac{24}{100}$$

By using (1.4-4) the conditional probabilities become

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{28}{62}$$

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = 0$$

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{24}{32}$$

TABLE 1.4-1
Numbers of resistors in a box having given resistance and tolerance

	Tolerance		
Resistance (Ω)	**5%**	**10%**	**Total**
22	10	14	24
47	28	16	44
100	24	8	32
Total	62	38	100

†It is reasonable that probabilities are related to the *number* of resistors in the box that satisfy an event, since each resistor is equally likely to be selected. An alternative approach would be based on elementary events similar to that used in Example 1.3-2. The reader may view the latter approach as more rigorous but less readily applied.

$P(A|B) = P(47\,\Omega|5\%)$ is the probability of drawing a 47-Ω resistor given that the resistor drawn is 5%. $P(A|C) = P(47\,\Omega|100\,\Omega)$ is the probability of drawing a 47-Ω resistor given that the resistor drawn is 100 Ω; this is clearly an impossible event so the probability of it is 0. Finally, $P(B|C) = P(5\%|100\,\Omega)$ is the probability of drawing a resistor of 5% tolerance given that the resistor is 100 Ω.

Total Probability

The probability $P(A)$ of any event A defined on a sample space S can be expressed in terms of conditional probabilities. Suppose we are given N mutually exclusive events B_n, $n = 1, 2, \ldots, N$, whose union equals S as illustrated in Figure 1.4-1. These events satisfy

$$B_m \cap B_n = \varnothing \qquad m \neq n = 1, 2, \ldots, N \tag{1.4-8}$$

$$\bigcup_{n=1}^{N} B_n = S \tag{1.4-9}$$

We shall prove that

$$P(A) = \sum_{n=1}^{N} P(A|B_n)P(B_n) \tag{1.4-10}$$

which is known as the *total probability* of event A.

Since $A \cap S = A$, we may start the proof using (1.4-9) and (1.2-8):

$$A \cap S = A \cap \left(\bigcup_{n=1}^{N} B_n\right) = \bigcup_{n=1}^{N} (A \cap B_n) \tag{1.4-11}$$

Now the events $A \cap B_n$ are mutually exclusive as seen from the Venn diagram (Fig. 1.4-1). By applying axiom 3 to these events, we have

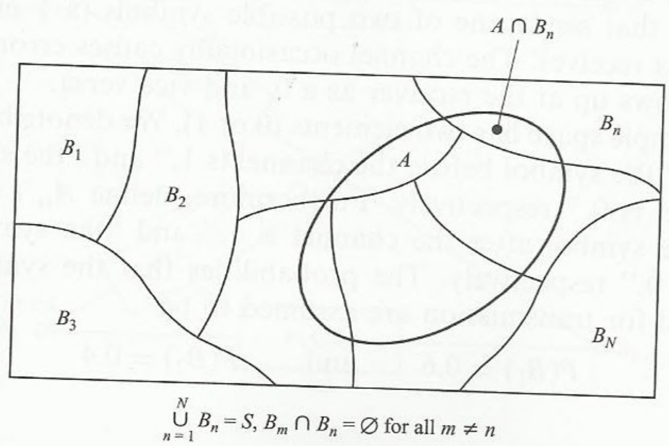


FIGURE 1.4-1
Venn diagram of N mutually exclusive events B_n and another event A.

$$P(A) = P(A \cap S) = P\left[\bigcup_{n=1}^{N}(A \cap B_n)\right] = \sum_{n=1}^{N} P(A \cap B_n) \tag{1.4-12}$$

where (1.4-11) has been used. Finally, (1.4-4) is substituted into (1.4-12) to obtain (1.4-10).

Bayes' Theorem†

The definition of conditional probability, as given by (1.4-4), applies to any two events. In particular, let B_n be one of the events defined above in the subsection on total probability. Equation (1.4-4) can be written

$$P(B_n|A) = \frac{P(B_n \cap A)}{P(A)} \tag{1.4-13}$$

if $P(A) \neq 0$, or, alternatively,

$$P(A|B_n) = \frac{P(A \cap B_n)}{P(B_n)} \tag{1.4-14}$$

if $P(B_n) \neq 0$. One form of Bayes' theorem is obtained by equating these two expressions:

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{P(A)} \tag{1.4-15}$$

Another form derives from a substitution of $P(A)$ as given by (1.4-10),

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{P(A|B_1)P(B_1) + \cdots + P(A|B_N)P(B_N)} \tag{1.4-16}$$

for $n = 1, 2, \ldots, N$.

An example will serve to illustrate Bayes' theorem and conditional probability.

EXAMPLE 1.4-2. An elementary binary communication system consists of a transmitter that sends one of two possible symbols (a 1 or a 0) over a channel to a receiver. The channel occasionally causes errors to occur so that a 1 shows up at the receiver as a 0, and vice versa.

The sample space has two elements (0 or 1). We denote by B_i, $i = 1, 2$, the events "the symbol before the channel is 1," and "the symbol before the channel is 0," respectively. Furthermore, define A_i, $i = 1, 2$, as the events "the symbol after the channel is 1," and "the symbol after the channel is 0," respectively. The probabilities that the symbols 1 and 0 are selected for transmission are assumed to be

$$P(B_1) = 0.6 \qquad \text{and} \qquad P(B_2) = 0.4$$

†The theorem is named for Thomas Bayes (1702–1761), an English theologian and mathematician.

Conditional probabilities describe the effect the channel has on the transmitted symbols. The reception probabilities given a 1 was transmitted are assumed to be

$$P(A_1|B_1) = 0.9$$
$$P(A_2|B_1) = 0.1$$

The channel is presumed to affect 0s in the same manner so

$$P(A_1|B_2) = 0.1$$
$$P(A_2|B_2) = 0.9$$

In either case, $P(A_1|B_i) + P(A_2|B_i) = 1$ because A_1 and A_2 are mutually exclusive and are the only "receiver" events (other than the uninteresting events $\varnothing$ and S) possible. The channel is often shown diagrammatically as illustrated in Figure 1.4-2. Because of its form it is usually called a *binary symmetric channel.*

From (1.4-10) we obtain the "received" symbol probabilities

$$P(A_1) = P(A_1|B_1)P(B_1) + P(A_1|B_2)P(B_2)$$
$$= 0.9(0.6) + 0.1(0.4) = 0.58$$
$$P(A_2) = P(A_2|B_1)P(B_1) + P(A_2|B_2)P(B_2)$$
$$= 0.1(0.6) + 0.9(0.4) = 0.42$$

From either (1.4-15) or (1.4-16) we have

$$P(B_1|A_1) = \frac{P(A_1|B_1)P(B_1)}{P(A_1)} = \frac{0.9(0.6)}{0.58} = \frac{0.54}{0.58} \approx 0.931$$

$$P(B_2|A_2) = \frac{P(A_2|B_2)P(B_2)}{P(A_2)} = \frac{0.9(0.4)}{0.42} = \frac{0.36}{0.42} \approx 0.857$$

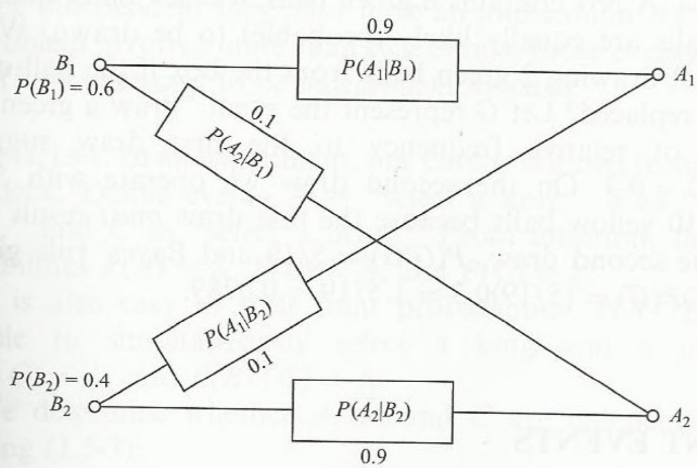


FIGURE 1.4-2
Binary symmetric communication system diagrammatical model applicable to Example 1.4-2.

$$P(B_1|A_2) = \frac{P(A_2|B_1)P(B_1)}{P(A_2)} = \frac{0.1(0.6)}{0.42} = \frac{0.06}{0.42} \approx 0.143$$

$$P(B_2|A_1) = \frac{P(A_1|B_2)P(B_2)}{P(A_1)} = \frac{0.1(0.4)}{0.58} = \frac{0.04}{0.58} \approx 0.069$$

These last two numbers are probabilities of system error while $P(B_1|A_1)$ and $P(B_2|A_2)$ are probabilities of correct system transmission of symbols.

In Bayes' theorem (1.4-16), the probabilities $P(B_n)$ are usually referred to as *a priori probabilities*, since they apply to the events B_n before the performance of the experiment. Similarly, the probabilities $P(A|B_n)$ are numbers typically known prior to conducting the experiment. Example 1.4-2 described such a case. The conditional probabilities are sometimes called *transition probabilities* in a communications context. On the other hand, the probabilities $P(B_n|A)$ are called *a posteriori probabilities*, since they apply after the experiment's performance when some event A is obtained.

EXAMPLE 1.4-3. A student takes a commuter train to get to a school's campus and to class. The probability the student will arrive at class on time is 0.95 provided the train is on time. If the train is known to be on schedule 70% of the time, what is the probability the student will be on time to class? Here we use Bayes' rule and interpret the relative frequency 70% as the probability of the train's being on time. Let C represent the event the student arrives in class and T represent the train arrives on time. The student will arrive on time if the joint event $C \cap T$ is true. The probability of this event is $P(C \cap T) = P(C|T)P(T) = 0.95(0.70) = 0.665$, from (1.4-13).

EXAMPLE 1.4-4. A box contains 6 green balls, 4 black balls, and 10 yellow balls. All balls are equally likely (probable) to be drawn. What is the probability of drawing 2 green balls from the box if the ball on the first draw is not replaced? Let G represent the event "draw a green ball." An application of relative frequency to the first draw suggests that $P(G) = 6/20 = 0.3$. On the second draw we operate with 5 green, 4 black, and 10 yellow balls because the first draw *must* result in a green ball. For the second draw, $P(G|G) = 5/19$ and Bayes' rule gives $P(G \cap G) = P(G|G)P(G) = (5/19)0.3 = 1.5/19 \approx 0.0789$.

1.5 INDEPENDENT EVENTS

In this section we introduce the concept of statistically independent events. Although a given problem may involve any number of events in general, it is most instructive to consider first the simplest possible case of two events.

Let two events A and B have nonzero probabilities of occurrence; that is, assume $P(A) \neq 0$ and $P(B) \neq 0$. We call the events *statistically independent* if the probability of occurrence of one event is not affected by the occurrence of the other event. Mathematically, this statement is equivalent to requiring

$$P(A|B) = P(A) \tag{1.5-1}$$

for statistically independent events. We also have

$$P(B|A) = P(B) \tag{1.5-2}$$

for statistically independent events. By substitution of (1.5-1) into (1.4-4), independence† also means that the probability of the joint occurrence (intersection) of two events must equal the product of the two event probabilities:

$$P(A \cap B) = P(A)P(B) \tag{1.5-3}$$

Not only is (1.5-3) [or (1.5-1)] necessary for two events to be independent but it is sufficient. As a consequence, (1.5-3) can, and often does, serve as a test of independence.

Statistical independence is fundamental to much of our later work. When events are independent, it will often be found that probability problems are greatly simplified.

It has already been stated that the joint probability of two mutually exclusive events is 0:

$$P(A \cap B) = 0 \tag{1.5-4}$$

If the two events have nonzero probabilities of occurrence, then, by comparison of (1.5-4) with (1.5-3), we easily establish that two events cannot be both mutually exclusive and statistically independent. Hence, in order for two events to be independent they *must* have an intersection $A \cap B \neq \varnothing$.

If a problem involves more than two events, those events satisfying either (1.5-3) or (1.5-1) are said to be *independent by pairs.*

EXAMPLE 1.5-1. In an experiment, one card is selected from an ordinary 52-card deck. Define events A as "select a king," B as "select a jack or queen," and C as "select a heart." From intuition, these events have probabilities $P(A) = \frac{4}{52}$, $P(B) = \frac{8}{52}$, and $P(C) = \frac{13}{52}$.

It is also easy to state joint probabilities. $P(A \cap B) = 0$ (it is not possible to simultaneously select a king and a jack or queen), $P(A \cap C) = \frac{1}{52}$, and $P(B \cap C) = \frac{2}{52}$.

We determine whether A, B, and C are independent by pairs by applying (1.5-3):

†We shall often use only the word independence to mean statistical independence.

$$P(A \cap B) = 0 \neq P(A)P(B) = \frac{32}{52^2}$$

$$P(A \cap C) = \frac{1}{52} = P(A)P(C) = \frac{1}{52}$$

$$P(B \cap C) = \frac{2}{52} = P(B)P(C) = \frac{2}{52}$$

Thus, A and C are independent as a pair, as are B and C. However, A and B are not independent, as we might have guessed from the fact that A and B are mutually exclusive.

In many practical problems, statistical independence of events is often *assumed*. The justification hinges on there being no apparent physical connection between the mechanisms leading to the events. In other cases, probabilities assumed for elementary events may lead to independence of other events defined from them (Cooper and McGillem, 1971, p. 24).

Multiple Events

When more than two events are involved, independence by pairs is not sufficient to establish the events as statistically independent, even if *every* pair satisfies (1.5-3).

In the case of three events A_1, A_2, and A_3, they are said to be independent if, and only if, they are independent by all pairs and are also independent as a triple; that is, they must satisfy the *four equations*:

$$P(A_1 \cap A_2) = P(A_1)P(A_2) \tag{1.5-5a}$$

$$P(A_1 \cap A_3) = P(A_1)P(A_3) \tag{1.5-5b}$$

$$P(A_2 \cap A_3) = P(A_2)P(A_3) \tag{1.5-5c}$$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) \tag{1.5-5d}$$

The reader may wonder if satisfaction of (1.5-5*d*) might be sufficient to guarantee independence by pairs, and therefore, satisfaction of all four conditions? The answer is no, and some further detail on this fact can be found in Davenport (1970, p. 83).

More generally, for N events $A_1, A_2, \ldots, A_N$ to be called statistically independent, we require that all the conditions

$$\begin{gathered} P(A_i \cap A_j) = P(A_i)P(A_j) \\ P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k) \\ \vdots \\ P(A_1 \cap A_2 \cap \cdots \cap A_N) = P(A_1)P(A_2) \cdots P(A_N) \end{gathered} \tag{1.5-6}$$

be satisfied for all $1 \le i < j < k < \cdots \le N$. There are $2^N - N - 1$ of these conditions (Davenport, 1970, p. 83).

EXAMPLE 1.5-2. Consider drawing four cards from an ordinary 52-card deck. Let events A_1, A_2, A_3, A_4 define drawing an ace on the first, second, third, and fourth cards, respectively. Consider two cases. First, draw the cards assuming each is replaced after the draw. Intuition tells us that these events are independent so $P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1)P(A_2)P(A_3)P(A_4) = (4/52)^4 \approx 3.50(10^{-5})$.

On the other hand, suppose we keep each card after it is drawn. We now expect these are not independent events. In the general case we may write

$$\begin{aligned}
&P(A_1 \cap A_2 \cap A_3 \cap A_4) \\
&= P(A_1)P(A_2 \cap A_3 \cap A_4 | A_1) \\
&= P(A_1)P(A_2|A_1)P(A_3 \cap A_4 | A_1 \cap A_2) \\
&= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)P(A_4|A_1 \cap A_2 \cap A_3) \\
&= \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \cdot \frac{1}{49} \approx 3.69(10^{-6})
\end{aligned}$$

Thus, we have approximately 9.5 times better chance of drawing four aces when cards are replaced than when kept. This is an intuitively satisfying result since replacing the ace drawn raises chances for an ace on the succeeding draw.

Properties of Independent Events

Many properties of independent events may be summarized by the statement: If N events $A_1, A_2, \ldots, A_N$ are independent, then any one of them is independent of any event formed by unions, intersections, and complements of the others (Papoulis, 1965, p. 42). Several examples of the application of this statement are worth listing for illustration.

For two independent events A_1 and A_2 it results that A_1 is independent of $\bar{A}_2$, $\bar{A}_1$ is independent of A_2, and $\bar{A}_1$ is independent of $\bar{A}_2$. These statements are proved as a problem at the end of this chapter.

For three independent events A_1, A_2, and A_3 any one is independent of the joint occurrence of the other two. For example

$$P[A_1 \cap (A_2 \cap A_3)] = P(A_1)P(A_2)P(A_3) = P(A_1)P(A_2 \cap A_3) \tag{1.5-7}$$

with similar statements possible for the other cases $A_2 \cap (A_1 \cap A_3)$ and $A_3 \cap (A_1 \cap A_2)$. Any one event is also independent of the union of the other two. For example

$$P[A_1 \cap (A_2 \cup A_3)] = P(A_1)P(A_2 \cup A_3) \tag{1.5-8}$$

This result and (1.5-7) do not necessarily hold if the events are only independent by pairs.

1.6 COMBINED EXPERIMENTS

All of our work up to this point is related to outcomes from a single experiment. Many practical problems arise where such a constrained approach does not apply. One example would be the simultaneous measurement of wind speed and barometric presure at some location and instant in time. *Two* experiments are actually being conducted; one has the outcome "speed"; the other outcome is "pressure." Still another type of problem involves conducting the *same* experiment several times, such as flipping a coin N times. In this case there are N performances of the same experiment. To handle these situations we introduce the concept of a combined experiment.

A *combined experiment* consists of forming a *single* experiment by suitably combining individual experiments, which we now call *subexperiments*. Recall that an experiment is defined by specifying three quantities. They are: (1) the applicable sample space, (2) the events defined on the sample space, and (3) the probabilities of the events. We specify these three quantities below, beginning with the sample space, for a combined experiment.

Combined Sample Space

Consider only two subexperiments first. Let S_1 and S_2 be the sample spaces of the two subexperiments and let s_1 and s_2 represent the elements of S_1 and S_2, respectively. We form a new space S, called the *combined sample space*,† whose elements are all the ordered pairs (s_1, s_2). Thus, if S_1 has M elements and S_2 has N elements, then S will have MN elements. The combined sample space is denoted

$$S = S_1 \times S_2 \tag{1.6-1}$$

EXAMPLE 1.6-1. If S_1 corresponds to flipping a coin, then $S_1 = \{H, T\}$, where H is the element "heads" and T represents "tails." Let $S_2 = \{1, 2, 3, 4, 5, 6\}$ corresponding to rolling a single die. The combined sample space $S = S_1 \times S_2$ becomes

$$S = \{(H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6), (T, 1), (T, 2),\\ (T, 3), (T, 4), (T, 5), (T, 6)\}$$

In the new space, elements are considered to be single objects, each object being a pair of items.

†Also called the *cartesian product space* in some texts.

EXAMPLE 1.6-2. We flip a coin twice, each flip being taken as one subexperiment. The applicable sample spaces are now

$$S_1 = \{H, T\}$$
$$S_2 = \{H, T\}$$
$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

In this last example, observe that the element (H, T) is considered different from the element (T, H); this fact emphasizes the elements of S are *ordered* pairs of objects.

The more general situation of N subexperiments is a direct extension of the above concepts. For N sample spaces S_n, $n = 1, 2, \ldots, N$, having elements s_n, the combined sample space S is denoted

$$S = S_1 \times S_2 \times \cdots \times S_N \tag{1.6-2}$$

and it is the set of all ordered N-tuples

$$(s_1, s_2, \ldots, s_N) \tag{1.6-3}$$

Events on the Combined Space

Events may be defined on the combined sample space through their relationship with events defined on the subexperiment sample spaces. Consider two subexperiments with sample spaces S_1 and S_2. Let A be any event defined on S_1 and B be any event defined on S_2, then

$$C = A \times B \tag{1.6-4}$$

is an event defined on S consisting of all pairs (s_1, s_2) such that

$$s_1 \in A \qquad \text{and} \qquad s_2 \in B \tag{1.6-5}$$

Since elements of A correspond to elements of the event $A \times S_2$ defined on S, and elements of B correspond to the event $S_1 \times B$ defined on S, we easily find that

$$A \times B = (A \times S_2) \cap (S_1 \times B) \tag{1.6-6}$$

Thus, the event defined by the subset of S given by $A \times B$ is the intersection of the subsets $A \times S_2$ and $S_1 \times B$. We consider all subsets of S of the form $A \times B$ as events. All intersections and unions of such events are also events (Papoulis, 1965, p. 50).

EXAMPLE 1.6-3. Let $S_1 = \{0 \le x \le 100\}$ and $S_2 = \{0 \le y \le 50\}$. The combined sample space is the set of all pairs of numbers (x, y) with $0 \le x \le 100$ and $0 \le y \le 50$ as illustrated in Figure 1.6-1. For events

$$A = \{x_1 < x < x_2\}$$
$$B = \{y_1 < y < y_2\}$$

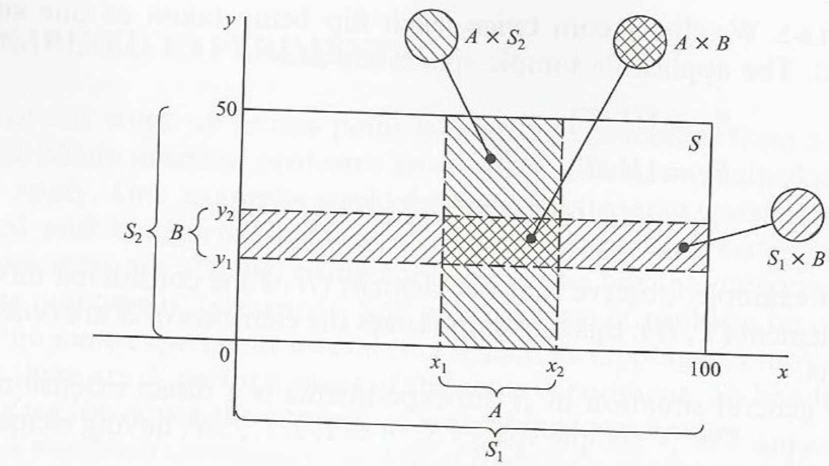


FIGURE 1.6-1
A combined sample space for two subexperiments.

where $0 \le x_1 < x_2 \le 100$ and $0 \le y_1 < y_2 \le 50$, the events $S_1 \times B$ and $A \times S_2$ are horizontal and vertical strips as shown. The event

$$A \times B = \{x_1 < x < x_2\} \times \{y_1 < y < y_2\}$$

is the rectangle shown. An event $S_1 \times \{y = y_1\}$ would be a horizontal line.

In the more general case of N subexperiments with sample spaces S_n on which events A_n are defined, the events on the combined sample space S will all be sets of the form

$$A_1 \times A_2 \times \cdots \times A_N \tag{1.6-7}$$

and unions and intersections of such sets (Papoulis, 1965, pp. 53–54).

Probabilities

To complete the definition of a combined experiment we must assign probabilities to the events defined on the combined sample space S. Consider only two subexperiments first. Since all events defined on S will be unions and intersections of events of the form $A \times B$, where $A \subset S_1$ and $B \subset S_2$, we only need to determine $P(A \times B)$ for any A and B. We shall only consider the case where

$$P(A \times B) = P(A)P(B) \tag{1.6-8}$$

Subexperiments for which (1.6-8) is valid are called *independent experiments*.

To see what elements of S correspond to elements of A and B, we only need substitute S_2 for B or S_1 for A in (1.6-8):

$$P(A \times S_2) = P(A)P(S_2) = P(A) \tag{1.6-9}$$
$$P(S_1 \times B) = P(S_1)P(B) = P(B) \tag{1.6-10}$$

Thus, elements in the set $A \times S_2$ correspond to elements of A, and those of $S_1 \times B$ correspond to those of B.

For N independent experiments, the generalization of (1.6-8) becomes

$$P(A_1 \times A_2 \times \cdots \times A_N) = P(A_1)P(A_2)\cdots P(A_N) \tag{1.6-11}$$

where $A_n \subset S_n$, $n = 1, 2, \ldots, N$.

With independent experiments, the above results show that probabilities for events defined on S are completely determined from probabilities of events defined in the subexperiments.

Permutations

Experiments often involve multiple trials in which outcomes are elements of a finite sample space and they are not replaced after each trial. For example, in drawing four cards from an ordinary 52-card deck, each of the "draws" is not replaced, so the sample spaces for the second, third, and fourth draws have only 51, 50, and 49 elements, respectively. In these and other types of problems, the number of possible sequences of the outcomes is often important.

For n total elements there are n possible outcomes on the first trial, $(n-1)$ on the second, and so forth. For r elements being drawn, the number of possible sequences of r elements from the original n is denoted by P_r^n and is given by

$$\left.\begin{array}{l}\text{Orderings of } r \text{ elements}\\ \text{taken from } n\end{array}\right\} = n(n-1)(n-2)\cdots(n-r+1)$$

$$= \frac{n!}{(n-r)!} = P_r^n, \qquad r = 1, 2, \ldots, n \tag{1.6-12}$$

This number is the number of *permutations*, or sequences, of r elements taken from n elements when order of occurrence is important. This last point is clear from the card experiment. Suppose the first two cards are both kings, say, a heart and a spade. Then the king of hearts followed by the king of spades is considered a different sequence from the spade followed by the heart.

EXAMPLE 1.6-4. How many permutations are there for four cards taken from a 52-card deck? From (1.6-12) $P_4^{52} = 52!/(52-4)! = 52(51)50(49) = 6{,}497{,}400$.

Combinations

If the order of elements in a sequence is *not* important, we reason that there are now fewer possible sequences of r elements taken from n elements without replacement. In fact, the number of permutations of (1.6-12) is reduced by a factor given by the number of permutations (orderings) of the r things, which is $P_r^r = r!$. The resulting number of sequences where order is not important is called the number of *combinations* of r things taken from n things. The notation $\binom{n}{r}$ is usually used to denote combinations, but other notations are also used in various sources. Thus,

$$\binom{n}{r} = \frac{P_r^n}{P_r^r} = \frac{n!}{(n-r)!r!} \tag{1.6-13}$$

The numbers $\binom{n}{r}$ are called *binomial coefficients* because they are central to the expansion of the binomial $(x+y)^n$ as given by

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r} \tag{1.6-14}$$

In computing factorials in (1.6-12) and (1.6-13) we define $0! = 1$, so $\binom{n}{0} = 1$ and $\binom{n}{n} = 1$.

EXAMPLE 1.6-5. A coach has five athletes from whom a 3-person team is to be selected for a competition. How many such teams could he select? The answer is the number of combinations of (1.6-13) with $r = 3$ and $n = 5$, or $5!/(3!2!) = 10$. Note that the same number occurs for a selection of 2-person teams: $5!/(2!3!) = 10$.

The preceding example points out the symmetry of binomial coefficients:

$$\binom{n}{r} = \binom{n}{n-r} \tag{1.6-15}$$

1.7 BERNOULLI TRIALS

We shall close this chapter on probability by considering a very practical problem. It involves any experiment for which there are only two possible outcomes on any trial. Examples of such an experiment are numerous: flipping a coin, hitting or missing the target in artillery, passing or failing an exam, receiving a 0 or a 1 in a computer bit stream, or winning or losing in a game of chance are just a few.

For this type of experiment, we let A be the elementary event having one of the two possible outcomes as its element. $\bar{A}$ is the only other possible elementary event. Specifically, we shall repeat the basic experiment N times and determine the probability that A is observed exactly k times out of the N trials. Such repeated experiments are called *Bernoulli trials*.† Those readers familiar with combined experiments will recognize this experiment as the combination of N identical subexperiments. For readers who omitted the section on combined experiments, we shall develop the problem so that the omission will not impair their understanding of the material.

†After the Swiss mathematician Jacob Bernoulli (1654–1705).

Assume that elementary events are statistically independent for every trial. Let event A occur on any given trial with probability

$$P(A) = p \tag{1.7-1}$$

The event $\bar{A}$ then has probability

$$P(\bar{A}) = 1 - p \tag{1.7-2}$$

After N trials of the basic experiment, one *particular* sequence of outcomes has A occurring k times, followed by $\bar{A}$ occurring $N - k$ times.† Because of assumed statistical independence of trials, the probability of this one sequence is

$$\underbrace{P(A)P(A)\cdots P(A)}_{k\text{ terms}}\,\underbrace{P(\bar{A})P(\bar{A})\cdots P(\bar{A})}_{N-k\text{ terms}} = p^k(1-p)^{N-k} \tag{1.7-3}$$

Now there are clearly other particular sequences that will yield k events A and $N - k$ events $\bar{A}$.‡ The probability of each of these sequences is given by (1.7-3). Since the sum of all such probabilities will be the desired probability of A occurring exactly k times in N trials, we only need find the number of such sequences. Some thought will reveal that this is the number of ways of taking k objects at a time from N objects. The number is known to be the binomial coefficient of (1.6-13).

$$\binom{N}{k} = \frac{N!}{k!(N-k)!} \tag{1.7-4}$$

From the product of (1.7-4) and (1.7-3) we finally obtain

$$P\{A \text{ occurs exactly } k \text{ times}\} = \binom{N}{k} p^k(1-p)^{N-k} \tag{1.7-5}$$

EXAMPLE 1.7-1. A submarine attempts to sink an aircraft carrier. It will be successful only if two or more torpedoes hit the carrier. If the sub fires three torpedoes and the probability of a hit is 0.4 for each torpedo, what is the probability that the carrier will be sunk?

Define the event $A = \{\text{torpedo hits}\}$. Then $P(A) = 0.4$, and $N = 3$. Probabilities are found from (1.7-5):

$$P\{\text{exactly no hits}\} = \binom{3}{0}(0.4)^0(1-0.4)^3 = 0.216$$

$$P\{\text{exactly one hit}\} = \binom{3}{1}(0.4)^1(1-0.4)^2 = 0.432$$

---

†This particular sequence corresponds to one N-dimensional element in the combined sample space S.

‡All such sequences define all the elements of S that satisfy the event $\{A$ occurs exactly k times in N trials$\}$ defined on the combined sample space.

$$P\{\text{exactly two hits}\} = \binom{3}{2}(0.4)^2(1-0.4)^1 = 0.288$$

$$P\{\text{exactly three hits}\} = \binom{3}{3}(0.4)^3(1-0.4)^0 = 0.064$$

The answer we desire is

$$\begin{aligned} P\{\text{carrier sunk}\} &= P\{\text{two or more hits}\} \\ &= P\{\text{exactly two hits}\} + P\{\text{exactly three hits}\} \\ &= 0.352 \end{aligned}$$

EXAMPLE 1.7-2. In a culture used for biological research the growth of unavoidable bacteria occasionally spoils results of an experiment that requires at least three out of four cultures to be unspoiled to obtain a single datum point. Experience has shown that about 6 of every 100 cultures are randomly spoiled by the bacteria. If the experiment requires three simultaneously derived, unspoiled data points for success, we find the probability of success for any given set of 12 cultures (three data points of four cultures each).

We treat individual datum points first as a Bernoulli trial problem with $N = 4$ and $p = P\{\text{good culture}\} = \frac{94}{100} = 0.94$. Here

$$P\{\text{valid datum point}\} = P\{3 \text{ good cultures}\} + P\{4 \text{ good cultures}\}$$

$$= \binom{4}{3}(0.94)^3(1-0.94)^1 + \binom{4}{4}(0.94)^4(1-0.94)^0 \approx 0.98$$

Finally, we treat the required three data points as a Bernoulli trial problem with $N = 3$ and $p = P\{\text{valid datum point}\} = 0.98$. Now

$$\begin{aligned} P\{\text{successful experiment}\} &= P\{3 \text{ valid data points}\} \\ &= \binom{3}{3}(0.98)^3(1-0.98)^0 \approx 0.941 \end{aligned}$$

Thus, the given experiment will be successful about 94.1 percent of the time.

When N, k, and $(N-k)$ are large, the factorials in (1.7-5) are difficult to evaluate, so approximations become useful. One approximation, called *Stirling's formula*, is

$$m! \approx (2\pi m)^{1/2} m^m e^{-m} \qquad m \text{ large} \tag{1.7-6}$$

It is exact for $m \to \infty$ in the sense that the ratio of $m!$ to the right side of (1.7-6) tends to unity. For other values of m its fractional error is on the order of $1/(12m)$, which is quite good (better than 1 percent) even for m as small as 10.

By applying Stirling's formula to the factorials, in (1.7-5), and then approximating some resulting factors by the first two terms in their series expansions, it can be shown (see Davenport, 1970, pp. 276–278) that

$$P\{A \text{ occurs exactly } k \text{ times}\} = \binom{N}{k} p^k (1-p)^{N-k}$$
$$\approx \frac{1}{\sqrt{2\pi Np(1-p)}} \exp\left[-\frac{(k-Np)^2}{2Np(1-p)}\right] \quad (1.7\text{-}7)$$

This equation, called the *De Moivre–Laplace*† *approximation*, holds for N, k, and $(N-k)$ large, k near Np such that its deviations from Np (higher or lower) are small in magnitude relative to both Np and $N(1-p)$. We illustrate these restrictions by example.

EXAMPLE 1.7-3. Suppose a certain machine gun fires rounds (cartridges) for 3 seconds at a rate of 2400 per minute, and the probability of any bullet's hitting a large target is 0.4. We find the probability that exactly 50 of the bullets hit the target.

Here $N = 3(2400/60) = 120$, $k = 50$, $p = 0.4$, $Np = 120(0.4) = 48$, and $N(1-p) = 120(0.6) = 72$. Thus, since N, k, and $(N-k) = 70$ are all large, while k is near Np and the deviation of k from Np, which is $50 - 48 = 2$, is much smaller than both $Np = 48$ and $N(1-p) = 72$, we can use (1.7-7):

$$P\{\text{exactly 50 bullets hit the target}\} = \binom{N}{k} p^k (1-p)^{N-k}$$
$$\approx \frac{1}{\sqrt{2\pi(48)0.6}} \exp\left[-\frac{(50-48)^2}{2(48)0.6}\right] = 0.0693$$

The approximation of (1.7-7) fails to be accurate when N becomes very large while p is very small. For these conditions another approximation is helpful. It is called the *Poisson*‡ *approximation*:

$$\binom{N}{k} p^k (1-p)^{N-k} \approx \frac{(Np)^k e^{-Np}}{k!} \qquad N \text{ large and } p \text{ small} \qquad (1.7\text{-}8)$$

1.8 SUMMARY

This chapter has developed the basics of probability, events, and random experiments by successively building on a basic foundation of set theory. It has also defined probability through the concept of a relative frequency of occurrence of events. Specifically, the topics developed were:

†Abraham De Moivre (1667–1754) was a French-born scientist who lived most of his life in England and contributed to the mathematics of probability. Marquis Pierre Simon De Laplace (1749–1827) was an outstanding French mathematician.

‡After the French mathematician Siméon Denis Poisson (1781–1840).

- Definitions of sets, characteristics of sets, and how they enter into definitions of probability.
- Introduction of probability defined through sets and through the use of the relative frequency concept.
- Development of special kinds of probability, such as applied to events, multiple events (joint probability), and events dependent on each other (conditional probability).
- Introduction of the statistical independence of events.
- Discussions of how to combine several separate random experiments such that they may be taken as a single (combined) experiment. Through Bernoulli trials these topics were applied to various practical problems involving success and failure outcomes of an experiment.

This chapter's topics form a firm basis to proceed to the important concept of a random variable in the next chapter.

PROBLEMS

1.1-1. Specify the following sets by the rule method.

$$A = \{1, 2, 3\}, B = \{8, 10, 12, 14\}, C = \{1, 3, 5, 7, \ldots\}$$

1.1-2. Use the tabular method to specify a class of sets for the sets of Problem 1.1-1.

1.1-3. State whether the following sets are countable or uncountable, or finite or infinite. $A = \{1\}$, $B = \{x = 1\}$, $C = \{0 < \text{integers}\}$, $D = \{\text{children in public school No. 5}\}$, $E = \{\text{girls in public school No. 5}\}$, $F = \{\text{girls in class in public school No. 5 at 3:00 A.M.}\}$, $G = \{\text{all lengths not exceeding one meter}\}$, $H = \{-25 \le x \le -3\}$, $I = \{-2, -1, 1 \le x \le 2\}$.

1.1-4. For each set of Problem 1.1-3, determine if it is equal to, or a subset of, any of the other sets.

1.1-5. State every possible subset of the set of letters $\{a, b, c, d\}$.

1.1-6. A thermometer measures temperatures from −40 to 130°F (−40 to 54.4°C).
(*a*) State a universal set to describe temperature measurements. Specify subsets for:
(*b*) Temperature measurements not exceeding water's freezing point, and
(*c*) Measurements exceeding the freezing point but not exceeding 100°F (37.8°C).

***1.1-7.** Prove that a set with N elements has 2^N subsets.

1.1-8. A random noise voltage at a given time may have any value from −10 to 10 V.
(*a*) What is the universal set describing noise voltage?
(*b*) Find a set to describe the voltages available from a half-wave rectifier for positive voltages that has a linear output-input voltage characteristic.
(*c*) Repeat parts (*a*) and (*b*) if a dc voltage of −3 V is added to the random noise.

1.1-9. Use the tabular method to define a set A that contains all integers with magnitudes not exceeding 7. Define a second set B having odd integers larger than -2 and not larger than 5. Determine if $A \subset B$ and $B \subset A$.

1.1-10. A set A has three elements a_1, a_2, and a_3. Determine all possible subsets of A.

1.1-11. Specify, by both the tabular and rule methods, each of the following sets: (*a*) all integers between 1 and 9, (*b*) all integers from 1 to 9, (*c*) the five values of equivalent resistance for n identical 10-Ω resistors in parallel where $n = 1, 2, \ldots, 5$, and (*d*) the six values of equivalent resistance for n identical 2.2-Ω resistors in series where $n = 1, 2, \ldots, 6$.

1.1-12. A box contains 100 capacitors (universal set) of which 40 are 0.01 μF with a 100-V voltage rating, 35 are 0.1 μF at a rating of 50 V, and 25 are 1.0 μF and have a 10-V rating. Determine the number of elements in the following sets:
(*a*) $A = \{\text{capacitors with capacitance} \geq 0.1\,\mu\text{F}\}$
(*b*) $B = \{\text{capacitors with voltage rating} > 5\,\text{V}\}$
(*c*) $C = \{\text{capacitors with both capacitance} \geq 0.1\,\mu\text{F}$ *and* voltage rating $\geq 50\,\text{V}\}$.

1.2-1. Show that $C \subset A$ if $C \subset B$ and $B \subset A$.

1.2-2. Two sets are given by $A = \{-6, -4, -0.5, 0, 1.6, 8\}$ and $B = \{-0.5, 0, 1, 2, 4\}$. Find:
(*a*) $A - B$ (*b*) $B - A$ (*c*) $A \cup B$ (*d*) $A \cap B$

1.2-3. A universal set is given as $S = \{2, 4, 6, 8, 10, 12\}$. Define two subsets as $A = \{2, 4, 10\}$ and $B = \{4, 6, 8, 10\}$. Determine the following:
(*a*) $\bar{A} = S - A$ (*b*) $A - B$ and $B - A$ (*c*) $A \cup B$ (*d*) $A \cap B$
(*e*) $\bar{A} \cap B$

1.2-4. Using Venn diagrams for three sets A, B, C, shade the areas corresponding to the sets:
(*a*) $(A \cup B) - C$ (*b*) $\bar{B} \cap A$ (*c*) $A \cap B \cap C$ (*d*) $(\overline{A \cup B}) \cap C$

1.2-5. Sketch a Venn diagram for three events where $A \cap B \neq \varnothing$, $B \cap C \neq \varnothing$, $C \cap A \neq \varnothing$, but $A \cap B \cap C = \varnothing$.

1.2-6. Use Venn diagrams to show that the following identities are true:
(*a*) $(\overline{A \cup B}) \cap C = C - [(A \cap C) \cup (B \cap C)]$
(*b*) $(A \cup B \cup C) - (A \cap B \cap C) = (\bar{A} \cap B) \cup (\bar{B} \cap C) \cup (\bar{C} \cap A)$
(*c*) $(\overline{A \cap B \cap C}) = \bar{A} \cup \bar{B} \cup \bar{C}$

1.2-7. Use Venn diagrams to prove De Morgan's laws $(\overline{A \cup B}) = \bar{A} \cap \bar{B}$ and $(\overline{A \cap B}) = \bar{A} \cup \bar{B}$.

1.2-8. A universal set is $S = \{-20 < s \leq -4\}$. If $A = \{-10 \leq s \leq -5\}$ and $B = \{-7 < s < -4\}$, find:
(*a*) $A \cup B$
(*b*) $A \cap B$
(*c*) A third set C such that the sets $A \cap C$ and $B \cap C$ are as large as possible while the smallest element in C is -9.

(*d*) What is the set $A \cap B \cap C$?

1.2-9. Use De Morgan's laws to show that:
(*a*) $\overline{A \cap (B \cup C)} = (\bar{A} \cup \bar{B}) \cap (\bar{A} \cup \bar{C})$
(*b*) $\overline{(A \cap B \cap C)} = \bar{A} \cup \bar{B} \cup \bar{C}$
In each case check your results using a Venn diagram.

1.2-10. Shade Venn diagrams to illustrate each of the following sets:
(*a*) $(A \cup \bar{B}) \cap \bar{C}$, (*b*) $\overline{(A \cap B)} \cup \bar{C}$, (*c*) $(A \cup \bar{B}) \cup (C \cap D)$,
(*d*) $(A \cap B \cap \bar{C}) \cup (\bar{B} \cap C \cap D)$.

1.2-11. A universal set S is composed of all points in a rectangular area defined by $0 \le x \le 3$ and $0 \le y \le 4$. Define three sets by $A = \{y \le 3(x-1)/2\}$, $B = \{y \ge 1\}$, and $C = \{y \ge 3 - x\}$. Shade in Venn diagrams corresponding to the sets (*a*) $A \cap B \cap C$, and (*b*) $C \cap B \cap \bar{A}$.

1.2-12. The take-off roll distance for aircraft at a certain airport can be any number from 80 m to 1750 m. Propeller aircraft require from 80 m to 1050 m while jets use from 950 m to 1750 m. The overall runway is 2000 m.
(*a*) Determine sets A, B, and C defined as "propeller aircraft take-off distances," "jet aircraft take-off distances," and "runway length safety margin," respectively.
(*b*) Determine the set $A \cap B$ and give its physical significance.
(*c*) What is the meaning of the set $\overline{A \cup B}$?
(*d*) What are the meanings of the sets $\overline{A \cup B \cup C}$ and $A \cup B$?

1.2-13. Prove that De Morgan's law (1.2-13) can be extended to N events A_i, $i = 1, 2, \ldots, N$ as follows:

$$\overline{(A_1 \cap A_2 \cap \cdots \cap A_N)} = (\bar{A}_1 \cup \bar{A}_2 \cup \cdots \cup \bar{A}_N)$$

1.2-14. Work Problem 1.2-13 for (1.2-12) to prove

$$\overline{(A_1 \cup A_2 \cup \cdots \cup A_N)} = (\bar{A}_1 \cap \bar{A}_2 \cap \cdots \cap \bar{A}_N)$$

1.2-15. Sets $A = \{1 \le s \le 14\}$, $B = \{3, 6, 14\}$, and $C = \{2 < s \le 9\}$ are defined on a sample space S. State if each of the following conditions is true or false.
(*a*) $C \subset B$, (*b*) $C \subset A$, (*c*) $B \cap C = \varnothing$, (*d*) $C \cup B = S$,
(*e*) $\bar{S} = \varnothing$, (*f*) $A \cap \bar{S} = \varnothing$, and (*g*) $C \subset A \subset B$.

1.2-16. Draw Venn diagrams and shade the areas corresponding to the sets (*a*) $(A \cup B \cup C) \cap (\bar{A} \cup \bar{B} \cup \bar{C})$, and (*b*) $[(A \cup \bar{B}) \cap C] \cup (\overline{A \cup C})$.

1.2-17. Work Problem 1.2-16 except assume sets (*a*) $(A \cap B \cap C) \cup (\overline{A \cup B \cup C})$, (*b*) $B - (A \cap B)$, and (*c*) $(A \cap B) \cup (A \cap C) \cup (B \cap C) - (A \cap B \cap C)$.

1.3-1. A die is tossed. Find the probabilities of the events $A = \{\text{odd number shows up}\}$, $B = \{\text{number larger than 3 shows up}\}$, $A \cup B$, and $A \cap B$.

1.3-2. In a game of dice, a "shooter" can win outright if the sum of the two numbers showing up is either 7 or 11 when two dice are thrown. What is his probability of winning outright?

1.3-3. A pointer is spun on a fair wheel of chance having its periphery labeled from 0 to 100.
(*a*) What is the sample space for this experiment?
(*b*) What is the probability that the pointer will stop between 20 and 35?
(*c*) What is the probability that the wheel will stop on 58?

1.3-4. An experiment has a sample space with 10 equally likely elements $S = \{a_1, a_2, \ldots, a_{10}\}$. Three events are defined as $A = \{a_1, a_5, a_9\}$, $B = \{a_1, a_2, a_6, a_9\}$, and $C = \{a_6, a_9\}$. Find the probabilities of:
(*a*) $A \cup C$
(*b*) $B \cup \bar{C}$
(*c*) $A \cap (B \cup C)$
(*d*) $\overline{A \cup B}$
(*e*) $(A \cup B) \cap C$

1.3-5. Let A be an arbitrary event. Show that $P(\bar{A}) = 1 - P(A)$.

1.3-6. An experiment consists of rolling a single die. Two events are defined as: $A = \{\text{a 6 shows up}\}$ and $B = \{\text{a 2 or a 5 shows up}\}$.
(*a*) Find $P(A)$ and $P(B)$.
(*b*) Define a third event C so that $P(C) = 1 - P(A) - P(B)$.

1.3-7. In a box there are 500 colored balls: 75 black, 150 green, 175 red, 70 white, and 30 blue. What are the probabilities of selecting a ball of each color?

1.3-8. A single card is drawn from a 52-card deck.
(*a*) What is the probability that the card is a jack?
(*b*) What is the probability the card will be a 5 or smaller?
(*c*) What is the probability that the card is a red 10?

1.3-9. A pair of fair dice are thrown in a gambling problem. Person A wins if the sum of numbers showing up is six or less *and* one of the dice shows four. Person B wins if the sum is five or more *and* one of the dice shows a four. Find: (*a*) The probability that A wins, (*b*) the probability of B winning, and (*c*) the probability that both A and B win.

1.3-10. You (person A) and two others (B and C) each toss a fair coin in a two-step gambling game. In step 1 the person whose toss is not a match to either of the other two is "odd man out." Only the remaining two whose coins match go on to step 2 to resolve the ultimate winner.
(*a*) What is the probability you will advance to step 2 after the first toss?
(*b*) What is the probability you will be out after the first toss?
(*c*) What is the probability that no one will be out after the first toss?

1.3-11. A particular electronic device is known to contain only 10-, 22-, and 48-Ω resistors, but these resistors may have 0.25-, 0.5-, or 1-W ratings, depending on how purchases are made to minimize cost. Historically, it is found that the probabilities of the 10-Ω resistors being 0.25, 0.5, or 1 W are 0.08, 0.10, and 0.01, respectively. For the 22-Ω resistors the similar probabilities are 0.20, 0.26, and 0.05. It is also historically found that the probabilities are 0.40, 0.51, and 0.09 that any resistors are 0.25, 0.50, and 1 W, respectively. What are the probabilities that the 48-Ω resistors are (*a*) 0.25, (*b*) 0.50, and (*c*) 1 W?

1.3-12. For the sample space defined in Example 1.3-2 find the probabilities that: (*a*) one die will show a 2 and the other will show a 3 or larger, and (*b*) the sum of the two numbers showing up will be 4 or less or will be 10 or more.

1.3-13. In a game two dice are thrown. Let one die be "weighted" so that a 4 shows up with probability $\frac{2}{7}$, while its other numbers all have probabilities of $\frac{1}{7}$. The same probabilities apply to the other die except the number 3 is "weighted." Determine the probability the shooter will win outright by having the sum of the numbers showing up be 7. What would be the probability for fair dice?

1.4-1. Two cards are drawn from a 52-card deck (the first is not replaced).
(*a*) Given the first card is a queen, what is the probability that the second is also a queen?
(*b*) Repeat part (*a*) for the first card a queen and the second card a 7.
(*c*) What is the probability that both cards will be a queen?

1.4-2. An ordinary 52-card deck is thoroughly shuffled. You are dealt four cards up. What is the probability that all four cards are sevens?

1.4-3. For the resistor selection experiment of Example 1.4-1, define event D as "draw a 22-Ω resistor," and E as "draw a resistor with 10% tolerance." find $P(D)$, $P(E)$, $P(D \cap E)$, $P(D|E)$, and $P(E|D)$.

1.4-4. For the resistor selection experiment of Example 1.4-1, define two mutually exclusive events B_1 and B_2 such that $B_1 \cup B_2 = S$.
(*a*) Use the total probability theorem to find the probability of the event "select a 22-Ω resistor," denoted D.
(*b*) Use Bayes' theorem to find the probability that the resistor selected had 5% tolerance, given it was 22 Ω.

1.4-5. In three boxes there are capacitors as shown in Table P1.4-5. An experiment consists of first randomly selecting a box, assuming each has the same likelihood of selection, and then selecting a capacitor from the chosen box.
(*a*) What is the probability of selecting a 0.01-μF capacitor, given that box 2 is selected?
(*b*) If a 0.01-μF capacitor is selected, what is the probability it came from box 3? (*Hint*: Use Bayes' and total probability theorems.)

TABLE P1.4-5
Capacitors

	Number in box			
Value (μF)	**1**	**2**	**3**	**Totals**
0.01	20	95	25	140
0.1	55	35	75	165
1.0	70	80	145	295
Totals	145	210	245	600

1.4-6. For Problem 1.4-5, list the nine conditional probabilities of capacitor selection, given certain box selections.

1.4-7. Rework Example 1.4-2 if $P(B_1) = 0.6$, $P(B_2) = 0.4$, $P(A_1|B_1) = P(A_2|B_2) = 0.95$, and $P(A_2|B_1) = P(A_1|B_2) = 0.05$.

1.4-8. Rework Example 1.4-2 if $P(B_1) = 0.7$, $P(B_2) = 0.3$, $P(A_1|B_1) = P(A_2|B_2) = 1.0$ and $P(A_2|B_1) = P(A_1|B_2) = 0$. What type of channel does this system have?

1.4-9. A company sells high fidelity amplifiers capable of generating 10, 25, and 50 W of audio power. It has on hand 100 of the 10-W units, of which 15% are defective, 70 of the 25-W units with 10% defective, and 30 of the 50-W units with 10% defective.
(*a*) What is the probability that an amplifier sold from the 10-W units is defective?
(*b*) If each wattage amplifier sells with equal likelihood, what is the probability of a randomly selected unit being 50 W and defective?
(*c*) What is the probability that a unit randomly selected for sale is defective?

1.4-10. A missile can be accidentally launched if two relays A and B both have failed. The probabilities of A and B failing are known to be 0.01 and 0.03, respectively. It is also known that B is more likely to fail (probability 0.06) if A has failed.
(*a*) What is the probability of an accidental missile launch?
(*b*) What is the probability that A will fail if B has failed?
(*c*) Are the events "A fails" and "B fails" statistically independent?

***1.4-11.** The communication system of Example 1.4-2 is to be extended to the case of three transmitted symbols 0, 1, and 2. Define appropriate events A_i and B_i, $i = 1, 2, 3$, to represent symbols after and before the channel, respectively. Assume channel transition probabilities are all equal at $P(A_i|B_j) = 0.1$, $i \neq j$, and are $P(A_i|B_j) = 0.8$ for $i = j = 1, 2, 3$, while symbol transmission probabilities are $P(B_1) = 0.5$, $P(B_2) = 0.3$, and $P(B_3) = 0.2$.
(*a*) Sketch the diagram analogous to Fig. 1.4-2.
(*b*) Compute received symbol probabilities $P(A_1)$, $P(A_2)$, and $P(A_3)$.
(*c*) Compute the a posteriori probabilities for this system.
(*d*) Repeat parts (*b*) and (*c*) for all transmission symbol probabilities equal. Note the effect.

1.4-12. A pharmaceutical product consists of 100 pills in a bottle. Two production lines used to produce the product are selected with probabilities 0.45 (line one) and 0.55 (line two). Each line can overfill or underfill bottles by at most 2 pills. Given that line one is observed, the probabilities are 0.02, 0.06, 0.88, 0.03, and 0.01 that the numbers of pills in a bottle will be 102, 101, 100, 99, and 98, respectively. For line two, the similar respective probabilities are 0.03, 0.08, 0.83, 0.04, and 0.02.
(*a*) Find the probability that a bottle of the product will contain 102 pills. Repeat for 101, 100, 99, and 98 pills.
(*b*) Given that a bottle contains the correct number of pills, what is the probability it came from line one?
(*c*) What is the probability that a purchaser of the product will receive less than 100 pills?

1.4-13. A manufacturing plant makes radios that each contains an integrated circuit (IC) supplied by three sources A, B, and C. The probability that the IC in a radio came from one of the sources is $\frac{1}{3}$, the same for all sources. ICs are known to be defective with probabilities 0.001, 0.003, and 0.002 for sources A, B, and C, respectively.
(*a*) What is the probability any given radio will contain a defective IC?
(*b*) If a radio contains a defective IC, find the probability it came from source A. Repeat for sources B and C.

1.4-14. There are three special decks of cards. The first, deck D_1, has all 52 cards of a regular deck. The second, D_2, has only the 16 face cards of a regular deck (only 4 each of jacks, queens, kings, and aces). The third, D_3, has only the 36 numbered cards of a regular deck (4 twos through 4 tens). A random experiment consists of first randomly choosing one of the three decks, then second, randomly choosing a card from the chosen deck. If $P(D_1) = \frac{1}{2}$, $P(D_2) = \frac{1}{3}$, and $P(D_3) = \frac{1}{6}$, find the probabilities: (*a*) of drawing an ace, (*b*) of drawing a three, and (*c*) of drawing a red card.

1.5-1. Determine whether the three events A, B, and C of Example 1.4-1 are statistically independent.

1.5-2. List the various equations that four events A_1, A_2, A_3, and A_4 must satisfy if they are to be statistically independent.

***1.5-3.** Given that two events A_1 and A_2 are statistically independent, show that:
(*a*) A_1 is independent of $\bar{A}_2$
(*b*) $\bar{A}_1$ is independent of A_2
(*c*) $\bar{A}_1$ is independent of $\bar{A}_2$

1.5-4. Show that there are $2^N - N - 1$ equations required in (1.5-6). (*Hint*: Recall that the binomial coefficient is the number of combinations of N things taken n at a time.)

1.5-5. In a communication system the signal sent from point a to point b arrives by two paths in parallel. Over each path the signal passes through two repeaters (in series). Each repeater in one path has a probability of failing (becoming an open circuit) of 0.005. This probability is 0.008 for each repeater on the other path. All repeaters fail independently of each other. Find the probability that the signal will not arrive at point b.

1.5-6. Work Problem 1.5-5, except assume the paths and repeaters of Figure P1.5-6, where the probabilities of the repeaters' failing (independently) are $p_1 = P(R_1) = 0.005$, $p_2 = P(R_2) = P(R_3) = P(R_4) = 0.01$, and $p_3 = P(R_5) = P(R_6) = 0.05$.

1.6-1. An experiment consists of randomly selecting one of five cities on Florida's west coast for a vacation. Another experiment consists of selecting at random one of four acceptable motels in which to stay. Define sample spaces S_1 and S_2 for the two experiments and a combined space $S = S_1 \times S_2$ for the combined experiment having the two subexperiments.

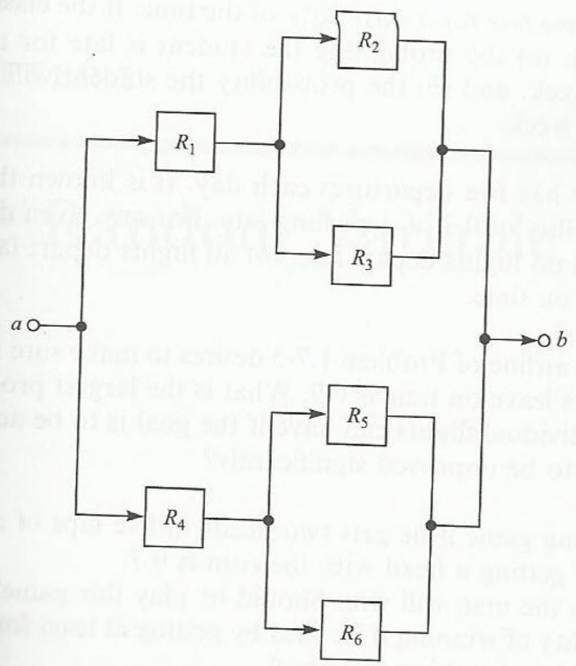


FIGURE P1.5-6

1.6-2. Sketch the area in the combined sample space of Example 1.6-3 corresponding to the event $A \times B$ where:
(*a*) $A = \{10 < x \le 15\}$ and $B = \{20 < y \le 50\}$
(*b*) $A = \{x = 40\}$ and $B = \{5 < y \le 40\}$

1.6-3. The six sides of a fair die are numbered from 1 to 6. The die is rolled four times. How many sequences of the four resulting numbers are possible?

1.6-4. In a 5-card poker game, a player is dealt 5 cards. How many poker hands are possible for an ordinary 52-card deck?

1.7-1. A production line manufactures 5-gal (18.93-liter) gasoline cans to a volume tolerance of 5%. The probability of any one can being out of tolerance is 0.03. If four cans are selected at random:
(*a*) What is the probability they are all out of tolerance?
(*b*) What is the probability of exactly two being out?
(*c*) What is the probability that all are in tolerance?

1.7-2. Spacecraft are expected to land in a prescribed recovery zone 80% of the time. Over a period of time, six spacecraft land.
(*a*) Find the probability that none lands in the prescribed zone.
(*b*) Find the probability that at least one will land in the prescribed zone.
(*c*) The landing program is called successful if the probability is 0.9 or more that three or more out of six spacecraft will land in the prescribed zone. Is the program successful?

1.7-3. In the submarine problem of Example 1.7-1, find the probabilities of sinking the carrier when fewer ($N = 2$) or more ($N = 4$) torpedoes are fired.

1.7-4. A student is known to arrive late for a class 40% of the time. If the class meets five times each week find: (*a*) the probability the student is late for at least three classes in a given week, and (*b*) the probability the student will not be late at all during a given week.

1.7-5. An airline in a small city has five departures each day. It is known that any given flight has a probability of 0.3 of departing late. For any given day find the probabilities that: (*a*) no flights depart late, (*b*) all flights depart late, and (*c*) three or more depart on time.

1.7-6. The local manager of the airline of Problem 1.7-5 desires to make sure that the probability that all flights leave on time is 0.9. What is the largest probability of being late that the individual flights can have if the goal is to be achieved? Will the operation have to be improved significantly?

1.7-7. A man wins in a gambling game if he gets two heads in five flips of a biased coin. The probability of getting a head with the coin is 0.7.
(*a*) Find the probability the man will win. Should he play this game?
(*b*) What is his probability of winning if he wins by getting at least four heads in five flips? Should he play this new game?

***1.7-8.** A rifleman can achieve a "marksman" award if he passes a test. He is allowed to fire six shots at a target's bull's eye. If he hits the bull's eye with at least five of his six shots he wins a set. He becomes a marksman only if he can repeat the feat three times straight, that is, if he can win three straight sets. If his probability is 0.8 of hitting a bull's eye on any one shot, find the probabilities of his: (*a*) winning a set, and (*b*) becoming a marksman.

1.7-9. A ship can successfully arrive at its destination if its engine and its satellite navigation system do not fail en route. If the engine and navigation system are known to fail independently with respective probabilities of 0.05 and 0.001, what is the probability of a successful arrival?

1.7-10. At a certain military installation six similar radars are placed in operation. It is known that a radar's probability of failing to operate before 500 hours of "on" time have accumulated is 0.06. What are the probabilities that before 500 hours have elapsed, (*a*) all will operate, (*b*) all will fail, and (*c*) only one will fail?

1.7-11. A particular model of automobile is recalled to fix a mechanical problem. The probability that a car will be properly repaired is 0.9. During the week a dealer has eight cars to repair.
(*a*) What is the probability that two or more of the eight cars will have to be repaired more than once?
(*b*) What is the probability all eight cars will be properly repaired?

1.7-12. In a large hotel it is known that 99% of all guests return room keys when checking out. If 250 engineers check out after a large conference, what is the probability that not more than three will fail to return their keys? [*Hint*: Use the approximation of (1.7-8).]

CHAPTER 2

The Random Variable

2.0 INTRODUCTION

In the previous chapter we introduced the concept of an event to describe characteristics of outcomes of an experiment. Events allowed us more flexibility in determining properties of an experiment than could be obtained by considering only the outcomes themselves. An event could be almost anything from "descriptive," such as "draw a spade," to numerical, such as "the outcome is 3."

In this chapter, we introduce a new concept that will allow events to be defined in a more consistent manner; they will always be numerical. The new concept is that of a *random variable*, and it will constitute a powerful tool in the solution of practical probabilistic problems.

2.1 THE RANDOM VARIABLE CONCEPT

Definition of a Random Variable

We define a real *random variable*† as a real *function* of the elements of a sample space S. We shall represent a random variable by a capital letter (such as W, X, or Y) and any particular value of the random variable by a lowercase letter (such as w, x, or y). Thus, given an experiment defined by a sample space S with elements s, we assign to every s a real number

†Complex random variables are considered in Chapter 5.

$$X(s) \tag{2.1-1}$$

according to some rule and call $X(s)$ a random variable.

A random variable X can be considered to be a function that maps all elements of the sample space into points on the real line or some parts thereof. We illustrate, by two examples, the mapping of a random variable.

EXAMPLE 2.1-1. An experiment consists of rolling a die and flipping a coin. The applicable sample space is illustrated in Figure 2.1-1. Let the random variable be a function X chosen such that (1) a coin head (H) outcome corresponds to positive values of X that are equal to the numbers that show up on the die, and (2) a coin tail (T) outcome corresponds to negative values of X that are equal in magnitude to *twice* the number that shows on the die. Here X maps the sample space of 12 elements into 12 values of X from -12 to 6 as shown in Figure 2.1-1.

EXAMPLE 2.1-2. Figure 2.1-2 illustrates an experiment where the pointer on a wheel of chance is spun. The possible outcomes are the numbers from 0 to 12 marked on the wheel. The sample space consists of the numbers in the set $\{0 < s \leq 12\}$. We define a random variable by the function

$$X = X(s) = s^2$$

Points in S now map onto the real line as the set $\{0 < x \leq 144\}$.

As seen in these two examples, a random variable is a function that maps each point in S into some point on the real line. It is not necessary that the sample-space points map uniquely, however. More than one point in S may map into a single value of X. For example, in the extreme case, we might map all six points in the sample space for the experiment "throw a die and observe the number that shows up" into the one point $X = 2$.

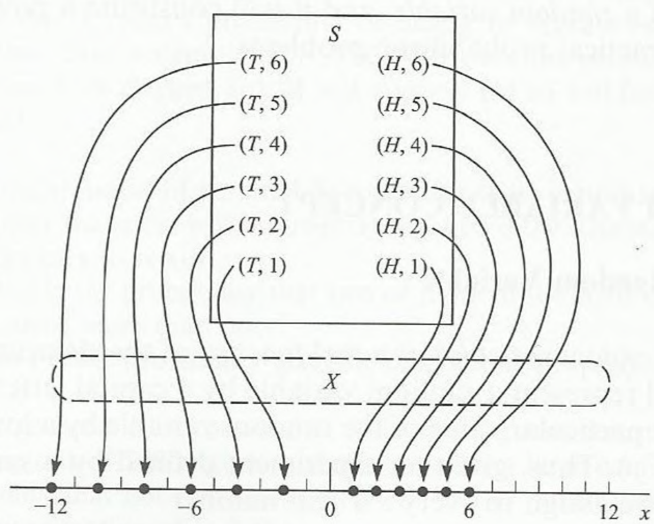


FIGURE 2.1-1
A random variable mapping of a sample space.

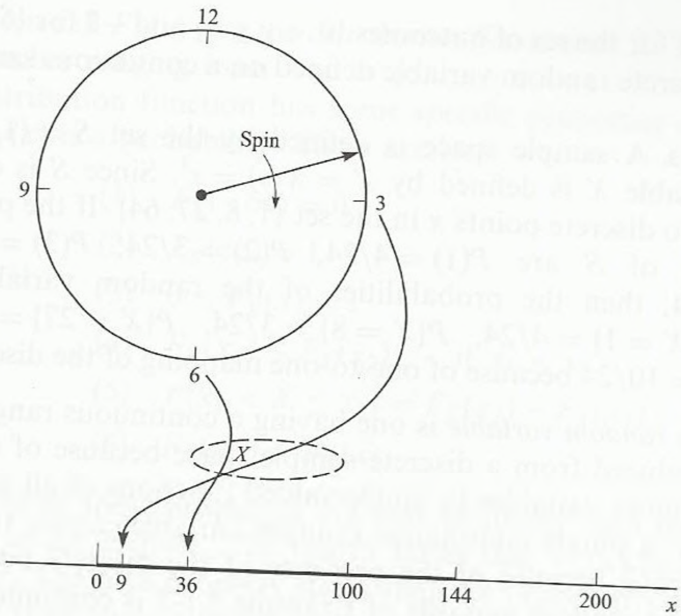


FIGURE 2.1-2
Mapping applicable to Example 2.1-2.

Conditions for a Function to Be a Random Variable

Thus, a random variable may be almost any function we wish. We shall, however, require that it not be multivalued. That is, every point in S must correspond to only one value of the random variable.

Moreover, we shall require that two additional conditions be satisfied in order that a function X be a random variable (Papoulis, 1965, p. 88). First, the set $\{X \leq x\}$ shall be an event for any real number x. The satisfaction of this condition will be no trouble in practical problems. This set corresponds to those points s in the sample space for which the random variable $X(s)$ does not exceed the number x. The probability of this event, denoted by $P\{X \leq x\}$, is equal to the sum of the probabilities of all the elementary events corresponding to $\{X \leq x\}$.

The second condition we require is that the probabilities of the events $\{X = \infty\}$ and $\{X = -\infty\}$ be 0:

$$P\{X = -\infty\} = 0 \qquad P\{X = \infty\} = 0 \tag{2.1-2}$$

This condition does not prevent X from being either $-\infty$ or ∞ for some values of s; it only requires that the probability of the set of those s be zero.

Discrete and Continuous Random Variables

A *discrete random variable* is one having only discrete values. Example 2.1-1 illustrated a discrete random variable. The sample space for a discrete random variable can be discrete, continuous, or even a mixture of discrete and continuous points. For example, the "wheel of chance" of Example 2.1-2 has a continuous sample space, but we could define a discrete random variable as having the value 1 for the set of outcomes $\{0 < s \leq 6\}$ and -1 for $\{6 < s \leq 12\}$. The result is a discrete random variable defined on a continuous sample space.

EXAMPLE 2.1-3. A sample space is defined by the set $S = \{1, 2, 3, 4\}$. A random variable X is defined by $X = X(s) = s^3$. Since S is discrete, its points map to discrete points x in the set $\{1, 8, 27, 64\}$. If the probabilities of elements of S are $P(1) = 4/24$, $P(2) = 3/24$, $P(3) = 7/24$, and $P(4) = 10/24$, then the probabilities of the random variable's values become $P\{X = 1\} = 4/24$, $P\{X = 8\} = 3/24$, $P\{X = 27\} = 7/24$, and $P\{X = 64\} = 10/24$ because of one-to-one mapping of the discrete points.

A *continuous random variable* is one having a continuous range of values. It cannot be produced from a discrete sample space because of our requirement that all random variables be single-valued functions of all sample-space points. Similarly, a purely continuous random variable cannot result from a mixed sample space because of the presence of the discrete portion of the sample space. The random variable of Example 2.1-2 is continuous.

EXAMPLE 2.1-4. Suppose the temperature at some geographical point is modeled as a continuous random variable T that is known to always exist from $-60°F$ to $+120°F$. Further, for ease of illustration, let us make the (nonrealistic) assumption that all values $\{-60 \leq t \leq 120\}$ are equally probable. Under these assumptions, we reason, using relative frequency arguments, that values t of T that fall in a small region dt centered anywhere in the range of $-60°F$ to $+120°F$ will have a probability $dt/[120 - (-60)] = dt/180$. This reasoning is extended to find the probability of any *single temperature* within dt by letting $dt \to 0$. It becomes zero.

Example 2.1-4 serves to demonstrate that the probability of occurrence of any discrete value of a continuous random variable is zero.

Mixed Random Variable

A *mixed random variable* is one for which some of its values are discrete and some are continuous. The mixed case is usually the least important type of random variable, but it occurs in some problems of practical significance.

2.2 DISTRIBUTION FUNCTION

The probability $P\{X \leq x\}$ is the probability of the event $\{X \leq x\}$. It is a number that depends on x; that is, it is a function of x. We call this function, denoted $F_X(x)$, the *cumulative probability distribution function* of the random variable X. Thus,

$$F_X(x) = P\{X \leq x\} \tag{2.2-1}$$

We shall often call $F_X(x)$ just the *distribution function* of X. The argument x is any real number ranging from $-\infty$ to ∞.

The distribution function has some specific properties derived from the fact that $F_X(x)$ is a probability. These are:†

$$(1) \quad F_X(-\infty) = 0 \tag{2.2-2a}$$

$$(2) \quad F_X(\infty) = 1 \tag{2.2-2b}$$

$$(3) \quad 0 \leq F_X(x) \leq 1 \tag{2.2-2c}$$

$$(4) \quad F_X(x_1) \leq F_X(x_2) \qquad \text{if } x_1 < x_2 \tag{2.2-2d}$$

$$(5) \quad P\{x_1 < X \leq x_2\} = F_X(x_2) - F_X(x_1) \tag{2.2-2e}$$

$$(6) \quad F_X(x^+) = F_X(x) \tag{2.2-2f}$$

The first three of these properties are easy to justify, and the reader should justify them as an exercise. The fourth states that $F_X(x)$ is a nondecreasing function of x. The fifth property states that the probability that X will have values larger than some number x_1 but not exceeding another number x_2 is equal to the difference in $F_X(x)$ evaluated at the two points. It is justified from the fact that the events $\{X \leq x_1\}$ and $\{x_1 < X \leq x_2\}$ are mutually exclusive, so the probability of the event $\{X \leq x_2\} = \{X \leq x_1\} \cup \{x_1 < X \leq x_2\}$ is the sum of the probabilities $P\{X \leq x_1\}$ and $P\{x_1 < X \leq x_2\}$. The sixth property states that $F_X(x)$ is a function continuous from the right.

Properties 1, 2, 4, and 6 may be used as tests to determine if some function, say, $G_X(x)$, could be a valid distribution function. If so, all four tests must be passed. [See Papoulis (1965), p. 99.]

If X is a discrete random variable, consideration of its distribution function defined by (2.2-1) shows that $F_X(x)$ must have a stairstep form, such as shown in Figure 2.2-1*a*. The amplitude of a step will equal the probability of occurrence of the value of X where the step occurs. If the values of X are denoted x_i, we may write $F_X(x)$ as

$$F_X(x) = \sum_{i=1}^{N} P\{X = x_i\} u(x - x_i) \tag{2.2-3}$$

where $u(\cdot)$ is the unit-step function defined by‡

$$u(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{2.2-4}$$

and N may be infinite for some random variables. By introducing the shortened notation

$$P(x_i) = P\{X = x_i\} \tag{2.2-5}$$

(2.2-3) can be written as

†We use the notation x^+ to imply $x + \varepsilon$ where $\varepsilon > 0$ is infinitesimally small; that is, $\varepsilon \to 0$.
‡This definition differs slightly from (A-5) by including the equality so that $u(x)$ satisfies (2.2-2*f*).

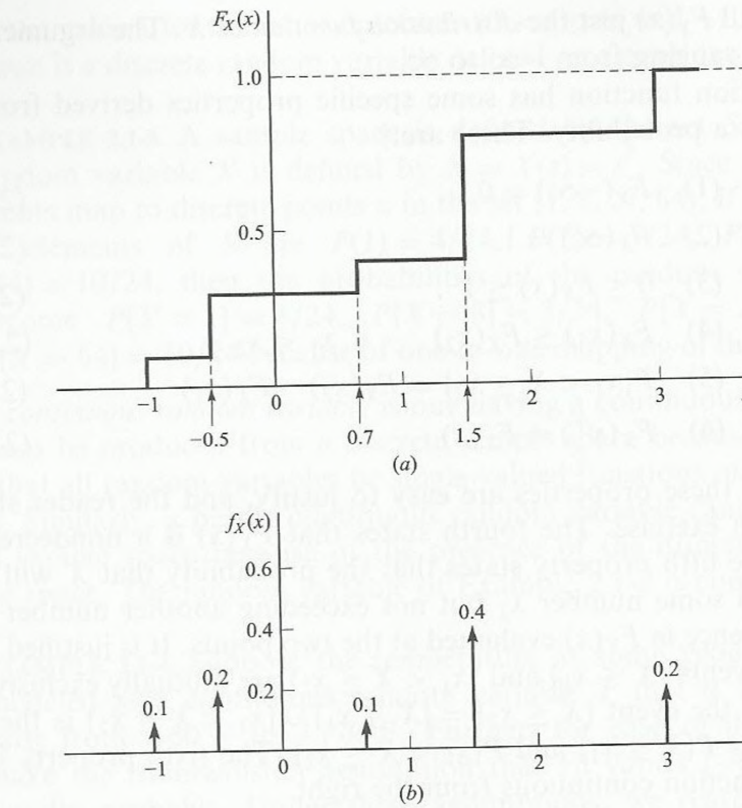


FIGURE 2.2-1
Distribution function (a) and density function (b) applicable to the discrete random variable of Example 2.2-1. [*Adapted from Peebles (1976) with permission of publishers Addison–Wesley, Advanced Book Program.*]

$$F_X(x) = \sum_{i=1}^{N} P(x_i)u(x - x_i) \tag{2.2-6}$$

We next consider an example that illustrates the distribution function of a discrete random variable.

> **EXAMPLE 2.2-1.** Let X have the discrete values in the set $\{-1, -0.5, 0.7, 1.5, 3\}$. The corresponding probabilities are assumed to be $\{0.1, 0.2, 0.1, 0.4, 0.2\}$. Now $P\{X < -1\} = 0$ because there are no sample space points in the set $\{X < -1\}$. Only when $X = -1$ do we obtain one outcome. Thus, there is an immediate jump in probability of 0.1 in the function $F_X(x)$ at the point $x = -1$. For $-1 < x < -0.5$, there are no additional sample space points so $F_X(x)$ remains constant at the value 0.1. At $x = -0.5$ there is another jump of 0.2 in $F_X(x)$. This process continues until all points are included. $F_X(x)$ then equals 1.0 for all x above the last point. Figure 2.2-1a illustrates $F_X(s)$ for this discrete random variable.

A continuous random variable will have a continuous distribution function. We consider an example for which $F_X(x)$ is the continuous function shown in Figure 2.2-2a.

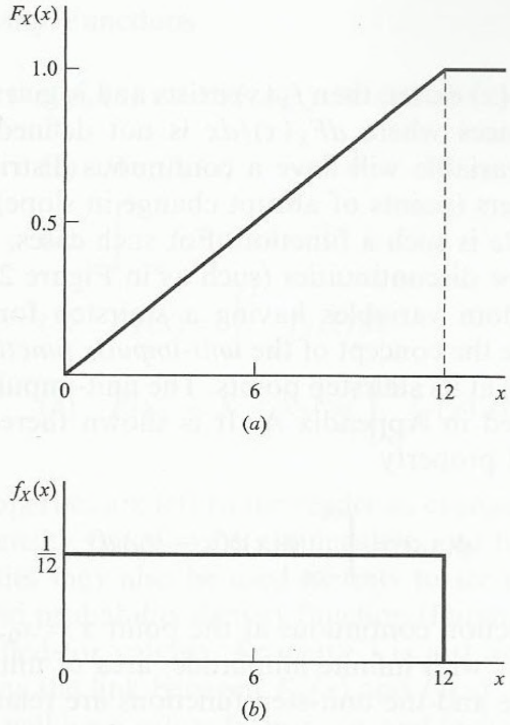


FIGURE 2.2-2
Distribution function (*a*) and density function (*b*) applicable to the continuous random variable of Example 2.2-2. [*Adapted from Peebles (1976) with permission of publishers Addison–Wesley, Advanced Book Program.*]

EXAMPLE 2.2-2. We return to the fair wheel-of-chance experiment. Let the wheel be numbered from 0 to 12 as shown in Figure 2.1-2. Clearly the probability of the event $\{X \leq 0\}$ is 0 because there are no sample space points in this set. For $0 < x \leq 12$ the probability of $\{0 < X \leq x\}$ will increase linearly with x for a fair wheel. Thus, $F_X(x)$ will behave as shown in Figure 2.2-2*a*.

The distribution function of a mixed random variable will be a sum of two parts, one of stairstep form, the other continuous.

2.3 DENSITY FUNCTION

The *probability density function*, denoted by $f_X(x)$, is defined as the derivative of the distribution function:

$$f_X(x) = \frac{dF_X(x)}{dx} \qquad (2.3\text{-}1)$$

We often call $f_X(x)$ just the *density function* of the random variable X.

Existence

If the derivative of $F_X(x)$ exists, then $f_X(x)$ exists and is given by (2.3-1). There may, however, be places where $dF_X(x)/dx$ is not defined. For example, a continuous random variable will have a continuous distribution $F_X(x)$, but $F_X(x)$ may have corners (points of abrupt change in slope). The distribution shown in Figure 2.2-2*a* is such a function. For such cases, we plot $f_X(x)$ as a function with step-type discontinuities (such as in Figure 2.2-2*b*).

For discrete random variables having a stairstep form of distribution function, we introduce the concept of the *unit-impulse function* $\delta(x)$ to describe the derivative of $F_X(x)$ at its stairstep points. The unit-impulse function and its properties are reviewed in Appendix A. It is shown there that $\delta(x)$ may be defined by its integral property

$$\phi(x_0) = \int_{-\infty}^{\infty} \phi(x)\delta(x - x_0)\, dx \tag{2.3-2}$$

where $\phi(x)$ is any function continuous at the point $x = x_0$; $\delta(x)$ can be interpreted as a "function" with infinite amplitude, area of unity, and zero duration. The unit-impulse and the unit-step functions are related by

$$\delta(x) = \frac{du(x)}{dx} \tag{2.3-3}$$

or

$$\int_{-\infty}^{x} \delta(\xi)\, d\xi = u(x) \tag{2.3-4}$$

The more general impulse function is shown symbolically as a vertical arrow occurring at the point $x = x_0$ and having an amplitude equal to the amplitude of the step function for which it is the derivative.

We return to the case of a discrete random variable and differentiate $F_X(x)$, as given by (2.2-6), to obtain

$$f_X(x) = \sum_{i=1}^{N} P(x_i)\delta(x - x_i) \tag{2.3-5}$$

Thus, the density function for a discrete random variable exists in the sense that we use impulse functions to describe the derivative of $F_X(x)$ at its stairstep points. Figure 2.2-1*b* is an example of the density function for the random variable having the function of Figure 2.2-1*a* as its distribution.

A physical interpretation of (2.3-5) is readily achieved. Clearly, the probability of X having one of its particular values, say, x_i, is a number $P(x_i)$. If this probability is assigned to the *point* x_i, then the *density* of probability is infinite because a point has no "width" on the x axis. The infinite "amplitude" of the impulse function describes this infinite density. The "size" of the density of probability at $x = x_i$ is accounted for by the scale factor $P(x_i)$ giving $P(x_i)\delta(x - x_i)$ for the density at the point $x = x_i$.

Properties of Density Functions

Several properties that $f_X(x)$ satisfies may be stated:

$$(1) \quad 0 \le f_X(x) \qquad \text{all } x \tag{2.3-6a}$$

$$(2) \quad \int_{-\infty}^{\infty} f_X(x)\,dx = 1 \tag{2.3-6b}$$

$$(3) \quad F_X(x) = \int_{-\infty}^{x} f_X(\xi)\,d\xi \tag{2.3-6c}$$

$$(4) \quad P\{x_1 < X \le x_2\} = \int_{x_1}^{x_2} f_X(x)\,dx \tag{2.3-6d}$$

Proofs of these properties are left to the reader as exercises. Properties 1 and 2 require that the density function be nonnegative and have an area of unity. These two properties may also be used as tests to see if some function, say, $g_X(x)$, can be a valid probability density function (Papoulis, 1965, p. 99). Both tests must be satisfied for validity. Property 3 is just another way of writing (2.3-1) and serves as the link between $F_X(x)$ and $f_X(x)$. Property 4 relates the probability that X will have values from x_1 to, and including, x_2 to the density function.

EXAMPLE 2.3-1. Let us test the function $g_X(x)$ shown in Figure 2.3-1a to see if it can be a valid density function. It obviously satisfies property 1 since it is nonnegative. Its area is $a\alpha$, which must equal unity to satisfy property 2. Therefore, $a = 1/\alpha$ is necessary if $g_X(x)$ is to be a density.

Suppose $a = 1/\alpha$. To find the applicable distribution function we first write

$$g_X(x) = \begin{cases} 0 & x_0 - \alpha > x \ge x_0 + \alpha \\ \dfrac{1}{\alpha^2}(x - x_0 + \alpha) & x_0 - \alpha \le x < x_0 \\ \dfrac{1}{\alpha} - \dfrac{1}{\alpha^2}(x - x_0) & x_0 \le x < x_0 + \alpha \end{cases}$$

Next, by using (2.3-6c), we obtain

$$G_X(x) = \begin{cases} 0 & x_0 - \alpha > x \\ \displaystyle\int_{x_0-\alpha}^{x} g_X(\xi)\,d\xi = \frac{1}{2\alpha^2}(x - x_0 + \alpha)^2 & x_0 - \alpha \le x < x_0 \\ \displaystyle\frac{1}{2} + \int_{x_0}^{x} g_X(\xi)\,d\xi = \frac{1}{2} + \frac{1}{\alpha}(x - x_0) - \frac{1}{2\alpha^2}(x - x_0)^2 & x_0 \le x < x_0 + \alpha \\ 1 & x_0 + \alpha \le x \end{cases}$$

This function is plotted in Figure 2.3-1b.

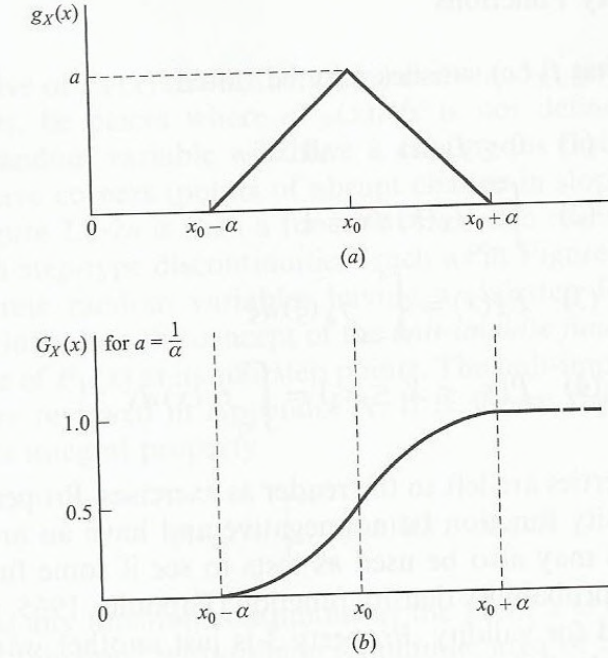


FIGURE 2.3-1
A possible probability density function (*a*) and a distribution function (*b*) applicable to Example 2.3-1.

EXAMPLE 2.3-2. Suppose a random variable is known to have the triangular probability density of the preceding example with $x_0 = 8$, $\alpha = 5$, and $a = 1/\alpha = \frac{1}{5}$. From the earlier work

$$f_X(x) = \begin{cases} 0 & 3 > x \geq 13 \\ (x-3)/25 & 3 \leq x < 8 \\ 0.2 - (x-8)/25 & 8 \leq x < 13 \end{cases}$$

We shall use this probability density in (2.3-6*d*) to find the probability that X has values greater than 4.5 but not greater than 6.7. The probability is

$$P\{4.5 < X \leq 6.7\} = \int_{4.5}^{6.7} [(x-3)/25]\, dx$$

$$= \frac{1}{25}\left[\frac{x^2}{2} - 3x\right]\Bigg|_{4.5}^{6.7} = 0.2288$$

Thus, the event $\{4.5 < X \leq 6.7\}$ has a probability of 0.2288 or 22.88%.

EXAMPLE 2.3-3. A random variable X is known to have a distribution function

$$F_X(x) = u(x)[1 - e^{-x^2/b}]$$

where $b > 0$ is a constant. Find its density function. By use of (2.3-1)

$$f_X(x) = \frac{dF_X(x)}{dx} = u(x)\frac{d}{dx}[1 - e^{-x^2/b}] + [1 - e^{-x^2/b}]\frac{du(x)}{dx}$$
$$= (1 - e^{-x^2/b})\delta(x) + u(x)\frac{2x}{b}e^{-x^2/b} = u(x)\frac{2x}{b}e^{-x^2/b}$$

The impulse term disappears because its coefficient is zero at $x = 0$ where the impulse "exists." [See (A-29).]

2.4 THE GAUSSIAN RANDOM VARIABLE

A random variable X is called *gaussian*† if its density function has the form

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-(x-a_X)^2/2\sigma_X^2} \tag{2.4-1}$$

where $\sigma_X > 0$ and $-\infty < a_X < \infty$ are real constants. This function is sketched in Figure 2.4-1a. Its maximum value $(2\pi\sigma_X^2)^{-1/2}$ occurs at $x = a_X$. Its "spread" about the point $x = a_X$ is related to σ_X. The function decreases to 0.607 times its maximum at $x = a_X + \sigma_X$ and $x = a_X - \sigma_X$. It was first derived by De Moivre some 200 years ago and later independently by both Gauss and Laplace (Kennedy and Neville, 1986, p. 175).

The gaussian density is the most important of all densities and it enters into nearly all areas of science and engineering. This importance stems from its accurate description of many practical and significant real-world quantities, especially when such quantities are the result of many small independent random effects acting to create the quantity of interest. For example, the voltage across a resistor at the output of an amplifier can be random (a noise voltage) due to a random current that is the result of many contributions from other random currents at various places within the amplifier. Random thermal agitation of electrons causes the randomness of the various currents. This type of noise is called *gaussian* because the random variable representing the noise voltage has the gaussian density.

The distribution function is found from (2.3-6c) using (2.4-1). The integral is

$$F_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \int_{-\infty}^{x} e^{-(\xi-a_X)^2/2\sigma_X^2}\, d\xi \tag{2.4-2}$$

This integral has no known closed-form solution and must be evaluated by numerical or approximation methods. To make the results generally available, we could develop a set of tables of $F_X(x)$ for various x with a_X and σ_X as

†After the German mathematician Johann Friedrich Carl Gauss (1777–1855). The gaussian density is often called the *normal density*.

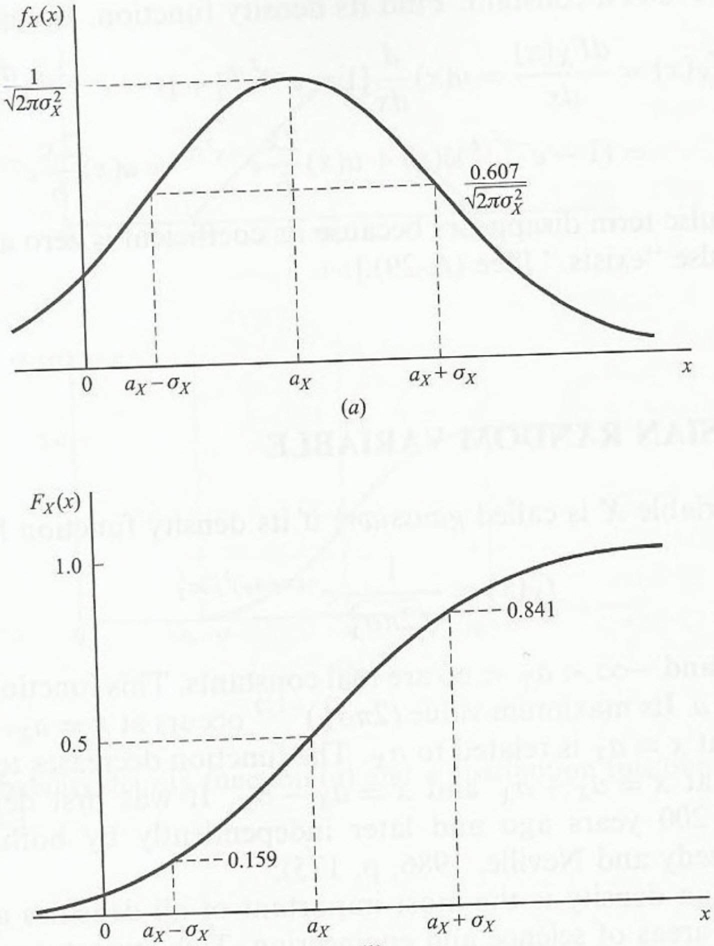


FIGURE 2.4-1
Density (*a*) and distribution (*b*) functions of a gaussian random variable.

parameters. However, this approach has limited value because there is an infinite number of possible combinations of a_X and σ_X, which requires an infinite number of tables. A better approach is possible where only one table of $F_X(x)$ is developed that corresponds to normalized (specific) values of a_X and σ_X. We then show that the one table can be used in the general case where a_X and σ_X can be arbitrary.

We start by first selecting the normalized case where $a_X = 0$ and $\sigma_X = 1$. Denote the corresponding distribution function by $F(x)$. From (2.4-2), $F(x)$ is

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\xi^2/2}\, d\xi \tag{2.4-3}$$

which is a function of x only. This function is tabularized in Appendix B for $x \geq 0$. For a negative value of x we use the relationship

$$F(-x) = 1 - F(x) \tag{2.4-4}$$

To show that the general distribution function $F_X(x)$ of (2.4-2) can be found in terms of $F(x)$ of (2.4-3), we make the variable change

$$u = (\xi - a_X)/\sigma_X \tag{2.4-5}$$

in (2.4-2) to obtain

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-a_X)/\sigma_X} e^{-u^2/2}\, du \tag{2.4-6}$$

From (2.4-3), this expression is clearly equivalent to

$$F_X(x) = F\left(\frac{x - a_X}{\sigma_X}\right) \tag{2.4-7}$$

Figure 2.4-1b depicts the behavior of $F_X(x)$.

We consider two examples to illustrate the application of (2.4-7).

EXAMPLE 2.4-1. We find the probability of the event $\{X \le 5.5\}$ for a gaussian random variable having $a_X = 3$ and $\sigma_X = 2$.

Here $(x - a_X)/\sigma_x = (5.5\text{-}3)/2 = 1.25$. From (2.4-7) and the definition of $F_X(x)$

$$P\{X \le 5.5\} = F_X(5.5) = F(1.25)$$

By using the table in Appendix B

$$P\{X \le 5.5\} = F(1.25) = 0.8944$$

EXAMPLE 2.4-2. Assume that the height of clouds above the ground at some location is a gaussian random variable X with $a_X = 1830$ m and $\sigma_X = 460$ m. We find the probability that clouds will be higher than 2750 m (about 9000 ft). From (2.4-7) and Appendix B:

$$\begin{aligned} P\{X > 2750\} &= 1 - P\{X \le 2750\} = 1 - F_X(2750) \\ &= 1 - F\left(\frac{2750 - 1830}{460}\right) = 1 - F(2.0) \\ &= 1 - 0.9772 = 0.0228 \end{aligned}$$

The probability that clouds are higher than 2750 m is therefore about 2.28 percent if their behavior is as assumed.

The function $F(x)$ can also be evaluated by approximation. First, we write $F(x)$ of (2.4-3) as

$$F(x) = 1 - Q(x) \tag{2.4-8}$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\xi^2/2}\, d\xi \tag{2.4-9}$$

is known as the *Q-function*. As with $F(x)$, $Q(x)$ has no known closed-form solution, but does have an excellent approximation given by

$$Q(x) \approx \left[\frac{1}{(1-a)x + a\sqrt{x^2+b}}\right]\frac{e^{-x^2/2}}{\sqrt{2\pi}} \qquad x \geq 0 \tag{2.4-10}$$

where a and b are constants. This approximation has been found to give minimum absolute relative error, for any $x \geq 0$, when $a = 0.339$ and $b = 5.510$ (see Börjesson and Sundberg, 1979). With these values of a and b, the approximation of (2.4-10) is said to equal the true value of $Q(x)$ within a maximum absolute error of 0.27% of $Q(x)$ for any $x \geq 0$. We consider a simple example.

EXAMPLE 2.4-3. We assume a gaussian random variable for which $a_X = 7$ and $\sigma_X = 0.5$ and find the probability of the event $\{X \leq 7.3\}$. From (2.4-7) and (2.4-8)

$$P\{X \leq 7.3\} = F_X(7.3) = F\left(\frac{7.3-7}{0.5}\right) = F(0.6) = 1 - Q(0.6)$$

$$\approx 1 - \left(\frac{1}{0.661(0.6) + 0.339\sqrt{(0.6)^2 + 5.51}}\right)\frac{e^{-(0.6)^2/2}}{\sqrt{2\pi}}$$

$$\approx 0.7264$$

From Table B-1 the answer is $F(0.6) = 0.7257$ so an absolute error of about $|0.7264 - 0.7257|/0.7257 = 0.00096$ (or 0.096%) exists.

2.5 OTHER DISTRIBUTION AND DENSITY EXAMPLES

Many distribution functions are important enough to have been given names. We give five examples. The first two are for discrete random variables; the remaining three are for continuous random variables. Other distributions are listed in Appendix F.

Binomial

Let $0 < p < 1$, and $N = 1, 2, \ldots$, then the function

$$f_X(x) = \sum_{k=0}^{N}\binom{N}{k}p^k(1-p)^{N-k}\delta(x-k) \tag{2.5-1}$$

is called the *binomial density function*. The quantity $\binom{N}{k}$ is the binomial coefficient defined in (1.7-4) as

$$\binom{N}{k} = \frac{N!}{k!(N-k)!} \tag{2.5-2}$$

The binomial density can be applied to the Bernoulli trial experiment of Chapter 1. It applies to many games of chance, detection problems in radar and sonar, and many experiments having only two possible outcomes on any given trial.

By integration of (2.5-1), the *binomial distribution function* is found:

$$F_X(x) = \sum_{k=0}^{N} \binom{N}{k} p^k (1-p)^{N-k} u(x-k) \tag{2.5-3}$$

Figure 2.5-1 illustrates the binomial density and distribution functions for $N = 6$ and $p = 0.25$.

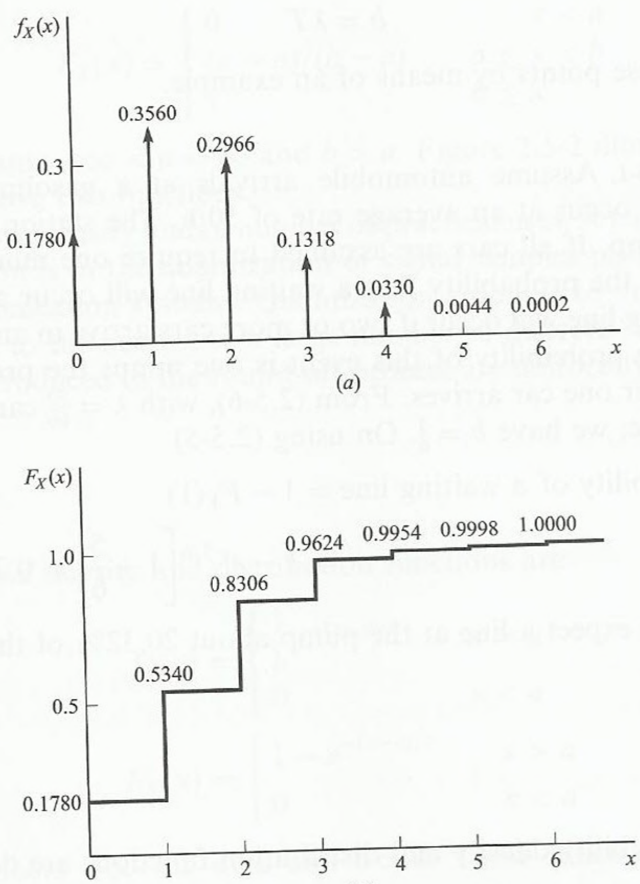


FIGURE 2.5-1
Binomial density (a) and distribution (b) functions for the case $N = 6$ and $p = 0.25$.

Poisson

The *Poisson* random variable X has a density and distribution given by

$$f_X(x) = e^{-b} \sum_{k=0}^{\infty} \frac{b^k}{k!} \delta(x-k) \tag{2.5-4}$$

$$F_X(x) = e^{-b} \sum_{k=0}^{\infty} \frac{b^k}{k!} u(x-k) \tag{2.5-5}$$

where $b > 0$ is a real constant. When plotted, these functions appear quite similar to those for the binomial random variable (Figure 2.5-1). In fact, if $N \to \infty$ and $p \to 0$ for the binomial case in such a way that $Np = b$, a constant, the Poisson case results.

The Poisson random variable applies to a wide variety of counting-type applications. It describes the number of defective units in a sample taken from a production line, the number of telephone calls made during a period of time, the number of electrons emitted from a small section of a cathode in a given time interval, etc. If the time interval of interest has duration T, and the events being counted are known to occur at an average rate λ and have a Poisson distribution, then b in (2.5-4) is given by

$$b = \lambda T \tag{2.5-6}$$

We illustrate these points by means of an example.

EXAMPLE 2.5-1. Assume automobile arrivals at a gasoline station are Poisson and occur at an average rate of 50/h. The station has only one gasoline pump. If all cars are assumed to require one minute to obtain fuel, what is the probability that a waiting line will occur at the pump?

A waiting line will occur if two or more cars arrive in any one-minute interval. The probability of this event is one minus the probability that either none or one car arrives. From (2.5-6), with $\lambda = \frac{50}{60}$ cars/minute and $T = 1$ minute, we have $b = \frac{5}{6}$. On using (2.5-5)

$$\begin{aligned}\text{Probability of a waiting line} &= 1 - F_X(1)\\ &= 1 - e^{-5/6}\left[1 + \frac{5}{6}\right] = 0.2032\end{aligned}$$

We therefore expect a line at the pump about 20.32% of the time.

Uniform

The *uniform* probability density and distribution functions are defined by:

$$f_X(x) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{elsewhere} \end{cases} \tag{2.5-7}$$

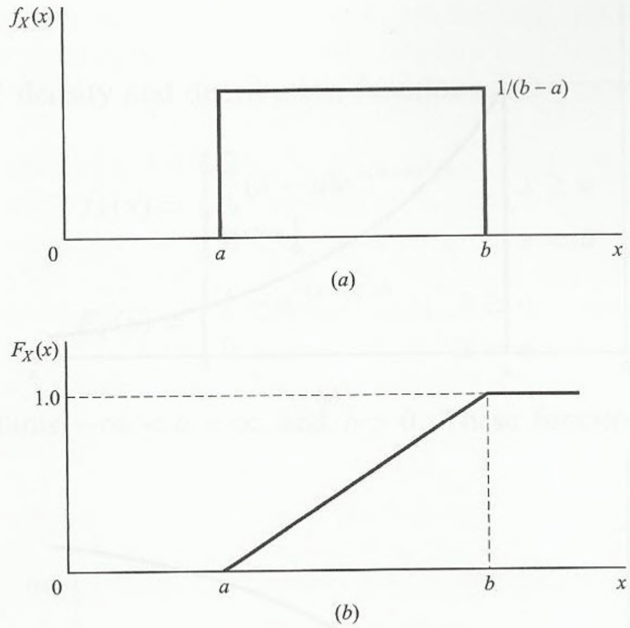


FIGURE 2.5-2
Uniform probability density function (*a*) and its distribution function (*b*).

$$F_X(x) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \leq x < b \\ 1 & b \leq x \end{cases} \tag{2.5-8}$$

for real constants $-\infty < a < \infty$ and $b > a$. Figure 2.5-2 illustrates the behavior of the above two functions.

The uniform density finds a number of practical uses. A particularly important application is in the quantization of signal samples prior to encoding in digital communication systems. Quantization amounts to "rounding off" the actual sample to the nearest of a large number of discrete "quantum levels." The errors introduced in the round-off process are uniformly distributed.

Exponential

The *exponential* density and distribution functions are:

$$f_X(x) = \begin{cases} \frac{1}{b} e^{-(x-a)/b} & x > a \\ 0 & x < a \end{cases} \tag{2.5-9}$$

$$F_X(x) = \begin{cases} 1 - e^{-(x-a)/b} & x > a \\ 0 & x < a \end{cases} \tag{2.5-10}$$

for real numbers $-\infty < a < \infty$ and $b > 0$. These functions are plotted in Figure 2.5-3.

The exponential density is useful in describing raindrop sizes when a large number of rainstorm measurements are made. It is also known to approxi-

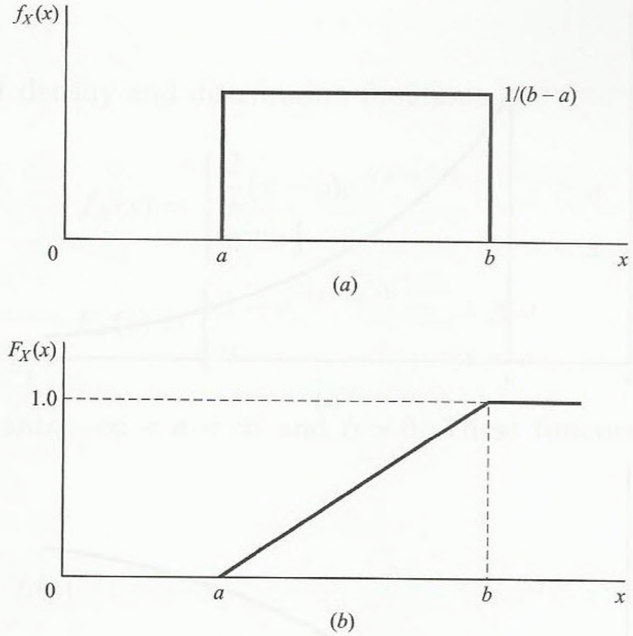


FIGURE 2.5-2
Uniform probability density function (a) and its distribution function (b).

$$F_X(x) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \le x < b \\ 1 & b \le x \end{cases} \tag{2.5-8}$$

for real constants $-\infty < a < \infty$ and $b > a$. Figure 2.5-2 illustrates the behavior of the above two functions.

The uniform density finds a number of practical uses. A particularly important application is in the quantization of signal samples prior to encoding in digital communication systems. Quantization amounts to "rounding off" the actual sample to the nearest of a large number of discrete "quantum levels." The errors introduced in the round-off process are uniformly distributed.

Exponential

The *exponential* density and distribution functions are:

$$f_X(x) = \begin{cases} \frac{1}{b} e^{-(x-a)/b} & x > a \\ 0 & x < a \end{cases} \tag{2.5-9}$$

$$F_X(x) = \begin{cases} 1 - e^{-(x-a)/b} & x > a \\ 0 & x < a \end{cases} \tag{2.5-10}$$

for real numbers $-\infty < a < \infty$ and $b > 0$. These functions are plotted in Figure 2.5-3.

The exponential density is useful in describing raindrop sizes when a large number of rainstorm measurements are made. It is also known to approxi-

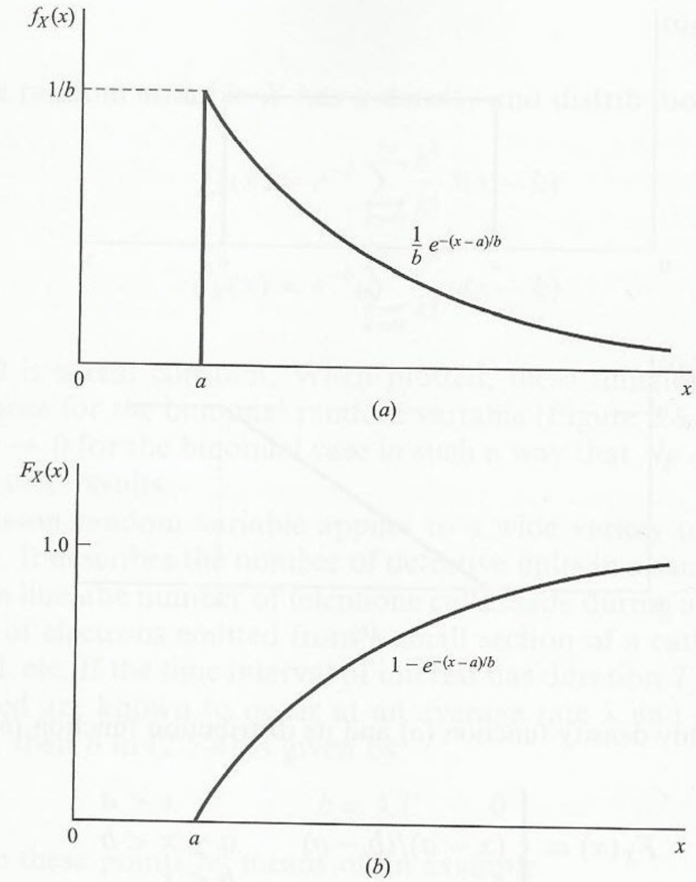


FIGURE 2.5-3
Exponential density (*a*) and distribution (*b*) functions.

mately describe the fluctuations in signal strength received by radar from certain types of aircraft as illustrated by the following example.

EXAMPLE 2.5-2. The power reflected from an aircraft of complicated shape that is received by a radar can be described by an exponential random variable P. The density of P is therefore

$$f_P(p) = \begin{cases} \dfrac{1}{P_0} e^{-p/P_0} & p > 0 \\ 0 & p < 0 \end{cases}$$

where P_0 is the average amount of received power. At some given time P may have a value different from its average value and we ask: what is the probability that the received power is larger than the power received on the average?

We must find $P\{P > P_0\} = 1 - P\{P \le P_0\} = 1 - F_P(P_0)$. From (2.5-10)

$$P\{P > P_0\} = 1 - (1 - e^{-P_0/P_0}) = e^{-1} \approx 0.368$$

In other words, the received power is larger than its average value about 36.8 percent of the time.

Rayleigh

The *Rayleigh*† density and distribution functions are:

$$f_X(x) = \begin{cases} \dfrac{2}{b}(x-a)e^{-(x-a)^2/b} & x \geq a \\ 0 & x < a \end{cases} \tag{2.5-11}$$

$$F_X(x) = \begin{cases} 1 - e^{-(x-a)^2/b} & x \geq a \\ 0 & x < a \end{cases} \tag{2.5-12}$$

for real constants $-\infty < a < \infty$ and $b > 0$. These functions are plotted in Figure 2.5-4.

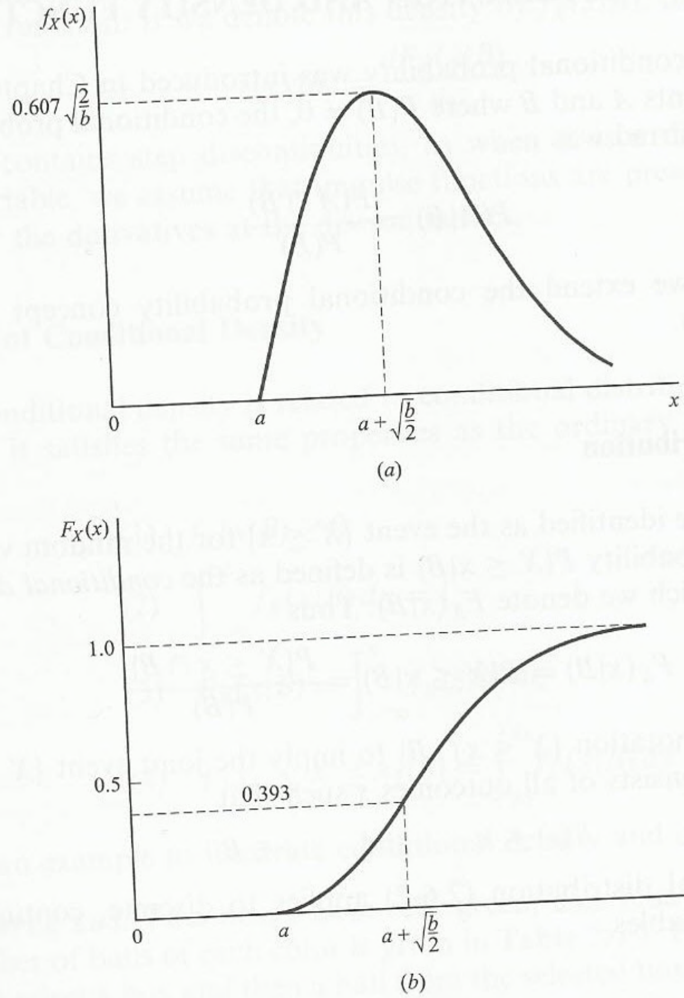


FIGURE 2.5-4
Rayleigh density (*a*) and distribution (*b*) functions.

†Named for the English physicist John William Strutt, Lord Rayleigh (1842–1919).

The Rayleigh density describes the envelope of one type of noise when passed through a bandpass filter. It also is important in analysis of errors in various measurement systems.

EXAMPLE 2.5-3. We find the value $x = x_0$ of a Rayleigh random variable for which $P\{X \leq x_0\} = P\{x_0 < X\}$. This value of X is called the *median* of the random variable. The probability condition requires $P\{X \leq x_0\} = F_X(x_0) = 0.5$. From (2.5-12) $F_X(x_0) = 1 - \exp[-(x_0 - a)^2/b] = 0.5$. The solution for x_0 follows the natural logarithm. We find $x_0 = a + [b\ln(2)]^{1/2}$. The median is similarly defined for random variables other than Rayleigh; it is the value of X for which the probability is 0.5 that values of X do not exceed the median.

2.6 CONDITIONAL DISTRIBUTION AND DENSITY FUNCTIONS

The concept of conditional probability was introduced in Chapter 1. Recall that, for two events A and B where $P(B) \neq 0$, the conditional probability of A given B had occurred was

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2.6-1}$$

In this section we extend the conditional probability concept to include random variables.

Conditional Distribution

Let A in (2.6-1) be identified as the event $\{X \leq x\}$ for the random variable X. The resulting probability $P\{X \leq x|B\}$ is defined as the *conditional distribution function* of X, which we denote $F_X(x|B)$. Thus

$$F_X(x|B) = P\{X \leq x|B\} = \frac{P\{X \leq x \cap B\}}{P(B)} \tag{2.6-2}$$

where we use the notation $\{X \leq x \cap B\}$ to imply the joint event $\{X \leq x\} \cap B$. This joint event consists of all outcomes s such that

$$X(s) \leq x \qquad \text{and} \qquad s \in B \tag{2.6-3}$$

The conditional distribution (2.6-2) applies to discrete, continuous, or mixed random variables.

Properties of Conditional Distribution

All the properties of ordinary distributions apply to $F_X(x|B)$. In other words, it has the following characteristics:

(1) $F_X(-\infty|B) = 0$ (2.6-4a)

(2) $F_X(\infty|B) = 1$ (2.6-4b)

(3) $0 \le F_X(x|B) \le 1$ (2.6-4c)

(4) $F_X(x_1|B) \le F_X(x_2|B) \quad \text{if} \quad x_1 < x_2$ (2.6-4d)

(5) $P\{x_1 < X \le x_2|B\} = F_X(x_2|B) - F_X(x_1|B)$ (2.6-4e)

(6) $F_X\{x^+|B) = F_X(x|B)$ (2.6-4f)

These characteristics have the same general meanings as described earlier following (2.2-2).

Conditional Density

In a manner similar to the ordinary density function, we define *conditional density function* of the random variable X as the derivative of the conditional distribution function. If we denote this density by $f_X(x|B)$, then

$$f_X(x|B) = \frac{dF_X(x|B)}{dx} \tag{2.6-5}$$

If $F_X(x|B)$ contains step discontinuities, as when X is a discrete or mixed random variable, we assume that impulse functions are present in $f_X(x|B)$ to account for the derivatives at the discontinuities.

Properties of Conditional Density

Because conditional density is related to conditional distribution through the derivative, it satisfies the same properties as the ordinary density function. They are:

(1) $f_X(x|B) \ge 0$ (2.6-6a)

(2) $\int_{-\infty}^{\infty} f_X(x|B)\,dx = 1$ (2.6-6b)

(3) $F_X(x|B) = \int_{-\infty}^{x} f_X(\xi|B)\,d\xi$ (2.6-6c)

(4) $P\{x_1 < X \le x_2|B\} = \int_{x_1}^{x_2} f_X(x|B)\,dx$ (2.6-6d)

We take an example to illustrate conditional density and distribution.

EXAMPLE 2.6-1. Two boxes have red, green, and blue balls in them; the number of balls of each color is given in Table 2.6-1. Our experiment will be to select a box and then a ball from the selected box. One box (number 2) is slightly larger than the other, causing it to be selected more frequently. Let B_2 be the event "select the larger box" while B_1 is the event "select the smaller box." Assume $P(B_1) = \frac{2}{10}$ and $P(B_2) = \frac{8}{10}$. (B_1 and B_2 are mutually exclusive and $B_1 \cup B_2$ is the certain event, since some box must be selected; therefore, $P(B_1) + P(B_2)$ must equal unity.)

x_i	Ball color	1	2	Totals
1	Red	5	80	85
2	Green	35	60	95
3	Blue	60	10	70
Totals		100	150	250

Now define a discrete random variable X to have values $x_1 = 1$, $x_2 = 2$, and $x_3 = 3$ when a red, green, or blue ball is selected, and let B be an event equal to either B_1 or B_2. From Table 2.6-1:

$$P(X = 1|B = B_1) = \frac{5}{100} \qquad P(X = 1|B = B_2) = \frac{80}{150}$$

$$P(X = 2|B = B_1) = \frac{35}{100} \qquad P(X = 2|B = B_2) = \frac{60}{150}$$

$$P(X = 3|B = B_1) = \frac{60}{100} \qquad P(X = 3|B = B_2) = \frac{10}{150}$$

The conditional probability density $f_X(x|B_1)$ becomes

$$f_X(x|B_1) = \frac{5}{100}\delta(x-1) + \frac{35}{100}\delta(x-2) + \frac{60}{100}\delta(x-3)$$

By direct integration of $f_X(x|B_1)$:

$$F_X(x|B_1) = \frac{5}{100}u(x-1) + \frac{35}{100}u(x-2) + \frac{60}{100}u(x-3)$$

For comparison, we may find the density and distribution of X by determining the probabilities $P(X = 1)$, $P(X = 2)$, and $P(X = 3)$. These are found from the total probability theorem embodied in (1.4-10):

$$P(X = 1) = P(X = 1|B_1)P(B_1) + P(X = 1|B_2)P(B_2)$$

$$= \frac{5}{100}\left(\frac{2}{10}\right) + \frac{80}{150}\left(\frac{8}{10}\right) = 0.437$$

$$P(X = 2) = \frac{35}{100}\left(\frac{2}{10}\right) + \frac{60}{150}\left(\frac{8}{10}\right) = 0.390$$

$$P(X = 3) = \frac{60}{100}\left(\frac{2}{10}\right) + \frac{10}{150}\left(\frac{8}{10}\right) = 0.173$$

Thus

$$f_X(x) = 0.437\,\delta(x-1) + 0.390\,\delta(x-2) + 0.173\,\delta(x-3)$$

and

$$F_X(x) = 0.437u(x-1) + 0.390\,u(x-2) + 0.173u(x-3)$$

These distributions and densities are plotted in Figure 2.6-1.

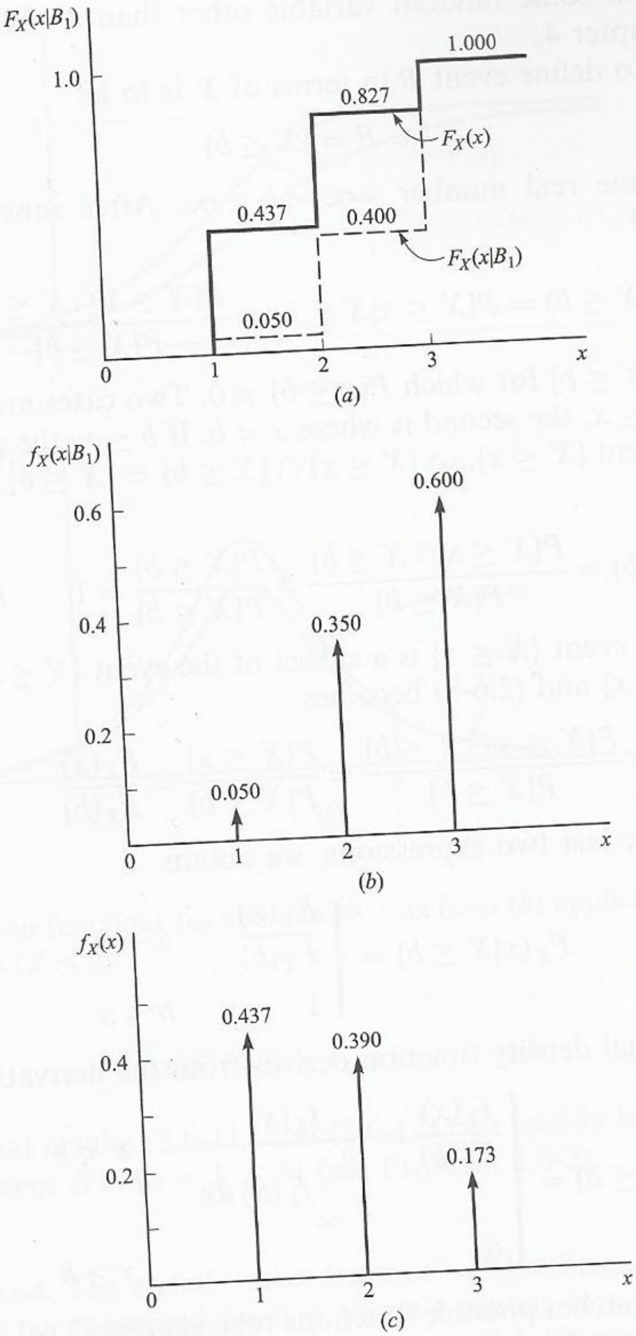


FIGURE 2.6-1
Distributions (*a*) and densities (*b*) and (*c*) applicable to Example 2.6-1.

*Methods of Defining Conditioning Event

The preceding example illustrates how the conditioning event B can be defined from some characteristic of the physical experiment. There are several other ways of defining B (Cooper and McGillem, 1971, p. 61). We shall consider two of these in detail.

In one method, event B is defined in terms of the random variable X. We discuss this case further in the next paragraph. In another method, event B

may depend on some random variable other than X. We discuss this case further in Chapter 4.

One way to define event B in terms of X is to let

$$B = \{X \leq b\} \tag{2.6-7}$$

where b is some real number $-\infty < b < \infty$. After substituting (2.6-7) in (2.6-2), we get†

$$F_X(x|X \leq b) = P\{X \leq x|X \leq b\} = \frac{P\{X \leq x \cap X \leq b\}}{P\{X \leq b\}} \tag{2.6-8}$$

for all events $\{X \leq b\}$ for which $P\{x \leq b\} \neq 0$. Two cases must be considered: one is where $b \leq x$; the second is where $x < b$. If $b \leq x$, the event $\{X \leq b\}$ is a subset of the event $\{X \leq x\}$, so $\{X \leq x\} \cap \{X \leq b\} = \{X \leq b\}$. Equation (2.6-8) becomes

$$F_X(x|X \leq b) = \frac{P\{X \leq x \cap X \leq b\}}{P\{X \leq b\}} = \frac{P\{X \leq b\}}{P\{X \leq b\}} = 1 \qquad b \leq x \tag{2.6-9}$$

When $x < b$ the event $\{X \leq x\}$ is a subset of the event $\{X \leq b\}$, so $\{X \leq x\} \cap \{X \leq b\} = \{X \leq x\}$ and (2.6-8) becomes

$$F_X(x|X \leq b) = \frac{P\{X \leq x \cap X \leq b\}}{P\{X \leq b\}} = \frac{P\{X \leq x\}}{P\{X \leq b\}} = \frac{F_X(x)}{F_X(b)} \qquad x < b \tag{2.6-10}$$

By combining the last two expressions, we obtain

$$F_X(x|X \leq b) = \begin{cases} \dfrac{F_X(x)}{F_X(b)} & x < b \\ 1 & b \leq x \end{cases} \tag{2.6-11}$$

The conditional density function derives from the derivative of (2.6-11):

$$f_X(x|X \leq b) = \begin{cases} \dfrac{f_X(x)}{F_X(b)} = \dfrac{f_X(x)}{\displaystyle\int_{-\infty}^{b} f_X(x)\,dx} & x < b \\ 0 & x \geq b \end{cases} \tag{2.6-12}$$

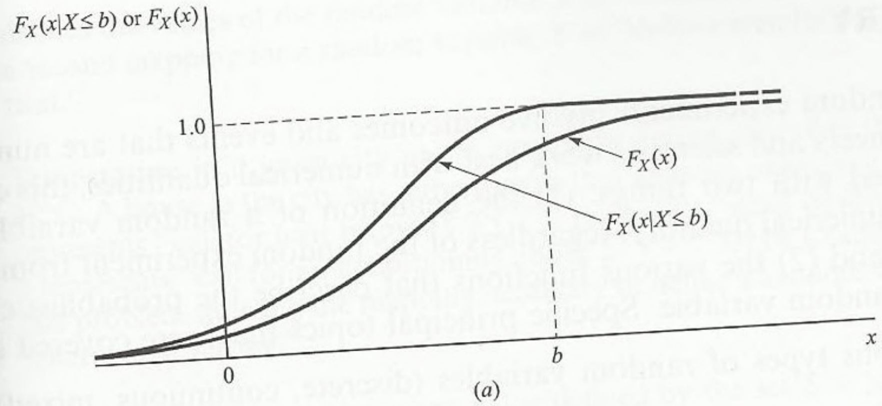
Figure 2.6-2 sketches possible functions representing (2.6-11) and (2.6-12).

From our assumptions that the conditioning event has nonzero probability, we have $0 < F_X(b) \leq 1$, so the expression of (2.6-11) shows that the conditional distribution function is never smaller than the ordinary distribution function:

$$F_X(x|X \leq b) \geq F_X(x) \tag{2.6-13}$$

A similar statement holds for the conditional density function of (2.6-12) wherever it is nonzero:

†Notation used has allowed for deletion of some braces for convenience. Thus, $F_X(x|\{X \leq b\})$ is written $F_X(x|X \leq b)$ and $P(\{X \leq x\} \cap \{X \leq b\})$ becomes $P\{X \leq x \cap X \leq b\}$.



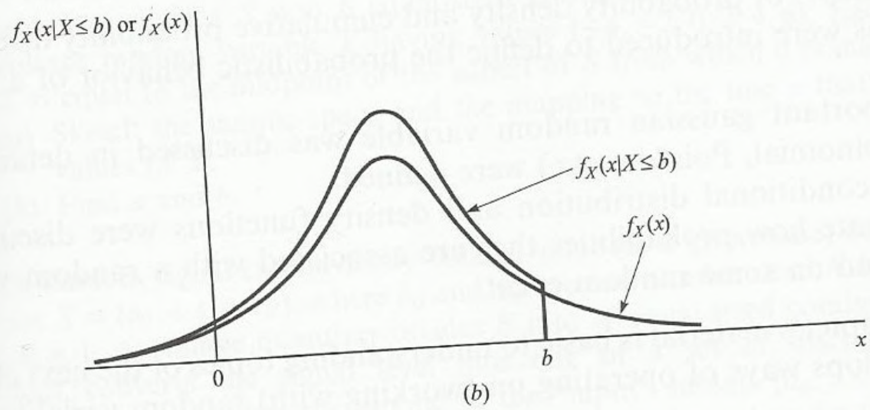


FIGURE 2.6-2
Possible distribution functions (*a*) and density functions (*b*) applicable to a conditioning event $B = \{X \le b\}$.

$$f_X(x|X \le b) \ge f_X(x) \qquad x < b \tag{2.6-14}$$

The principal results (2.6-11) and (2.6-12) can readily be extended to the more general event $B = \{a < X \le b\}$ (see Problem 2.6-2).

EXAMPLE 2.6-2. The radial "miss-distance" of landings from parachuting sky divers, as measured from a target's center, is a Rayleigh random variable with $b = 800\,\text{m}^2$ and $a = 0$. From (2.5-12) we have

$$F_X(x) = [1 - e^{-x^2/800}]u(x)$$

The target is a circle of 50-m radius with a bull's eye of 10-m radius. We find the probability of a parachuter hitting the bull's eye given that the landing is on the target.

The required probability is given by (2.6-11) with $x = 10$ and $b = 50$:

$$\begin{aligned} P(\text{bull's eye}|\text{landing on target}) &= F_X(10)/F_X(50) \\ &= (1 - e^{-100/800})/(1 - e^{-2500/800}) = 0.1229 \end{aligned}$$

Parachuter accuracy is such that about 12.29% of landings falling on the target will actually hit the bull's eye.

2.7 SUMMARY

Not all random experiments involve outcomes and events that are numerical. Since engineers and scientists work best with numerical quantities, this chapter is concerned with two things: (1) the definition of a random varaible; it is always a numerical quantity, regardless of the random experiment from which it derives; and (2) the various functions that describe the probabilistic behavior of a random variable. Specific principal topics that were covered are:

- The various types of random variables (discrete, continuous, mixed) were defined.
- The concepts of probability density and cumulative probability distribution functions were introduced to define the probabilistic behavior of a random variable.
- The important gaussian random variable was discussed in detail, while others (binomial, Poisson, etc.) were defined.
- Finally, conditional distribution and density functions were discussed to demonstrate how probabilities that are associated with a random variable can depend on some random event.

The chapter's material is basic to understanding topics of the next chapter, which develops ways of operating on (working with) random variables.

PROBLEMS

2.1-1. The sample space for an experiment is $S = \{0, 1, 2.5, 6\}$. List all possible values of the following random variables:
(*a*) $X = 2s$
(*b*) $X = 5s^2 - 1$
(*c*) $X = \cos(\pi s)$
(*d*) $X = (1 - 3s)^{-1}$

2.1-2. Work Problem 2.1-1 for $S = \{-2 < s \leq 5\}$.

2.1-3. Given that a random variable X has the following possible values, state if X is discrete, continuous, or mixed.
(*a*) $\{-20 < x < -5\}$
(*b*) $\{10, 12 < x \leq 14, 15, 17\}$
(*c*) $\{-10 \text{ for } s > 2 \text{ and } 5 \text{ for } s \leq 2, \text{ where } 1 < s \leq 6\}$
(*d*) $\{4, 3.1, 1, -2\}$

2.1-4. A random variable X is a function. So is probability P. Recall that the *domain* of a function is the set of values its argument may take on while its *range* is the set of corresponding values of the function. In terms of sets, events, and sample spaces, state the domain and range for X and P.

2.1-5. A man matches coin flips with a friend. He wins \$2 if coins match and loses \$2 if they do not match. Sketch a sample space showing possible outcomes for this experiment and illustrate how the points map onto the real line x that defines the values of the random variable $X =$ "dollars won on a trial." Show a second mapping for a random variable $Y =$ "dollars won by the friend on a trial."

2.1-6. Temperature in a given city varies randomly during any year from −21 to 49°C. A house in the city has a thermostat that assumes only three positions: 1 represents "call for heat below 18.3°C," 2 represents "dead or idle zone," and 3 represents "call for air conditioning above 21.7°C." Draw a sample space for this problem showing the mapping necessary to define a random variable $X =$ "thermostat setting."

2.1-7. A random voltage can have any value defined by the set $S = \{a \leq s \leq b\}$. A quantizer divides S into 6 equal-sized contiguous subsets and generates a voltage random variable X having values $\{-4, -2, 0, 2, 4, 6\}$. Each value of X is equal to the midpoint of the subset of S from which it is mapped.
(*a*) Sketch the sample space and the mapping to the line x that defines the values of X.
(*b*) Find a and b.

***2.1-8.** A random signal can have any voltage value (at a given time) defined by the set $S = \{a_0 < s \leq a_N\}$, where a_0 and a_N are real numbers and N is any integer $N \geq 1$. A voltage quantizer divides S into N equal-sized contiguous subsets and converts the signal level into one of a set of discrete levels a_n, $n = 1, 2, \ldots, N$, that correspond to the "input" subsets $\{a_{n-1} < s \leq a_n\}$. The set $\{a_1, a_2, \ldots, a_N\}$ can be taken as the discrete values of an "output" random variable X of the quantizer. If the smallest "input" subset is defined by $\Delta = a_1 - a_0$ and other subsets by $a_n - a_{n-1} = 2^{n-1}\Delta$, determine Δ and the quantizer levels a_n in terms of a_0, a_N, and N.

2.1-9. An honest coin is tossed three times.
(*a*) Sketch the applicable sample space S showing all possible elements. Let X be a random variable that has values representing the number of heads obtained on any triple toss. Sketch the mapping of S onto the real line defining X.
(*b*) Find the probabilities of the values of X.

2.1-10. Work Problem 2.1-9 for a biased coin for which $P\{\text{head}\} = 0.6$.

2.1-11. Resistor R_2 in Figure P2.1-11 is randomly selected from a box of resistors containing 180-Ω, 470-Ω, 1000-Ω, and 2200-Ω resistors. All resistor values have the same likelihood of being selected. The voltage E_2 is a discrete random variable. Find the set of values E_2 can have and give their probabilities.

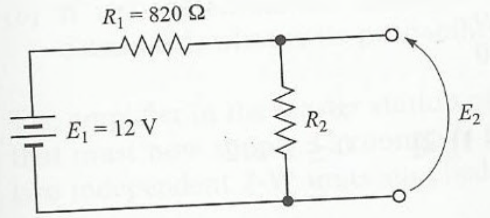


FIGURE P2.1-11

2.1-12. A sample space is defined by $S = \{1, 2 \leq s \leq 3, 4, 5\}$. A random variable is defined by $X = 2$ for $0 \leq s \leq 2.5$, $X = 3$ for $2.5 < s < 3.5$, and $X = 5$ for $3.5 \leq s \leq 6$.
(*a*) Is X discrete, continuous, or mixed?
(*b*) Give a set that defines the values X can have.

2.1-13. A gambler flips a fair coin three times.
(*a*) Draw a sample space S for this experiment. A random variable X representing his winnings is defined as follows: He loses \$1 if he gets no heads in three flips; he wins \$1, \$2, and \$3 if he obtains 1, 2, or 3 heads, respectively. Show how elements of S map to values of X.
(*b*) What are the probabilities of the various values of X?

2.1-14. A random current is described by the sample space $S = \{-4 \leq i \leq 12\}$. A random variable X is defined by

$$X(i) = \begin{cases} -2 & i \leq -2 \\ i & -2 < i \leq 1 \\ 1 & 1 < i \leq 4 \\ 6 & 4 < i \end{cases}$$

(*a*) Show, by a sketch, the value x into which the values of i are mapped by X.
(*b*) What type of random variable is X?

2.2-1. Bolts made on a production line are nominally designed to have a 760-mm length. A go-no-go testing device eliminates all bolts less than 650 mm and over 920 mm in length. The surviving bolts are then made available for sale and their lengths are known to be described by a uniform probability density function. A certain buyer orders all bolts that can be produced with a ±5% tolerance about the nominal length. What fraction of the production line's output is he purchasing?

2.2-2. Find and sketch the density and distribution functions for the random variables of parts (*a*), (*b*), and (*c*) in Problem 2.1-1 if the sample space elements have equal likelihoods of occurrence.

2.2-3. If temperature in Problem 2.1-6 is uniformly distributed, sketch the density and distribution functions of the random variable X.

2.2-4. For the uniform random variable defined by (2.5-7) find:
(*a*) $P\{0.9a + 0.1b < X \leq 0.7a + 0.3b\}$
(*b*) $P\{(a + b)/2 < X \leq b\}$

2.2-5. Determine which of the following are valid distribution functions:

(*a*) $$G_X(x) = \begin{cases} 1 - e^{-x/2} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(*b*) $$G_X(x) = \begin{cases} 0 & x < 0 \\ 0.5 + 0.5\sin[\pi(x - 1)/2] & 0 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

(*c*) $$G_X(x) = \frac{x}{a}[u(x - a) - u(x - 2a)]$$

2.2-6. A function $G_X(x) = a[1 + (2/\pi)\sin^{-1}(x/c)]\,\text{rect}(x/2c) + (a+b)u(x-c)$ is defined for all $\infty < x < \infty$, where $c > 0$, b, and a are real constants and rect$(\cdot)$ is defined by (E-2). Find any conditions on a, b, and c that will make $G_X(x)$ a valid probability distribution function. Discuss what choices of constants correspond to a continuous, discrete, or mixed random variable.

2.2-7. (*a*) Generalize Problem 2.2-5(*a*) by finding values of real constants a and b such that

$$G_X(x) = [1 - a\exp(-x/b)]u(x)$$

is a valid distribution function.

(*b*) Are there any values of a and b such that $G_X(x)$ corresponds to a mixed random variable X?

2.2-8. (*a*) Find the probabilities associated with all values of the random variable X of Problem 2.1-14.

(*b*) Sketch the probability distribution function of the random variable X.

2.2-9. A random variable X has the distribution function

$$F_X(x) = \sum_{n=1}^{12} \frac{n^2}{650} u(x-n)$$

Find the probabilities: (*a*) $P\{-\infty < X \le 6.5\}$, (*b*) $P\{X > 4\}$, and (*c*) $P\{6 < X \le 9\}$.

2.2-10. If the function

$$G_X(x) = K \sum_{n=1}^{N} n^3 u(x-n)$$

must be a valid probability distribution function, determine K to make it valid. (*Hint*: Use a series from Appendix C.)

2.3-1. Determine the real constant a, for arbitrary real constants m and $0 < b$, such that

$$f_X(x) = ae^{-|x-m|/b}$$

is a valid density function (called the *Laplace density*).

2.3-2. An intercom system master station provides music to six hospital rooms. The probability that any one room will be switched on and draw power at any time is 0.4. When on, a room draws 0.5 W.

(*a*) Find and plot the density and distribution functions for the random variable "power delivered by the master station."

(*b*) If the master-station amplifier is overloaded when more than 2 W is demanded, what is its probability of overload?

***2.3-3.** The amplifier in the master station of Problem 2.3-2 is replaced by a 4-W unit that must now supply 12 rooms. Is the probability of overload better than if two independent 2-W units supplied six rooms each?

2.3-4. Jusify that a distribution function $F_X(x)$ satisfies (2.2-2*a*, *b*, *c*).

2.3-5. Use the definition of the impulse function to evaluate the following integrals. (*Hint*: Refer to Appendix A.)

(*a*) $\int_{3}^{4} (3x^2 + 2x - 4)\delta(x - 3.2)\,dx$

(*b*) $\int_{-\infty}^{\infty} \cos(6\pi x)\delta(x - 1)\,dx$

(*c*) $\int_{-\infty}^{\infty} \frac{24\delta(x-2)\,dx}{x^4 + 3x^2 + 2}$

(*d*)† $\int_{-\infty}^{\infty} \delta(x - x_0)e^{-j\omega x}\,dx$

(*e*) $\int_{-3}^{3} u(x - 2)\delta(x - 3)\,dx$

2.3-6. Show that the properties of a density function $f_X(x)$, as given by (2.3-6), are valid.

2.3-7. For the random variable defined in Example 2.3-1, find:
(*a*) $P\{x_0 - 0.6\alpha < X \le x_0 + 0.3\alpha\}$
(*b*) $P\{X = x_0\}$

2.3-8. Find a constant $b > 0$ so that the function

$$f_X(x) = \begin{cases} e^{3x}/4 & 0 \le x \le b \\ 0 & \text{elsewhere} \end{cases}$$

is a valid probability density.

2.3-9. Given the function

$$g_X(x) = 4\cos(\pi x/2b)\,\mathrm{rect}\,(x/2b)$$

find a value of b so that $g_X(x)$ is a valid probability density.

2.3-10. A random variable X has the density function

$$f_X(x) = (\tfrac{1}{2})u(x)\exp(-x/2)$$

Define events $A = \{1 < X \le 3\}$, $B = \{X \le 2.5\}$, and $C = A \cap B$. Find the probabilities of events (*a*) A, (*b*) B, and (*c*) C.

***2.3-11.** Let $\phi(x)$ be a continuous, but otherwise arbitrary real function, and let a and b be real constants. Find $G(a, b)$ defined by

$$G(a, b) = \int_{-\infty}^{\infty} \phi(x)\,\delta(ax + b)\,dx$$

(*Hint*: Use the definition of the impulse function.)

†The quantity j is the unit-imaginary; that is, $j = \sqrt{-1}$.

2.3-12. For real constants $b > 0$, $c > 0$, and any a, find a condition on constant a and a relationship between c and a (for given b) such that the function

$$f_X(x) = \begin{cases} a[1 - (x/b)] & 0 \le x \le c \\ 0 & \text{elsewhere} \end{cases}$$

is a valid probability density.

2.3-13. Use the properties or definition of the impulse function (Appendix A) to evaluate the following integrals:

(a) $\int_{-\infty}^{\infty} \delta(x+5)\frac{x^2}{1+x^2}\,dx$

(b) $\int_{-\infty}^{\infty} \delta(x-3)\cos(\pi x/6)\,dx$

(c) $\int_{-\infty}^{\infty} e^{-4(x+1)}\delta(x+1)\,dx$

2.3-14. Work Problem 2.3-13 except for the following integrals:

(a) $\int_{-2}^{6} [\delta(x-1) + \delta(x+3) + \delta(x-5)]\,dx$

(b) $\int_{-\infty}^{6} \delta(x-7)u(x+3)\,dx$

(c) $\int_{-3}^{2} [\delta(x-1) - \delta(x+2)]\frac{e^{-2x^2}}{1+x^2+x^4}\,dx$

2.3-15. Find a value for constant A such that

$$f_X(x) = \begin{cases} 0 & x < -1 \\ A(1-x^2)\cos(\pi x/2) & -1 \le x \le 1 \\ 0 & 1 < x \end{cases}$$

is a valid probability density function.

2.4-1. A random variable X is gaussian with $a_X = 0$ and $\sigma_X = 1$.
(a) What is the probability that $|X| > 2$?
(b) What is the probability that $X > 2$?

2.4-2. Work Problem 2.4-1 if $a_X = 4$ and $\sigma_X = 2$.

2.4-3. For the gaussian density function of (2.4-1), show that

$$\int_{-\infty}^{\infty} x f_X(x)\,dx = a_X$$

2.4-4. For the gaussian density function of (2.4-1), show that

$$\int_{-\infty}^{\infty} (x - a_X)^2 f_X(x)\,dx = \sigma_X^2$$

2.4-5. A production line manufactures 1000-Ω resistors that must satisfy a 10% tolerance.

(*a*) If resistance is adequately described by a gaussian random variable X for which $a_X = 1000\,\Omega$ and $\sigma_X = 40\,\Omega$, what fraction of the resistors is expected to be rejected?
(*b*) If a machine is not properly adjusted, the product resistances change to the case where $a_X = 1050\,\Omega$ (5% shift). What fraction is now rejected?

2.4-6. Cannon shell impact position, as measured along the line of fire from the target point, can be described by a gaussian random variable X. It is found that 15.15% of shells fall 11.2 m or farther from the target in a direction toward the cannon, while 5.05% fall farther than 95.6 m beyond the target. what are a_X and σ_X for X?

2.4-7. A gaussian random variable X has $a_X = 2$, and $\sigma_X = 2$.
(*a*) Find $P\{X > 1.0\}$.
(*b*) Find $P\{X \leq -1.0\}$.

2.4-8. In a certain "junior" olympics, javelin throw distances are well approximated by a gaussian distribution for which $a_X = 30$ m and $\sigma_X = 5$ m. In a qualifying round, contestants must throw farther than 26 m to qualify. In the main event the record throw is 42 m.
(*a*) What is the probability of being disqualified in the qualifying round?
(*b*) In the main event what is the probability the record will be broken?

2.4-9. Suppose height to the bottom of clouds is a gaussian random variable X for which $a_X = 4000$ m, and $\sigma_X = 1000$ m. A person bets that cloud height tomorrow will fall in the set $A = \{1000\text{ m} < X \leq 3300\text{ m}\}$ while a second person bets that height will be satisfied by $B = \{2000\text{ m} < X \leq 4200\text{ m}\}$. A third person bets they are both correct. Find the probabilities that each person will win the bet.

2.4-10. The output voltage X from the receiver in a particular binary digital communication system, when a binary zero is being received, is gaussian (noise only) as defined by $a_X = 0$ and $\sigma_X = 0.3$. When a binary one is being received it is also gaussian (signal-plus-noise now), but as defined by $a_X = 0.9$ and $\sigma_X = 0.3$. The receiver's decision logic specifies that at the end of a binary (bit) interval, if $X > 0.45$ a binary one is being received. If $X \leq 0.45$ a binary zero is decided. If it is given that a binary zero is truly being received, find the probabilities that (*a*) a binary one (mistake) will be decided, and (*b*) a binary zero is decided (correct decision).

2.4-11. A gaussian random voltage X for which $a_X = 0$ and $\sigma_X = 4.2$ V appears across a 100-Ω resistor with a power rating of 0.25 W. What is the probability that the voltage will cause an instantaneous power that exceeds the resistor's rating?

2.4-12. Work Problem 2.4-11 except assume a 0.5-W resistor.

2.4-13. For the gaussian random variable, show that the curve's points of inflection (where the first derivative of the probability density function with respect to x has a zero slope) occur at $a_X \pm \sigma_X$.

2.4-14. A random variable X is known to be gaussian with $a_X = 1.6$ and $\sigma_X = 0.4$. Find: (*a*) $P\{1.4 < X \leq 2.0\}$, and (*b*) $P\{-0.6 < (X - 1.6) \leq 0.6\}$.

2.4-15. The radial distance to the impact points for shells fired over land by a cannon is well-approximated as a gaussian random variable with $a_X = 2000$ m and $\sigma_X = 40$ m when the cannon is aimed at a target located at 1980 m distance.
(*a*) Find the probability that shells will fall within ± 68 m of the target.
(*b*) Find the probability that shells will fall at distances of 2050 m or more.

2.4-16. Assume that the time of arrival of birds at a particular place on a migratory route, as measured in days from the first of the year (January 1 is the first day), is approximated as a gaussian random variable X with $a_X = 200$ and $\sigma_X = 20$ days.
(*a*) What is the probability the birds arrive after 160 days but on or before the 210th day?
(*b*) What is the probability the birds will arrive after the 231st day?

2.4-17. Assume fluorescent lamps made by a manufacturer have a probability of 0.05 of being inoperable when new. A person purchases eight of the lamps for home use.
(*a*) Plot the probability distribution function for a random variable "the number of inoperable lamps."
(*b*) What is the probability that exactly one lamp is inoperable of the eight?
(*c*) What is the probability that all eight lamps are functional?
(*d*) Determine the probability that one or more lamps are not operable.

2.5-1. (*a*) Use the exponential density of (2.5-9) and solve for I_2 defined by

$$I_2 = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx$$

(*b*) Solve for I_1 defined by

$$I_1 = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

(*c*) Verify that I_1 and I_2 satisfy the equation $I_2 - I_1^2 = b^2$.

2.5-2. Verify that the maximum value of $f_X(x)$ for the Rayleigh density function of (2.5-11) occurs at $x = a + \sqrt{b/2}$ and is equal to $\sqrt{2/b}\,\exp(-\frac{1}{2}) \approx 0.607\sqrt{2/b}$. This value of x is called the *mode* of the random variable. (In general, a random variable may have more than one such value—explain.)

2.5-3. The lifetime of a system expressed in weeks is a Rayleigh random variable X for which

$$f_X(x) = \begin{cases} (x/200)e^{-x^2/400} & 0 \leq x \\ 0 & x < 0 \end{cases}$$

(*a*) What is the probability that the system will not last a full week?
(*b*) What is the probability the system lifetime will exceed one year?

2.5-4. The *Cauchy*† random variable has the probability density function

$$f_X(x) = \frac{b/\pi}{b^2 + (x-a)^2}$$

for real numbers $0 < b$ and $-\infty < a < \infty$. Show that the distribution function of X is

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x-a}{b}\right)$$

2.5-5. The *log-normal density* function is given by

$$f_X(x) = \begin{cases} \dfrac{\exp\{-[\ln(x-b) - a_X]^2/2\sigma_X^2\}}{\sqrt{2\pi}\sigma_X(x-b)} & x \geq b \\ 0 & x < b \end{cases}$$

for real constants $0 < \sigma_X$, $-\infty < a_X < \infty$, and $-\infty < b < \infty$, where $\ln(x)$ denotes the natural logarithm of x. Show that the corresponding distribution function is

$$F_X(x) = \begin{cases} F\left[\dfrac{\ln(x-b) - a_X}{\sigma_X}\right] & x \geq b \\ 0 & x < b \end{cases}$$

where $F(\cdot)$ is given by (2.4-3).

2.5-6. A random variable X is known to be Poisson with $b = 4$.
(*a*) Plot the density and distribution functions for this random variable.
(*b*) What is the probability of the event $\{0 \leq X \leq 5\}$?

2.5-7. The number of cars arriving at a certain bank drive-in window during any 10-min period is a Poisson random variable X with $b = 2$. Find:
(*a*) The probability that more than 3 cars will arrive during any 10-min period.
(*b*) The probability that no cars will arrive.

2.5-8. Let X be a Rayleigh random variable with $a = 0$. Find the probability that X will have values larger than its mode (see Problem 2.5-2).

2.5-9. A certain large city averages three murders per week and their occurrences follow a Poisson distribution.
(*a*) What is the probability that there will be five or more murders in a given week?
(*b*) On the average, how many weeks a year can this city expect to have no murders?
(*c*) How many weeks per year (average) can the city expect the number of murders per week to equal or exceed the average number per week?

2.5-10. A certain military radar is set up at a remote site with no repair facilities. If the radar is known to have a *mean-time-between-failures* (MTBF) of 200 h, find

†After the French mathematician Augustin Louis Cauchy (1789–1857).

the probability that the radar is still in operation one week later when picked up for maintenance and repairs.

2.5-11. If the radar of Problem 2.5-10 is permanently located at the remote site, find the probability that it will be operational as a function of time since its setup.

2.5-12. A computer undergoes downtime if a certain critical component fails. This component is known to fail at an average rate of once per four weeks. No significant downtime occurs if replacement components are on hand because repair can be made rapidly. There are three components on hand, and ordered replacements are not due for six weeks.
(*a*) What is the probability of significant downtime occurring before the ordered components arrive?
(*b*) If the shipment is delayed two weeks, what is the probability of significant downtime occurring before the shipment arrives?

2.5-13. The envelope (amplitude) of the output signal of a radar system that is receiving only noise (no signal) is a Rayleigh random voltage X for which $a = 0$ and $b = 2$ V. The system gets a false target detection if X exceeds a threshold level V volts. How large must V be to make the probability of false detection 0.001?

2.6-1. Rework Example 2.6-1 to find $f_X(x|B_2)$ and $F_X(x|B_2)$. Sketch the two functions.

*2.6-2. Extend the analysis of the text that leads to (2.6-11) and (2.6-12) to the more general event $B = \{a < X \leq b\}$. Specifically, show that now

$$F_X(x|a < X \leq b) = \begin{cases} 0 & x < a \\ \dfrac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)} & a \leq x < b \\ 1 & b \leq x \end{cases}$$

and

$$f_X(x|a < X \leq b) = \begin{cases} 0 & x < a \\ \dfrac{f_X(x)}{F_X(b) - F_X(a)} = \dfrac{f_X(x)}{\displaystyle\int_a^b f_X(x)\,dx} & a \leq x < b \\ 0 & b \leq x \end{cases}$$

*2.6-3. Consider the system having a lifetime defined by the random variable X in Problem 2.5-3. Given that the system will survive beyond 20 weeks, find the probability that it will survive beyond 26 weeks.

*2.6-4. Assume the lifetime of a laboratory research animal is defined by a Rayleigh density with $a = 0$ and $b = 30$ weeks in (2.5-11) and (2.5-12). If for some clinical reasons it is known that the animal will live at *most* 20 weeks, what is the probability it will live 10 weeks or less?

*2.6-5. Suppose the depth of water, measured in meters, behind a dam is described by an exponential random variable having a density

$$f_X(x) = (1/13.5)u(x)\exp(-x/13.5)$$

There is an emergency overflow at the top of the dam that prevents the dept from exceeding 40.6 m. There is a pipe placed 32.0 m below the overflo (ignore the pipe's finite diameter) that feeds water to a hydroelectric generato
(*a*) What is the probability that water is wasted through emergency overflow
(*b*) Given that water is not wasted in overflow, what is the probability th generator will have water to drive it?
(*c*) What is the probability that water will be too low to produce power?

***2.6-6.** In Problem 2.6-5 find and sketch the distribution and density functions c water depth given that water will be deep enough to generate power but n water is wasted by emergency overflow. Also sketch for comparisons th distribution and density of water depth without any conditions.

***2.6-7.** In Example 2.6-2 a parachuter is an "expert" if he hits the bull's eye. If he fal outside the bull's eye but within a circle of 25-m radius, he is called "qualified for competition. Given that a parachuter is not an expert but hits the targe what is the probability of being "qualified"?

2.6-8. In a game show contestants choose one of three doors to determine what pri they win. History shows that the three doors, 1, 2, and 3, are chosen wit probabilities 0.30, 0.45, and 0.25, respectively. It is also known that given doo 1 is chosen, the probabilities of winning prizes of \$0, \$100, and \$1000 are 0.1 0.20, and 0.70. For door 2 the respective probabilities are 0.50, 0.35, and 0.1 and for door 3 they are 0.80, 0.15, and 0.05. If X is a random variab describing dollars won, and D describes the door selected (values of D ar $D_1 = 1, D_2 = 2$, and $D_3 = 3$), find: (*a*) $F_X(x|D = D_1)$ and $f_X(x|D = D_1)$, ($f_X(x|D = D_2)$, (*c*) $f_X(x|D = D_3)$, and (*d*) $f_X(x)$.

2.6-9. Again consider the game show of Problem 2.6-8 and find the probabilities c winning (*a*) \$0, (*b*) \$100, and (*c*) \$1000.

***2.6-10.** Divers return each day to the site of a sunken treasure ship. Due to rando navigational errors, they arrive with a radial positional error (from the tru site) described by a random variable X (in kilometers) defined by

$$F_X(x) = [1 - e^{-x^3/2}]u(x)$$

(this is the *Weibull*† distribution; see Appendix F). If they must arrive with a error of not more than 1.2 km to prevent having to move to a new positio and within 0.6 km for optimum use of air tanks, what is the probability c optimum use of tanks given they arrive on site on the first effort?

***2.6-11.** For the navigational errors of Problem 2.6-10 find and plot the condition density $f_X(x|X \leq 1.2\,\text{km})$.

†After Ernst Hjalmar Waloddi Weibull (1887–1979), a Swedish applied physicist.

CHAPTER 3

Operations on One Random Variable—Expectation

3.0 INTRODUCTION

The random variable was introduced in Chapter 2 as a means of providing a systematic definition of events defined on a sample space. Specifically, it formed a mathematical model for describing characteristics of some real, physical world random phenomenon. In this chapter, we extend our work to include some important *operations* that may be performed on a random variable. Most of these operations are based on a single concept—expectation.

3.1 EXPECTATION

Expectation is the name given to the process of averaging when a random variable is involved. For a random variable X, we use the notation $E[X]$, which may be read "the mathematical *expectation* of X," "the *expected* value of X," "the *mean* value of X," or "the *statistical average* of X." Occasionally we also use the notation $\bar{X}$, which is read the same way as $E[X]$; that is, $\bar{X} = E[X]$.†

Nearly everyone is familiar with averaging procedures. An example that serves to tie a familiar problem to the new concept of expectation may be the easiest way to proceed.

†Up to this point in this book an overbar has represented the complement of a set or event. Henceforth, unless specifically stated otherwise, the overbar will always represent a mean value.

EXAMPLE 3.1-1. Ninety people are randomly selected and the fractional dollar value of coins in their pockets is counted. If the count goes above a dollar, the dollar value is discarded and only the portion from 0¢ to 99¢ is accepted. It is found that 8, 12, 28, 22, 15, and 5 people had 18¢, 45¢, 64¢, 72¢, 77¢, and 95¢ in their pockets, respectively.

Our everyday experiences indicate that the average of these values is

$$\begin{aligned} \text{Average } \$ &= 0.18\left(\frac{8}{90}\right) + 0.45\left(\frac{12}{90}\right) + 0.64\left(\frac{28}{90}\right) + 0.72\left(\frac{22}{90}\right) \\ &\quad + 0.77\left(\frac{15}{90}\right) + 0.95\left(\frac{5}{90}\right) \\ &\approx \$0.632 \end{aligned}$$

Expected Value of a Random Variable

The everyday averaging procedure used in the above example carries over directly to random variables. In fact, if X is the discrete random variable "fractional dollar value of pocket coins," it has 100 discrete values x_i that occur with probabilities $P(x_i)$, and its expected value $E[X]$ is found in the same way as in the example:

$$E[X] = \sum_{i=1}^{100} x_i P(x_i) \tag{3.1-1}$$

The values x_i identify with the fractional dollar values in the example, while $P(x_i)$ is identified with the ratio of the number of people for the given dollar value to the total number of people. If a large number of people had been used in the "sample" of the example, all fractional dollar values would have shown up and the ratios would have approached $P(x_i)$. Thus, the average in the example would have become more like (3.1-1) for many more than 90 people.

In general, the expected value of any random variable X is defined by

$$E[X] = \bar{X} = \int_{-\infty}^{\infty} x f_X(x)\,dx \tag{3.1-2}$$

If X happens to be discrete with N possible values x_i having probabilities $P(x_i)$ of occurrence, then

$$f_X(x) = \sum_{i=1}^{N} P(x_i)\delta(x - x_i) \tag{3.1-3}$$

from (2.3-5). Upon substitution of (3.1-3) into (3.1-2), we have

$$E[X] = \sum_{i=1}^{N} x_i P(x_i) \qquad \text{discrete random variable} \tag{3.1-4}$$

Hence, (3.1-1) is a special case of (3.1-4) when $N = 100$. For some discrete random variables, N may be infinite in (3.1-3) and (3.1-4).

EXAMPLE 3.1-2. We determine the mean value of the continuous, exponentially distributed random variable for which (2.5-9) applies:

$$f_X(x) = \begin{cases} \frac{1}{b}e^{-(x-a)/b} & x > a \\ 0 & x < a \end{cases}$$

From (3.1-2) and an integral from Appendix C:

$$E[X] = \int_a^\infty \frac{x}{b}e^{-(x-a)/b}\,dx = \frac{e^{a/b}}{b}\int_a^\infty xe^{-x/b}\,dx = a+b$$

If a random variable's density is symmetrical about a line $x = a$, then $E[X] = a$; that is,

$$E[X] = a \quad \text{if} \quad f_X(x+a) = f_X(-x+a) \tag{3.1-5}$$

Expected Value of a Function of a Random Variable

As will be evident in the next section, many useful parameters relating to a random variable X can be derived by finding the expected value of a real function $g(\cdot)$ of X. It can be shown (see Papoulis, 1965, p. 142) that this expected value is given by

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)f_X(x)\,dx \tag{3.1-6}$$

If X is a discrete random variable, (3.1-3) applies and (3.1-6) reduces to

$$E[g(X)] = \sum_{i=1}^{N} g(x_i)P(x_i) \qquad \text{discrete random variable} \tag{3.1-7}$$

where N may be infinite for some random variables.

EXAMPLE 3.1-3. It is known that a particular random voltage can be represented as a Rayleigh random variable V having a density function given by (2.5-11) with $a = 0$ and $b = 5$. The voltage is applied to a device that generates a voltage $Y = g(V) = V^2$ that is equal, numerically, to the power in V (in a 1-Ω resistor). We find the average power in V by means of (3.1-6):

$$\text{Power in } V = E[g(V)] = E[V^2] = \int_0^\infty \frac{2v^3}{5}e^{-v^2/5}\,dv$$

By letting $\xi = v^2/5$, $d\xi = 2v\,dv/5$, we obtain

$$\text{Power in } V = 5\int_0^\infty \xi e^{-\xi}\,d\xi = 5\text{ W}$$

after using (C-46).

Note that if $g(X)$ in (3.1-6) is a sum of N functions $g_n(X)$, $n = 1, 2, \ldots, N$, then the expected value of the sum of N functions of a random variable X is the sum of the N expected values of the individual functions of the random variable.

EXAMPLE 3.1-4. A problem in communication systems is how to define the information of a source. Consider modeling a source capable of issuing any one of L distinct symbols (messages) represented as values x_i, $i = 1, 2, \ldots, L$, of a discrete random variable X ($L = 2$ is the binary case). Let $P(x_i)$ be the probabilities of the symbols $X = x_i$. We ask what is the information contained in this source, on the average. We form three considerations.

First, we reason that information should be largest for source outputs with small probabilities. After all, it conveys little information to predict hot, dry weather for the Sahara desert since these conditions prevail almost all the time. But to predict cool heavy rain carries much "information." Next, information from two independent sources should reasonably add. Finally, information should be a positive quantity (a choice we make) and should be zero for an event that is certain to occur. The only function with these characteristics is the logarithm [Carlson (1975), p. 343]. Now since two quantities represent the smallest measure of choice, the logarithm to the base 2 is chosen for measuring information, and its unit is called the *bit*.

For our source we are led to define the information in symbol x_i as $\log_2[1/P(x_i)] = -\log_2[P(x_i)]$. By use of (3.1-7) we obtain the average information, or *entropy*, of a discrete source as

$$H = -\sum_{i=1}^{L} P(x_i)\log_2[P(x_i)] = \frac{-1}{\ln(2)}\sum_{i=1}^{L} P(x_i)\ln[P(x_i)]$$

where $\ln(\cdot)$ is the natural logarithm (to base e). The unit of H is bits/symbol; it results from averaging the information over all source symbols.

*Conditional Expected Value

If, in (3.1-2), $f_X(x)$ is replaced by the conditional density $f_X(x|B)$, where B is any event defined on the sample space, we have the *conditional expected value* of X, denoted $E[X|B]$:

$$E[X|B] = \int_{-\infty}^{\infty} x f_X(x|B)\,dx \tag{3.1-8}$$

One way to define event B, as shown in Chapter 2, is to let it depend on the random variable X by defining

$$B = \{X \le b\} \qquad -\infty < b < \infty \tag{3.1-9}$$

We showed there that

$$f_X(x|X \le b) = \begin{cases} \dfrac{f_X(x)}{\int_{-\infty}^{b} f_X(x)\,dx} & x < b \\ 0 & x \ge b \end{cases} \tag{3.1-10}$$

Thus, by substituting (3.1-10) into (3.1-8):

$$E[X|X \le b] = \frac{\int_{-\infty}^{b} x f_X(x)\,dx}{\int_{-\infty}^{b} f_X(x)\,dx} \tag{3.1-11}$$

which is the mean value of X when X is constrained to the set $\{X \le b\}$.

3.2 MOMENTS

An immediate application of the expected value of a function $g(\cdot)$ of a random variable X is in calculating moments. Two types of moments are of interest, those about the origin and those about the mean.

Moments about the Origin

The function

$$g(X) = X^n \qquad n = 0, 1, 2, \ldots \tag{3.2-1}$$

when used in (3.1-6) gives the moments about the origin of the random variable X. Denote the nth moment by m_n. Then,

$$m_n = E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\,dx \tag{3.2-2}$$

Clearly $m_0 = 1$, the area of the function $f_X(x)$, while $m_1 = \bar{X}$, the expected value of X.

Central Moments

Moments about the mean value of X are called *central moments* and are given the symbol μ_n. They are defined as the expected value of the function

$$g(X) = (X - \bar{X})^n \qquad n = 0, 1, 2, \ldots \tag{3.2-3}$$

which is

$$\mu_n = E[(X - \bar{X})^n] = \int_{-\infty}^{\infty} (x - \bar{X})^n f_X(x)\,dx \tag{3.2-4}$$

The moment $\mu_0 = 1$, the area of $f_X(x)$, while $\mu_1 = 0$. (Why?)

Variance and Skew

The second central moment μ_2 is so important we shall give it the name *variance* and the special notation σ_X^2. Thus, variance is given by†

$$\sigma_X^2 = \mu_2 = E[(X - \bar{X})^2] = \int_{-\infty}^{\infty} (x - \bar{X})^2 f_X(x)\, dx \tag{3.2-5}$$

The positive square root σ_X of variance is called the *standard deviation* of X; it is a measure of the spread in the function $f_X(x)$ about the mean.

Variance can be found from a knowledge of first and second moments. By expanding (3.2-5), we have‡

$$\begin{aligned}\sigma_X^2 &= E[X^2 - 2\bar{X}X + \bar{X}^2] = E[X^2] - 2\bar{X}E[X] + \bar{X}^2 \\ &= E[X^2] - \bar{X}^2 = m_2 - m_1^2\end{aligned} \tag{3.2-6}$$

EXAMPLE 3.2-1. Let X have the exponential density function given in Example 3.1-2. By substitution into (3.2-5), the variance of X is

$$\sigma_X^2 = \int_a^{\infty} (x - \bar{X})^2 \frac{1}{b} e^{-(x-a)/b}\, dx$$

By making the change of variable $\xi = x - \bar{X}$ we obtain

$$\sigma_X^2 = \frac{e^{-(\bar{X}-a)/b}}{b} \int_{a-\bar{X}}^{\infty} \xi^2 e^{-\xi/b}\, d\xi = (a + b - \bar{X})^2 + b^2$$

after using an integral from Appendix C. However, from Example 3.1-2, $\bar{X} = E[X] = (a + b)$, so

$$\sigma_X^2 = b^2$$

The reader may wish to verify this result by finding the second moment $E[X^2]$ and using (3.2-6).

The third central moment $\mu_3 = E[(X - \bar{X})^3]$ is a measure of the asymmetry of $f_X(x)$ about $x = \bar{X} = m_1$. It will be called the *skew* of the density function. If a density is symmetric about $x = \bar{X}$, it has zero skew. In fact, for this case $\mu_n = 0$ for all odd values of n. (Why?) The normalized third central moment μ_3/σ_X^3 is known as the *skewness* of the density function, or, alternatively, as the *coefficient of skewness*.

EXAMPLE 3.2-2. We continue Example 3.2-1 and compute the skew and coefficient of skewness for the exponential density. From (3.2-4) with $n = 3$ we have

†The subscript indicates that σ_X^2 is the variance of a random variable X. For a random variable Y its variance would be σ_Y^2.

‡We use the fact that the expected value of a sum of functions of X equals the sum of expected values of individual functions, as previously noted.

$$\mu_3 = E[(X-\bar{X})^3] = E[X^3 - 3\bar{X}X^2 + 3\bar{X}^2X - \bar{X}^3]$$
$$= \overline{X^3} - 3\bar{X}\overline{X^2} + 2\bar{X}^3 = \overline{X^3} - 3\bar{X}(\sigma_X^2 + \bar{X}^2) + 2\bar{X}^3$$
$$= \overline{X^3} - 3\bar{X}\sigma_X^2 - \bar{X}^3$$

Next, we have

$$\overline{X^3} = \int_a^\infty \frac{x^3}{b} e^{-(x-a)/b}\, dx = a^3 + 3a^2b + 6ab^2 + 6b^3$$

after using (C-48). On substituting $\bar{X} = a + b$ and $\sigma_X^2 = b^2$ from the earlier example, and reducing the algebra we find

$$\mu_3 = 2b^3$$
$$\frac{\mu_3}{\sigma_X^3} = 2$$

This density has a relatively large coefficient of skewness, as can be seen intuitively from Figure 2.5-3.

Chebychev's Inequality

A useful tool in some probability problems is Chebychev's inequality.† For a random variable X with mean value $\bar{X}$ and variance σ_X^2, it states that

$$P\{|X - \bar{X}| \geq \epsilon\} \leq \sigma_X^2/\epsilon^2 \tag{3.2-7}$$

for any $\epsilon > 0$. This expression can be demonstrated by integration of the probability density, using (2.3-6c):

$$P\{|X-\bar{X}| \geq \epsilon\} = \int_{-\infty}^{\bar{X}-\epsilon} f_X(x)\,dx + \int_{\bar{X}+\epsilon}^{\infty} f_X(x)\,dx = \int_{|x-\bar{X}|\geq\epsilon}^{\infty} f_X(x)\,dx \tag{3.2-8}$$

But since

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x-\bar{X})^2 f_X(x)\,dx \geq \int_{|x-\bar{X}|\geq\epsilon}^{\infty} (x-\bar{X})^2 f_X(x)\,dx$$
$$\geq \epsilon^2 \int_{|x-\bar{X}|\geq\epsilon}^{\infty} f_X(x)\,dx = \epsilon^2 P\{|x-\bar{X}| \geq \epsilon\} \tag{3.2-9}$$

must be true, we solve to show the validity of (3.2-7).

EXAMPLE 3.2-3. We find the largest probability that any random variable's values are smaller than its mean by 3 standard deviations or larger than its mean by the same amount. This probability is $P\{X \geq \bar{X} + 3\sigma_X\} + P\{X \leq \bar{X} - 3\sigma_X\} = P\{|X - \bar{X}| \geq 3\sigma_X\}$. From (3.2-7) with $\epsilon = 3\sigma_X$ we have $P\{|X-\bar{X}| \geq 3\sigma_X\} \leq \sigma_X^2/(3\sigma_X)^2 = 1/9$, or about 11.1%.

†After the Russian mathematician Pafnuty Lvovich Chebychev (1821–1894).

By a procedure similar to that above, an alternative form of Chebychev's inequality can be proved. It is

$$P\{|X - \bar{X}| < \epsilon\} \geq 1 - (\sigma_X^2/\epsilon^2) \tag{3.2-10}$$

for any $\epsilon > 0$. An interesting result derives immediately from (3.2-10). If $\sigma_X^2 \to 0$ for a random variable, then $P\{|X - \bar{X}| < \epsilon\} \to 1$, for any ϵ. For arbitrarily small ϵ we have $P\{|X - \bar{X}| \to 0\} \to 1$ or $P\{X = \bar{X}\} \to 1$. In other words, if the variance of a random variable X approaches zero, the probability approaches 1 that X will equal its mean value.

Markov's Inequality

Another inequality that is useful in probability problems is *Markov's inequality*, which applies to a nonnegative random variable X; it is

$$P\{X \geq a\} \leq E[X]/a \qquad a > 0 \tag{3.2-11}$$

The restriction to nonnegative random variables is relieved by *Chernoff's inequality*, which is developed in Example 3.3-3 below.

*3.3 FUNCTIONS THAT GIVE MOMENTS

Two functions can be defined that allow moments to be calculated for a random variable X. They are the characteristic function and the moment generating function.

*Characteristic Function

The *characteristic function* of a random variable X is defined by

$$\Phi_X(\omega) = E[e^{j\omega X}] \tag{3.3-1}$$

where $j = \sqrt{-1}$. It is a function of the real variable $-\infty < \omega < \infty$. If (3.3-1) is written in terms of the density function, $\Phi_X(\omega)$ is seen to be the *Fourier transform*† (with the sign of ω reversed) of $f_X(x)$:

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x) e^{j\omega x}\, dx \tag{3.3-2}$$

Because of this fact, if $\Phi_X(\omega)$ is known, $f_X(x)$ can be found from the *inverse Fourier transform* (with sign of x reversed)

†Readers unfamiliar with Fourier transforms should interpret $\Phi_X(\omega)$ as simply the expected value of the function $g(X) = \exp(j\omega X)$. Appendix D is included as a review for others wishing to refresh their background in Fourier transform theory.

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_X(\omega)e^{-j\omega x}\, d\omega \tag{3.3-3}$$

By formal differentiation of (3.3-2) n times with respect to ω and setting $\omega = 0$ in the derivative, we may show that the nth moment of X is given by

$$m_n = (-j)^n \frac{d^n\Phi_X(\omega)}{d\omega^n}\bigg|_{\omega=0} \tag{3.3-4}$$

A major advantage of using $\Phi_X(\omega)$ to find moments is that $\Phi_X(\omega)$ always exists (Davenport, 1970, p. 426), so the moments can always be found if $\Phi_X(\omega)$ is known, provided, of course, both the moments and the derivatives of $\Phi_X(\omega)$ exist.

It can be shown that the maximum magnitude of a characteristic function is unity and occurs at $\omega = 0$; that is,

$$|\Phi_X(\omega)| \le \Phi_X(0) = 1 \tag{3.3-5}$$

(See Problem 3.3-1.)

EXAMPLE 3.3-1. Again we consider the random variable with the exponential density of Example 3.1-2 and find its characteristic function and first moment. By substituting the density function into (3.3-2), we get

$$\Phi_X(\omega) = \int_a^{\infty} \frac{1}{b}e^{-(x-a)/b}e^{j\omega x}\, dx = \frac{e^{a/b}}{b}\int_a^{\infty} e^{-(1/b-j\omega)x}\, dx$$

Evaluation of the integral follows the use of an integral from Appendix C:

$$\Phi_X(\omega) = \frac{e^{a/b}}{b}\left[\frac{e^{-(1/b-j\omega)x}}{-(1/b-j\omega)}\bigg|_a^{\infty}\right]$$

$$= \frac{e^{j\omega a}}{1-j\omega b}$$

The derivative of $\Phi_X(\omega)$ is

$$\frac{d\Phi_X(\omega)}{d\omega} = e^{j\omega a}\left[\frac{ja}{1-j\omega b} + \frac{jb}{(1-j\omega b)^2}\right]$$

so the first moment becomes

$$m_1 = (-j)\frac{d\Phi_X(\omega)}{d\omega}\bigg|_{\omega=0} = a + b$$

in agreement with m_1 found in Example 3.1-2.

*Moment Generating Function

Another statistical average closely related to the characteristic function is the *moment generating function*, defined by

$$M_X(\nu) = E[e^{\nu X}] \tag{3.3-6}$$

where ν is a real number $-\infty < \nu < \infty$. Thus, $M_X(\nu)$ is given by

$$M_X(\nu) = \int_{-\infty}^{\infty} f_X(x) e^{\nu x}\,dx \tag{3.3-7}$$

The main advantage of the moment generating function derives from its ability to give the moments. Moments are related to $M_X(\nu)$ by the expression:

$$m_n = \left.\frac{d^n M_X(\nu)}{d\nu^n}\right|_{\nu=0} \tag{3.3-8}$$

The main disadvantage of the moment generating function, as opposed to the characteristic function, is that it may not exist for all random variables and all values of ν. However, if $M_X(\nu)$ exists for all values of ν in the neighborhood of $\nu = 0$ the moments are given by (3.3-8) (Wilks, 1962, p. 114).

EXAMPLE 3.3-2. To illustrate the calculation and use of the moment generating function, let us reconsider the exponential density of the earlier examples. On use of (3.3-7) we have

$$\begin{aligned} M_X(\nu) &= \int_a^{\infty} \frac{1}{b} e^{-(x-a)/b} e^{\nu x}\,dx \\ &= \frac{e^{a/b}}{b} \int_a^{\infty} e^{[\nu-(1/b)]x}\,dx \\ &= \frac{e^{a\nu}}{1-b\nu} \end{aligned}$$

In evaluating $M_X(\nu)$ we have used an integral from Appendix C.

By differentiation we have the first moment

$$\begin{aligned} m_1 &= \left.\frac{dM_X(\nu)}{d\nu}\right|_{\nu=0} \\ &= \left.\frac{e^{a\nu}[a(1-b\nu)+b]}{(1-b\nu)^2}\right|_{\nu=0} = a + b \end{aligned}$$

which, of course, is the same as previously found.

*Chernoff's Inequality and Bound

As another example of an application of the moment generating function, we develop Chernoff's inequality through an example.

EXAMPLE 3.3-3. Let X be any random variable, nonnegative or not. For any real $\nu > 0$ it is clear from some sketches that

$$\exp[\nu(x-a)] \geq u(x-a) \tag{1}$$

where $u(\cdot)$ is the unit-step function and a is an arbitrary real constant. Since

$$P\{X \geq a\} = \int_a^{\infty} f_X(x)\,dx = \int_{-\infty}^{\infty} f_X(x) u(x-a)\,dx \tag{2}$$

we have

$$P\{X \geq a\} \leq \int_{-\infty}^{\infty} f_X(x) e^{v(x-a)}\, dx = e^{-va} M_X(v) \tag{3}$$

from (1), (2), and (3.3-7). Equation (3) is called *Chernoff's inequality*. Because the right side is a function of parameter v, it can be minimized with respect to this parameter. The minimum value is called *Chernoff's bound* [Viniotis (1998), p. 144].

3.4 TRANSFORMATIONS OF A RANDOM VARIABLE

Quite often one may wish to transform (change) one random variable X into a new random variable Y by means of a transformation

$$Y = T(X) \tag{3.4-1}$$

Typically, the density function $f_X(x)$ or distribution function $F_X(x)$ of X is known, and the problem is to determine either the density function $f_Y(y)$ or distribution function $F_Y(y)$ of Y. The problem can be viewed as a "black box" with input X, output Y, and "transfer characteristic" $Y = T(X)$, as illustrated in Figure 3.4-1.

In general, X can be a discrete, continuous, or mixed random variable. In turn, the transformation T can be linear, nonlinear, segmented, staircase, etc. Clearly, there are many cases to consider in a general study, depending on the form of X and T. In this section we shall consider only three cases: (1) X continuous and T continuous and either monotonically increasing or decreasing with X; (2) X continuous and T continuous but nonmonotonic; (3) X discrete and T continuous. Note that the transformation in all three cases is assumed continuous. The concepts introduced in these three situations are broad enough that the reader should have no difficulty in extending them to other cases (see Problem 3.4-2).

Monotonic Transformations of a Continuous Random Variable

A transformation T is called *monotonically increasing* if $T(x_1) < T(x_2)$ for any $x_1 < x_2$. It is *monotonically decreasing* if $T(x_1) > T(x_2)$ for any $x_1 < x_2$.

Consider first the increasing transformation. We assume that T is continuous and differentiable at all values of x for which $f_X(x) \neq 0$. Let Y have a

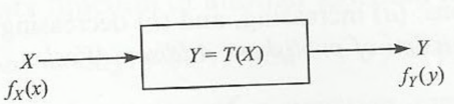


FIGURE 3.4-1
Transformation of a random variable X to a new random variable Y.

particular value y_0 corresponding to the particular value x_0 of X as shown in Figure 3.4-2a. The two numbers are related by

$$y_0 = T(x_0) \qquad \text{or} \qquad x_0 = T^{-1}(y_0) \tag{3.4-2}$$

where T^{-1} represents the inverse of the transformation T. Now the probability of the event $\{Y \le y_0\}$ must equal the probability of the event $\{X \le x_0\}$ because of the one-to-one correspondence between X and Y. Thus,

$$F_Y(y_0) = P\{Y \le y_0\} = P\{X \le x_0\} = F_X(x_0) \tag{3.4-3}$$

or

$$\int_{-\infty}^{y_0} f_Y(y)\,dy = \int_{-\infty}^{x_0=T^{-1}(y_0)} f_X(x)\,dx \tag{3.4-4}$$

Next, we differentiate both sides of (3.4-4) with respect to y_0 using Leibniz's rule† to get

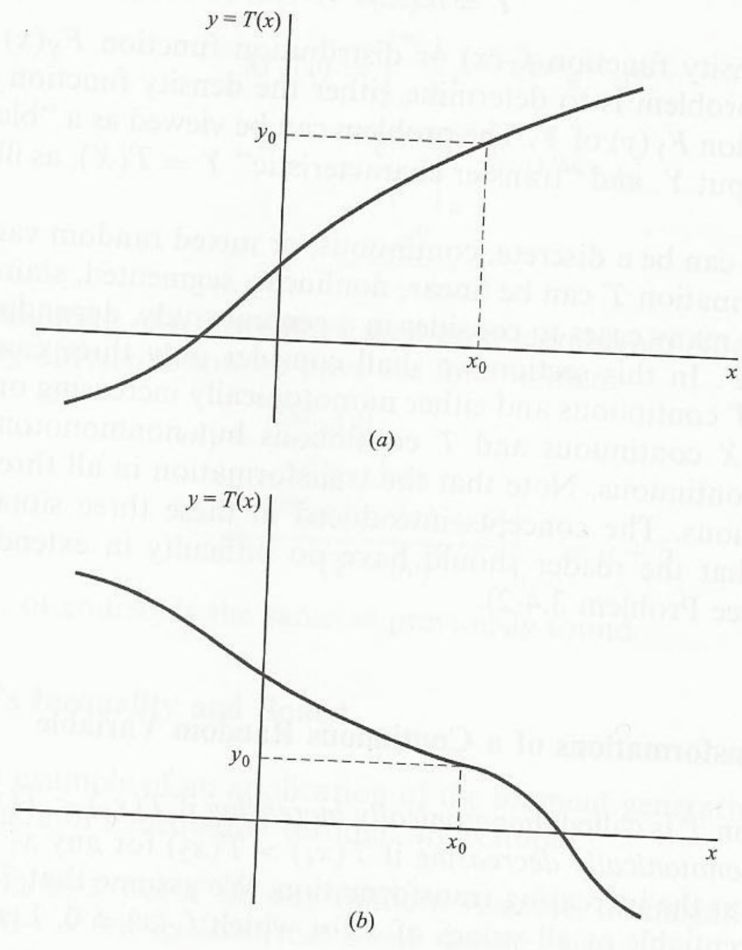


FIGURE 3.4-2
Monotonic transformations: (a) increasing, and (b) decreasing. [*Adapted from Peebles (1976) with permission of publishers Addison–Wesley, Advanced Book Program.*]

†Leibniz's rule is given in (G-2) of Appendix G.

$$f_Y(y_0) = f_X[T^{-1}(y_0)]\frac{dT^{-1}(y_0)}{dy_0} \tag{3.4-5}$$

Since this result applies for any y_0, we may now drop the subscript and write

$$f_Y(y) = f_X[T^{-1}(y)]\frac{dT^{-1}(y)}{dy} \tag{3.4-6}$$

or, more compactly,

$$f_Y(y) = f_X(x)\frac{dx}{dy} \tag{3.4-7}$$

In (3.4-7) it is understood that x is a function of y through (3.4-2).

A consideration of Figure 3.4-2b for the decreasing transformation verifies that

$$F_Y(y_0) = P\{Y \le y_0\} = P\{X \ge x_0\} = 1 - F_X(x_0) \tag{3.4-8}$$

A repetition of the steps leading to (3.4-6) will again produce (3.4-6) except that the right side is negative. However, since the slope of $T^{-1}(y)$ is also negative, we conclude that for either type of monotonic transformation

$$f_Y(y) = f_X[T^{-1}(y)]\left|\frac{dT^{-1}(y)}{dy}\right| \tag{3.4-9}$$

or simply

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| \tag{3.4-10}$$

EXAMPLE 3.4-1. If we take T to be the linear transformation $Y = T(X) = aX + b$, where a and b are any real constants, then $X = T^{-1}(Y) = (Y - b)/a$ and $dx/dy = 1/a$. From (3.4-9)

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right)\left|\frac{1}{a}\right|$$

If X is assumed to be gaussian with the density function given by (2.4-1), we get

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-[(y-b)/a - a_X]^2/2\sigma_X^2}\left|\frac{1}{a}\right| \\ &= \frac{1}{\sqrt{2\pi a^2\sigma_X^2}} e^{-[y-(aa_X+b)]^2/2a^2\sigma_X{}^2} \end{aligned}$$

which is the density function of another gaussian random variable having

$$a_Y = aa_X + b \qquad \text{and} \qquad \sigma_Y^2 = a^2\sigma_X^2$$

Thus, *a linear transformation of a gaussian random variable produces another gaussian random variable.* A linear amplifier having a random voltage X as its input is one example of a linear transformation.

Nonmonotonic Transformations of a Continuous Random Variable

A transformation may not be monotonic in the more general case. Figure 3.4-3 illustrates one such transformation. There may now be more than one interval of values of X that correspond to the event $\{Y \leq y_0\}$. For the value of y_0 shown in the figure, the event $\{Y \leq y_0\}$ corresponds to the event $\{X \leq x_1$ and $x_2 \leq X \leq x_3\}$. Thus, the probability of the event $\{Y \leq y_0\}$ now equals the probability of the event $\{x$ values yielding $Y \leq y_0\}$, which we shall write as $\{x | Y \leq y_0\}$. In other words

$$F_Y(y_0) = P\{Y \leq y_0\} = P\{x | Y \leq y_0\} = \int_{\{x|Y \leq y_0\}} f_X(x)\,dx \tag{3.4-11}$$

Formally, one may differentiate to obtain the density function of Y:

$$f_Y(y_0) = \frac{d}{dy_0} \int_{\{x|Y \leq y_0\}} f_X(x)\,dx \tag{3.4-12}$$

Although we shall not give a proof, the density function is also given by (Papoulis, 1965, p. 126)

$$f_Y(y) = \sum_n \frac{f_X(x_n)}{\left|\dfrac{dT(x)}{dx}\right|_{x=x_n}} \tag{3.4-13}$$

where the sum is taken so as to include all the roots x_n, $n = 1, 2, \ldots$, which are the real solutions of the equation†

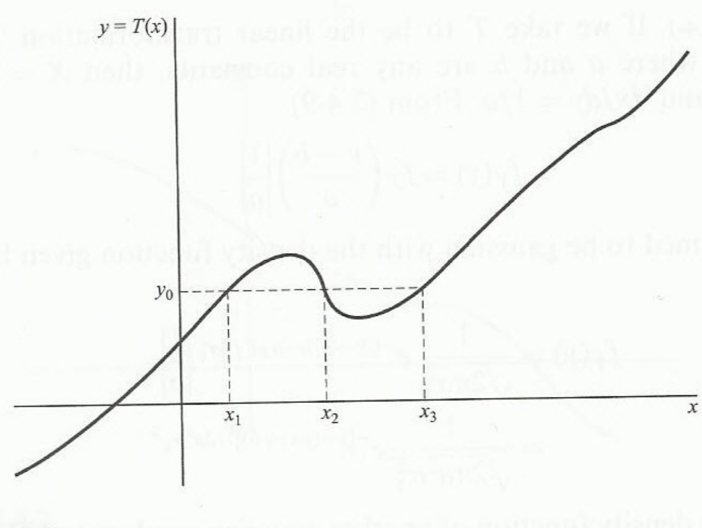


FIGURE 3.4-3
A nonmonotonic transformation. [*Adapted from Peebles (1976) with permission of publishers Addison–Wesley, Advanced Book Program.*]

†If $y = T(x)$ has no real roots for a given value of y, then $f_Y(y) = 0$.

$$y = T(x) \tag{3.4-14}$$

We illustrate the above concepts by an example.

EXAMPLE 3.4-2. We find $f_Y(y)$ for the square-law transformation

$$Y = T(X) = cX^2$$

shown in Figure 3.4-4, where c is a real constant $c > 0$. We shall use both the procedure leading to (3.4-12) and that leading to (3.4-13).

In the former case, the event $\{Y \leq y\}$ occurs when $\{-\sqrt{y/c} \leq x \leq \sqrt{y/c}\} = \{x | Y \leq y\}$, so (3.4-12) becomes

$$f_Y(y) = \frac{d}{dy} \int_{-\sqrt{y/c}}^{\sqrt{y/c}} f_X(x)\, dx \qquad y \geq 0$$

Upon use of Leibniz's rule we obtain

$$f_Y(y) = f_X\left(\sqrt{y/c}\right) \frac{d(\sqrt{y/c})}{dy} - f_X\left(-\sqrt{y/c}\right) \frac{d(-\sqrt{y/c})}{dy}$$

$$= \frac{f_X(\sqrt{y/c}) + f_X(-\sqrt{y/c})}{2\sqrt{cy}} \qquad y \geq 0$$

In the latter case where we use (3.4-13), we have $X = \pm\sqrt{Y/c}$, $Y \geq 0$, so $x_1 = -\sqrt{y/c}$ and $x_2 = \sqrt{y/c}$. Furthermore, $dT(x)/dx = 2cx$ so

$$\left.\frac{dT(x)}{dx}\right|_{x=x_1} = 2cx_1 = -2c\sqrt{\frac{y}{c}} = -2\sqrt{cy}$$

$$\left.\frac{dT(x)}{dx}\right|_{x=x_2} = 2\sqrt{cy}$$

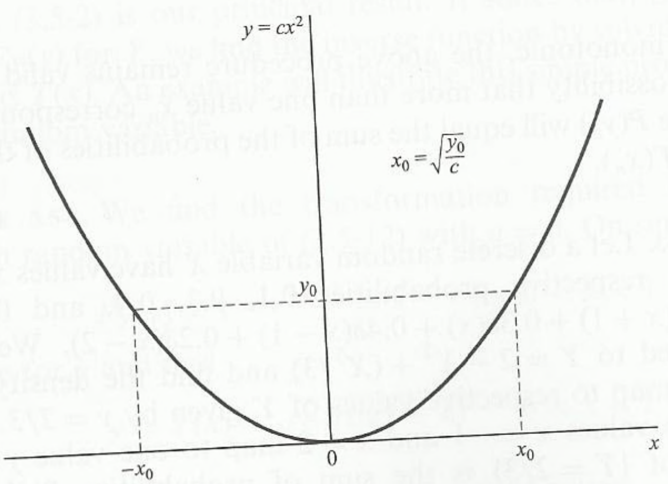


FIGURE 3.4-4
A square-law transformation. [*Adapted from Peebles (1976) with permission of publishers Addison–Wesley, Advanced Book Program.*]

From (3.4-13) we again have

$$f_Y(y) = \frac{f_X(\sqrt{y/c}) + f_X(-\sqrt{y/c})}{2\sqrt{cy}} \qquad y \geq 0$$

Transformation of a Discrete Random Variable

If X is a discrete random variable while $Y = T(X)$ is a continuous transformation, the problem is especially simple. Here

$$f_X(x) = \sum_n P(x_n)\delta(x - x_n) \tag{3.4-15}$$

$$F_X(x) = \sum_n P(x_n)u(x - x_n) \tag{3.4-16}$$

where the sum is taken to include all the possible values x_n, $n = 1, 2, \ldots$, of X.

If the transformation is monotonic, there is a one-to-one correspondence between X and Y so that a set $\{y_n\}$ corresponds to the set $\{x_n\}$ through the equation $y_n = T(x_n)$. The probability $P(y_n)$ equals $P(x_n)$. Thus,

$$f_Y(y) = \sum_n P(y_n)\delta(y - y_n) \tag{3.4-17}$$

$$F_Y(y) = \sum_n P(y_n)u(y - y_n) \tag{3.4-18}$$

where

$$y_n = T(x_n) \tag{3.4-19}$$

$$P(y_n) = P(x_n) \tag{3.4-20}$$

If T is not monotonic, the above procedure remains valid except there now exists the possibility that more than one value x_n corresponds to a value y_n. In such a case $P(y_n)$ will equal the sum of the probabilities of the various x_n for which $y_n = T(x_n)$.

EXAMPLE 3.4-3. Let a discrete random variable X have values $x = -1, 0, 1$, and 2 with respective probabilities 0.1, 0.3, 0.4, and 0.2, so that $f_X(x) = 0.1\delta(x+1) + 0.3\delta(x) + 0.4\delta(x-1) + 0.2\delta(x-2)$. We assume X is transformed to $Y = 2 - X^2 + (X^3/3)$ and find the density of Y. The values of X map to respective values of Y given by $y = 2/3$, 2, 4/3, and 2/3. The two values $x = -1$ and $x = 2$ map to one value $y = 2/3$. The probability of $\{Y = 2/3\}$ is the sum of probabilities $P\{X = -1\}$ and $P\{X = 2\}$, so

$$f_Y(y) = 0.3\delta[y - (2/3)] + 0.4\delta[y - (4/3)] + 0.3\delta(y - 2)$$

3.5 COMPUTER GENERATION OF ONE RANDOM VARIABLE

A digital computer is often used to simulate systems in order to estimate their performance with noise prior to the actual construction of the system. These simulations usually require that random numbers be generated that are values of random variables having prescribed distributions. If software (a computer program, or subroutine, that can be called up on demand) exists for the specified distribution, there is no problem. However, if the computer "library" does not contain the desired program, it is necessary for the simulation to generate its own random numbers. In this section we briefly describe how to generate a random variable with specified probability distribution, given mainly that the computer is able to generate random numbers that are values of a random variable with uniform distribution on (0,1), a commonly satisfied condition in most cases.

The problem, then, is to find the transformation $T(X)$ in Figure 3.4-1 that will create a random variable Y of prescribed distribution function when X has a uniform distribution on (0,1). We assume initially that $T(X)$ is a monotonically nondecreasing function so that (3.4-3) applies. Our work will show that this condition is automatically satisfied. From (3.4-3) we have (for any x and y)

$$F_Y[y = T(x)] = F_X(x) \tag{3.5-1}$$

But for uniform X, $F_X(x) = x$ when $0 < x < 1$, from (2.5-8). Thus, we solve for the inverse in (3.5-1) to get

$$y = T(x) = F_Y^{-1}(x) \qquad 0 < x < 1 \tag{3.5-2}$$

Since any distribution function $F_Y(y)$ is nondecreasing, its inverse is nondecreasing, and the initial assumption is always satisfied.

Equation (3.5-2) is our principal result. It states that, given a specified distribution $F_Y(y)$ for Y, we find the inverse function by solving $F_Y(y) = x$ for y. The result is $T(x)$. An example will illustrate this simple procedure to create a Rayleigh random variable.

EXAMPLE 3.5-1. We find the transformation required to generate the Rayleigh random variable of (2.5-12) with $a = 0$. On setting

$$F_Y(y) = 1 - e^{-y^2/b} = x \qquad \text{for } 0 < x < 1$$

we solve for y and find

$$y = T(x) = \sqrt{-b\ln(1-x)} \qquad 0 < x < 1$$

We give another example for the arcsine distribution.

EXAMPLE 3.5-2. A computer software program can generate random numbers on request that are values of a random variable uniformly distributed on (0,1). We find the required transformation to convert these values to those of a random variable with the arcsine distribution of (F-23) in Appendix F. From (F-23) and (3.5-1) we require

$$F_Y(y) = \begin{cases} 0, & y \le -a \\ 0.5 + \dfrac{1}{\pi}\sin^{-1}(y/a), & -a < y < a \\ 1, & a \le y \end{cases} = \begin{cases} 0, & 0 < x < 0 \\ x, & 0 < x < 1 \\ 1, & 1 \le x \end{cases}$$

The solution for y is direct. We obtain

$$y = a\sin[\pi(2x - 1)/2] \qquad 0 < x < 1$$

Equation (3.5-2) can be readily applied to any distribution for which its inverse can be analytically determined (see Problems 3.5-1 through 3.5-3 for other examples). For other distributions the required inverse can be stored in the computer for a number of points (y, x), and the simulation can then use interpolation between computed points to obtain values of y for any value of x.

The gaussian random variable is an important example of a distribution for which the inverse cannot be found analytically. Because computer simulations often require gaussian random numbers to be generated, we show in Section 5.6 how this important problem can be solved by extension of the methods of this section.

EXAMPLE 3.5-3. We use MATLAB®† to generate a sequence of 100 random numbers x_i that correspond to a random variable uniformly distributed on (0,1). The results of Example 3.5-1 are then used to convert the random numbers to a Rayleigh random variable by means of the transformation $y_i = [-\ln(1 - x_i)]^{1/2}$. We generate a *histogram* of the uniform values by classifying them into 10 "bins" of width 0.1. A histogram is a plot of the number of values falling in a bin divided by the total number of values generated. Such histograms approximate the probability density of the random variable for which the values apply. For the Rayleigh values a histogram using 22 bins of width 0.1 from 0 to 2.2 is created. Finally, to show that the histograms equal, approximately, the probability densities of the random variables, the true densities are also plotted for comparison.

The applicable MATLAB code is shown in Figure 3.5-1. Results for the uniform random variable are shown in Figure 3.5-2, while the plots for the Rayleigh case are shown in Figure 3.5-3. If more than 100 values of y_i were used, the histograms would tend to approximate the true densities more closely.

†MATLAB is a registered trademark of The Mathworks, Inc., 3 Apple Hill Drive, Natick, MA 01960-2098, USA.

```
%%%%%%%%%%%%%% Example 3.5-3 %%%%%%%%%%%%%%%%%

clear

N = 100; % number of random variables to generate
stp = 0.1; % step size
b = 1; % Rayleigh parameter

x = rand(1,N); % uniformly distributed random numbers
y = sqrt(-b*log(1-x)); % Rayleigh distributed random numbers

f = find(y > 2.2); % find values out of the range of interest
y(f) = []; % remove these values

xcenter = [0.05:stp:1]; % centers of the bins for the histogram
ycenter = [0.05:stp:2.2];

xabscissa = 0:stp:1; % abscissa used for analytic results
yabscissa = 0:stp:2.5;

xhist = hist(x,xcenter); % compute histograms (not normalized)
yhist = hist(y,ycenter);

xtrue = ones(size(xabscissa)); % compute the analytic values
ytrue = 2*yabscissa/b.*exp(-yabscissa.^2/b);

% plot results
clf
bar(xcenter,xhist./(N*stp),1,'w') % plot normalized histogram
hold on
plot(xabscissa,xtrue,'k') % uniform distribution

xlabel('Magnitude Bins')
ylabel('Relative Number of Samples')
title('Histogram of Uniform Distribution')

figure
bar(ycenter,yhist./(N*stp),1,'w')
hold on
plot(yabscissa,ytrue,'k') % Rayleigh distribution

xlabel('Magnitude Bins')
ylabel ('Relative Number of Samples')
title('Histogram of Rayleigh Distribution')
```

FIGURE 3.5-1
MATLAB code used in Example 3.5-3.

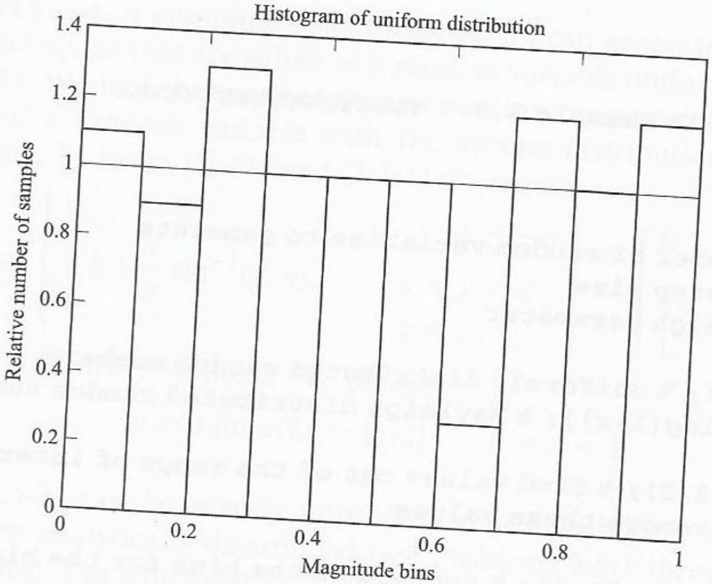


FIGURE 3.5-2
Histogram and true density function for the uniform random variable of Example 3.5-3.

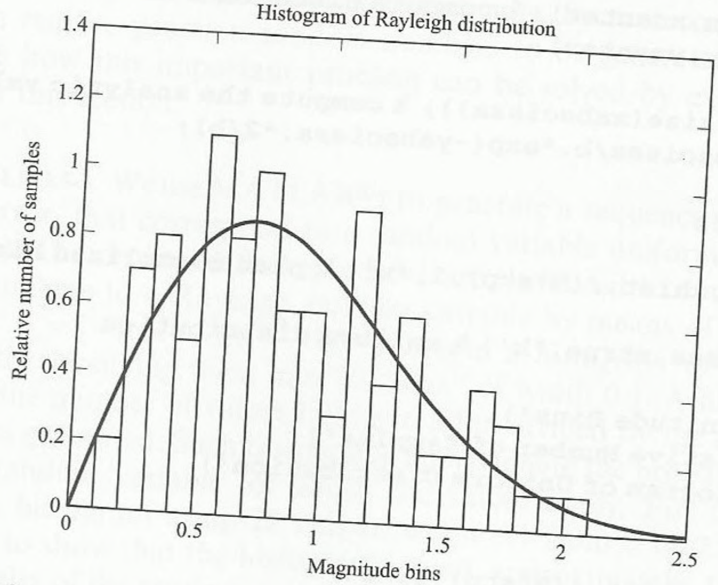


FIGURE 3.5-3
Histogram and true density function for the Rayleigh random variable of Example 3.5-3.

3.6 SUMMARY

Although the preceding chapter discussed the random variable and how to calculate probabilities of its having particular values, these topics are not sufficient for many practical applications. One must also be able to determine useful *characteristics* of random variables, and know how to convert (transform) one random variable into another. These were the subjects of this chapter. All these subjects were developed from a single concept: *expectation*, which is nothing more than an averaging procedure for random quantities. Items covered were:

- Expectation was defined in general for a random variable or some function of a random variable.
- Moments (about the origin, and central) were developed as valuable measures of a random variable's characteristics. Of particular value were mean value, variance, and skew.
- The characteristic function and moment generating function were given as convenient methods for finding moments.
- Methods were given to transform one random variable into another, and to find distribution and density functions of the new random variable.
- The important concepts of how to generate a specified random variable by computer were developed and an example and chapter-end problems were included that are based on use of MATLAB software.

PROBLEMS

3.1-1. A discrete random variable X has possible values $x_i = i^2$, $i = 1, 2, 3, 4, 5$, which occur with probabilities 0.4, 0.25, 0.15, 0.1, and 0.1, respectively. Find the mean value $\bar{X} = E[X]$ of X.

3.1-2. The natural numbers are the possible values of a random variable X: that is, $x_n = n$, $n = 1, 2, \ldots$. These numbers occur with probabilities $P(x_n) = (\frac{1}{2})^n$. Find the expected value of X.

3.1-3. If the probabilities in Problem 3.1-2 are $P(x_n) = p^n$, $0 < p < 1$, show that $p = \frac{1}{2}$ is the only value of p that is allowed for the problem as formulated. (*Hint*: Use the fact that $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ is necessary.)

3.1-4. (*a*) Find the average amount the gambler in Problem 2.1-13 can expect to win.
(*b*) What is his probability of winning on any given playing of the game?

3.1-5. The *arcsine* probability density is defined by

$$f_X(x) = \frac{\text{rect}(x/2a)}{\pi\sqrt{a^2 - x^2}}$$

for any real constant $a > 0$. Show that $\bar{X} = 0$ and $\overline{X^2} = a^2/2$ for this density.

***3.1-6.** For the animal described in Problem 2.6-4 find its expected lifetime given that it will not live beyond 20 weeks.

3.1-7. Find the expected value of the function $g(X) = X^3$ where X is a random variable defined by the density

$$f_X(x) = (\tfrac{1}{2})u(x)\exp(-x/2)$$

3.1-8. A random variable X represents the value of coins (in cents) given in change when purchases are made at a particular store. Suppose the probabilities of 1¢, 5¢, 10¢, 25¢, and 50¢ being present in change are 0.35, 0.25, 0.20, 0.15, and 0.05, respectively.
(*a*) Write an expression for the probability density function of X.
(*b*) Find the mean of X.

3.1-9. In the circuit of Figure P2.1-11 of Chapter 2 let the resistance of R_1 be a random variable uniformly distributed on $(R_0 - \Delta R, R_0 + \Delta R)$ where R_0 and ΔR are constants.
(*a*) Find an expression for the power dissipated in R_2 for any constant voltage E_1.
(*b*) Find the mean value of power when R_1 is random.
(*c*) Evaluate the mean power for $E_1 = 12\,\text{V}$, $R_2 = 1000\,\Omega$, $R_0 = 1500\,\Omega$, and $\Delta R = 100\,\Omega$.

***3.1-10.** The power (in milliwatts) returned to a radar from a certain class of aircraft has the probability density function

$$f_P(p) = \frac{1}{10} e^{-p/10} u(p)$$

Suppose a given aircraft belongs to this class but is known to not produce a power larger than 15 mW.
(*a*) Find the probability density function of P conditional on $P \leq 15\,\text{mW}$.
(*b*) Find the conditional mean value of P.

3.1-11. A random variable X has a probability density

$$f_X(x) = \begin{cases} (1/2)\cos(x) & -\pi/2 < x < \pi/2 \\ 0 & \text{elsewhere in } x \end{cases}$$

Find the mean value of the function $g(X) = 4X^2$.

3.1-12. Work Problem 3.1-11, except assume a function $g(X) = 4X^4$.

3.1-13. An information source can emit (generate) any one of 128 levels where each is equally probable and independent of all others. What average information does the source represent? (*Hint*: Use the results of Example 3.1-4.)

3.1-14. A random variable X is uniformly distributed on the interval $(-5, 15)$. Another random variable $Y = e^{-X/5}$ is formed. Find $E[Y]$.

3.1-15. A gaussian voltage random variable X [see (2.4-1)] has a mean value $\bar{X} = a_X = 0$ and variance $\sigma_X^2 = 9$. The voltage X is applied to a square-law, full-wave diode detector with a transfer characteristic $Y = 5X^2$. Find the mean value of the output voltage Y.

***3.1-16.** For the system having a lifetime specified in Problem 2.5-3 of Chapter 2, determine the expected lifetime of the system given that the system has survived 20 weeks.

3.2-1. Give an example of a random variable where its mean value might not equal any of its possible values.

3.2-2. Find:
(*a*) the expected value, and
(*b*) the variance of the random variable with the triangular density of Figure 2.3-1*a* if $a = 1/\alpha$.

3.2-3. Show that the mean value and variance of the random variable having the uniform density function of (2.5-7) are:

$$\bar{X} = E[X] = (a+b)/2$$

and

$$\sigma_X^2 = (b-a)^2/12$$

3.2-4. A pointer is spun on a fair wheel of chance numbered from 0 to 100 around its circumference.
(*a*) What is the average value of all possible pointer positions?
(*b*) What deviation from its average value will the pointer position take on the average; that is, what is the pointer's root-mean-squared deviation from its mean? (*Hint*: Use results of Problem 3.2-3.)

3.2-5. Find:
(*a*) the mean value, and
(*b*) the variance of the random variable X defined by Problems 2.1-6 and 2.2-3 of Chapter 2.

***3.2-6.** For the *binomial density* of (2.5-1), show that

$$E[X] = \bar{X} = Np$$

and

$$\sigma_X^2 = Np(1-p)$$

3.2-7. (*a*) Let resistance be a random variable in Problem 2.1-11 of Chapter 2. Find the mean value of resistance.
(*b*) What is the output voltage E_2 if an *average* resistor were used in the circuit?
(*c*) For the resistors specified, what is the mean value of E_2? Does the voltage of part (*b*) equal this value? Explain your results.

3.2-8. (*a*) Use the symmetry of the density function given by (2.4-1) to justify that the parameter a_X in the *gaussian density* is the mean value of the random variable: $\bar{X} = a_X$.
(*b*) Prove that the parameter σ_X^2 is the variance. (*Hint*: Use an equation from Appendix C.)

3.2-9. Show that the mean value $E[X]$ and variance σ_X^2 of the Rayleigh random variable, with density given by (2.5-11), are

$$E[X] = a + \sqrt{\pi b/4}$$

and

$$\sigma_X^2 = b(4-\pi)/4$$

3.2-10. What is the expected lifetime of the system defined in Problem 2.5-3 of Chapter 2?

3.2-11. Find:
(*a*) the mean value, and
(*b*) the variance for a random variable with the *Laplace* density

$$f_X(x) = \frac{1}{2b} e^{-|x-m|/b}$$

where b and m are real constants, $b > 0$, and $-\infty < m < \infty$.

3.2-12. Determine the mean value of the *Cauchy* random variable in Problem 2.5-4 of Chapter 2. What can you say about the variance of this random variable?

***3.2-13.** For the *Poisson* random variable defined in (2.5-4) show that:
(*a*) the mean value is b and
(*b*) the variance also equals b.

3.2-14. (*a*) Use (3.2-2) to find the first three moments m_1, m_2, and m_3 for the exponential density of Example 3.1-2.
(*b*) Find m_1, m_2, and m_3 from the characteristic function found in Example 3.3-1. Verify that they agree with those of part (*a*).

3.2-15. Find the expressions for all the moments about the origin and central moments for the uniform density of (2.5-7).

3.2-16. Define a function $g(\cdot)$ of a random variable X by

$$g(X) = \begin{cases} 1 & x \geq x_0 \\ 0 & x < x_0 \end{cases}$$

where x_0 is a real number $-\infty < x_0 < \infty$. Show that

$$E[g(X)] = 1 - F_X(x_0)$$

3.2-17. Show that the second moment of any random variable X about an arbitrary point a is minimum when $a = \bar{X}$; that is, show that $E[(X-a)^2]$ is minimum for $a = \bar{X}$.

3.2-18. For any discrete random variable X with values x_i having probabilities of occurrence $P(x_i)$, show that the moments of X are

$$m_n = \sum_{i=1}^{N} x_i^n P(x_i)$$

$$\mu_n = \sum_{i=1}^{N} (x_i - \bar{X})^n P(x_i)$$

where N may be infinite for some X.

3.2-19. Prove that central moments μ_n are related to moments m_k about the origin by

$$\mu_n = \sum_{k=0}^{n} \binom{n}{k} (-\bar{X})^{n-k} m_k$$

3.2-20. A random variable X has a density function $f_X(x)$ and moments m_n. If the density is shifted higher in x by an amount $\alpha > 0$ to a new origin, show that the moments of the shifted density, denoted m_n', are related to the moments m_n by

$$m_n' = \sum_{k=0}^{n} \binom{n}{k} \alpha^{n-k} m_k$$

3.2-21. Continue Problem 3.1–14 by finding all moments of Y. (*Hint*: Treat Y^n as a function of Y, not as a transformation.)

3.2-22. Reconsider the production line that manufactures bolts in Problem 2.2-1.
(*a*) What is the average length of bolts that are placed up for sale?
(*b*) What is the standard deviation of length of bolts sold?
(*c*) What percentage of all bolts sold are expected to have a length within one standard deviation of the average length?
(*d*) By what tolerance (as a percentage) does the average length of bolts sold match the nominally desired length of 760 mm?

3.2-23. A random variable X has a probability density

$$f_X(x) = \begin{cases} (\pi/16)\cos(\pi x/8) & -4 \le x \le 4 \\ 0 & \text{elsewhere} \end{cases}$$

Find: (*a*) its mean value $\bar{X}$, (*b*) its second moment $\overline{X^2}$, and (*c*) its variance.

3.2-24. A certain meter is designed to measure small dc voltages but makes errors because of noise. The errors are accurately represented as a gaussian random variable with a mean of zero and a standard deviation of 10^{-3} V. When the dc voltage is disconnected it is found that the probability is 0.5 that the meter reading is positive due to noise. With the dc voltage present, this probability becomes 0.2514. What is the dc voltage?

3.2-25. Find the skew and coefficient of skewness for a Rayleigh random variable for which $a = 0$ in (2.5-11).

3.2-26. A random variable X has the density

$$f_X(x) = \begin{cases} (\frac{3}{32})(-x^2 + 8x - 12) & 2 \le x \le 6 \\ 0 & \text{elsewhere} \end{cases}$$

Find the following moments: (*a*) m_0, (*b*) m_1, (*c*) m_2, and (*d*) μ_2.

3.2-27. The *chi-square density* with N degrees of freedom is defined by

$$f_X(x) = \frac{x^{(N/2)-1}}{2^{N/2}\Gamma(N/2)} u(x) e^{-x/2}$$

where $\Gamma(\cdot)$ is the gamma function

$$\Gamma(z) = \int_0^\infty \xi^{z-1} e^{-\xi}\, d\xi \qquad \text{real part of } z > 0$$

and $N = 1, 2, \ldots$. Show that (*a*) $\bar{X} = N$, (*b*) $\overline{X^2} = N(N+2)$, and (*c*) $\sigma_X^2 = 2N$ for this density.

3.2-28. For the density of Problem 3.2-27 find its arbitrary moment $\overline{X^n}$, $n = 0, 1, 2, \ldots$.

3.2-29. A random variable X is called *Weibull* if its density has the form

$$f_X(x) = abx^{b-1}\exp(-ax^b)u(x)$$

where $a > 0$ and $b > 0$ are real constants. Use the definition of the gamma function of Problem 3.2-27 to find (*a*) the mean value, (*b*) the second moment, and (*c*) the variance of X.

***3.2-30.** Show that the characteristic function of a random variable having the binomial density of (2.5-1) is

$$\Phi_X(\omega) = [1 - p + pe^{j\omega}]^N$$

***3.2-31.** Show that the characteristic function of a Poisson random variable defined by (2.5-4) is

$$\Phi_X(\omega) = \exp[-b(1 - e^{j\omega})]$$

***3.2-32.** The *Erlang*† *random variable* X has a characteristic function

$$\Phi_X(\omega) = \left[\frac{a}{a - j\omega}\right]^N$$

for $a > 0$ and $N = 1, 2, \ldots$. Show that $\bar{X} = N/a$, $\overline{X^2} = N(N+1)/a^2$, and $\sigma_X^2 = N/a^2$.

3.2-33. A random variable X has $\bar{X} = -3$, $\overline{X^2} = 11$, and $\sigma_X^2 = 2$. For a new random variable $Y = 2X - 3$, find (*a*) $\bar{Y}$, (*b*) $\overline{Y^2}$, and (*c*) σ_Y^2.

3.2-34. A random variable has a probability density

$$f_X(x) = \begin{cases} (5/4)(1 - x^4) & 0 < x \le 1 \\ 0 & \text{elsewhere in } x \end{cases}$$

Find: (*a*) $E[X]$, (*b*) $E[4X + 2]$, and (*c*) $E[X^2]$.

3.2-35. Use the definition of the gamma function as given by (F–1*f*) in Appendix F to obtain an expression for the moments $E[X^n]$, $n = 0, 1, 2, \ldots$, for the gamma density defined by (F–50). Use the expression to prove that (F-52) and (F-53) are true.

3.2-36. Suppose it is found that the function

$$f_X(x) = \frac{16/\pi}{(4 + x^2)^2}$$

is a good empirical fit to the probability density function of some random experimental data represented by a random variable X. Find the mean, second moment, and variance of X.

†A. K. Erlang (1878–1929) was a Danish engineer.

***3.3-1.** Show that any characteristic function $\Phi_X(\omega)$ satisfies

$$|\Phi_X(\omega)| \le \Phi_X(0) = 1$$

***3.3-2.** The characteristic function for a gaussian random variable X, having a mean value of 0, is

$$\Phi_X(\omega) = \exp(-\sigma_X^2 \omega^2/2)$$

Find all the moments of X using $\Phi_X(\omega)$.

***3.3-3.** Work Problem 3.3-2 using the moment generating function

$$M_X(v) = \exp(\sigma_X^2 v^2/2)$$

for the zero-mean gaussian random variable.

***3.3-4.** A discrete random variable X can have $N+1$ values $x_k = k\Delta$, $k = 0, 1, \ldots, N$, where $\Delta > 0$ is a real number. Its values occur with equal probability. Show that the characteristic function of X is

$$\Phi_X(\omega) = \frac{1}{N+1}\,\frac{\sin[(N+1)\omega\Delta/2]}{\sin(\omega\Delta/2)}\,e^{jN\omega\Delta/2}$$

***3.3-5.** The characteristic function of the Laplace density of Problem 3.2-11 is known to be

$$\Phi_X(\omega) = \frac{e^{jm\omega}}{1 + (b\omega)^2}$$

Use this result to find the mean, second moment, and variance of the random variable X.

***3.3-6.** The chi-square density of Problem 3.2-27 has a characteristic function

$$\Phi_X(\omega) = \frac{1}{(1 - j2\omega)^{N/2}}$$

Use this function with (3.3-4) to verify the mean and a second moment found in Problem 3.2-27.

***3.3-7.** Solve for the Chernoff bound for a gaussian random variable with zero mean and variance one. [*Hint*: First find $M(v)$ by use of (C-51).]

3.4-1. A random variable X is uniformly distributed on the interval $(-\pi/2, \pi/2)$. X is transformed to the new random variable $Y = T(X) = a\tan(X)$, where $a > 0$. Find the probability density function of Y.

3.4-2. Work Problem 3.4-1 if X is uniform on the interval $(-\pi, \pi)$.

3.4-3. A random variable X undergoes the transformation $Y = a/X$, where a is a real number. Find the density function of Y.

3.4-4. A random variable X is uniformly distributed on the interval $(-a, a)$. It is transformed to a new variable Y by the transformation $Y = cX^2$ defined in Example 3.4-2. Find and sketch the density function of Y.

3.4-5. A zero-mean gaussian random variable X is transformed to the random variable Y determined by

$$Y = \begin{cases} cX & X > 0 \\ 0 & X \leq 0 \end{cases}$$

where c is a real constant, $c > 0$. Find and sketch the density function of Y.

3.4-6. If the transformation of Problem 3.4-5 is applied to a Rayleigh random variable with $a \geq 0$, what is its effect?

***3.4-7.** A random variable Θ is uniformly distributed on the interval (θ_1, θ_2) where θ_1 and θ_2 are real and satisfy

$$0 \leq \theta_1 < \theta_2 < \pi$$

Find and sketch the probability density function of the transformed random variable $Y = \cos(\Theta)$.

3.4-8. A random variable X can have values -4, -1, 2, 3, and 4, each with probability $\frac{1}{5}$. Find:
(*a*) the density function,
(*b*) the mean, and
(*c*) the variance of the random $Y = 3X^3$.

3.4-9. A gaussian random variable, for which

$$f_X(x) = \left(2/\sqrt{\pi}\right)\exp(-4x^2)$$

is applied to a square-law device to produce a new (output) random variable $Y = X^2/2$. (*a*) Find the density of Y. (*b*) Find the moments $m_n = E[Y^n]$, $n = 0, 1, \ldots$. (*Hint*: Put your answer in terms of the gamma function defined in Problem 3.2-27.)

3.4-10. A gaussian random variable, for which $\bar{X} = 0.6$ and $\sigma_X = 0.8$, is transformed to a new random variable by the transformation

$$Y = T(X) = \begin{cases} 4 & 1.0 \leq X < \infty \\ 2 & 0 \leq X < 1.0 \\ -2 & -1.0 \leq X < 0 \\ -4 & -\infty < X < -1.0 \end{cases}$$

(*a*) Find the density function of Y.
(*b*) Find the mean and variance of Y.

3.4-11. Work Problem 3.4-1 except assume a transformation $Y = T(X) = a\sin(X)$ with $a > 0$.

3.4-12. Let X be a gaussian random variable with density given by (2.4-1). If X is transformed to a new random variable $Y = b + e^X$, where b is a real constant, show that the density of Y is log-normal as defined in Problem 2.5-5. This transformation allows log-normal random numbers to be generated from gaussian random numbers by a digital computer.

3.4-13. A random variable X is uniformly distributed on (0, 6). If X is transformed to a new random variable $Y = 2(X-3)^2 - 4$, find: (*a*) the density of Y, (*b*) $\bar{Y}$, (*c*) σ_Y^2.

3.4-14. It is known that the envelope of the bandpass noise that emerges from a communication or radar receiver can be modeled as a Rayleigh random variable X with the probability density of (2.5-11) when $a = 0$, $b = 2\sigma_X^2$, and σ_X^2 is the power in the bandpass noise. If the envelope is transformed to a new variable $Y = cX^2$, where c is a constant, find the density of Y. This transformation is equivalent to a diode envelope detector where the noise level is small and the diode behaves approximately as a square-law device.

3.4-15. A certain "soft" limiter accepts a random input voltage X and limits the amplitudes of an output random variable Y according to

$$Y = \begin{cases} V(1-e^{-X/a}) & 0 \le X \\ -V(1-e^{X/a}) & X \le 0 \end{cases}$$

where $V > 0$ and $a > 0$ are constants. Show that the probability density of Y is

$$f_Y(y) = \frac{a}{(V-y)} f_X\left[a \ln\left(\frac{V}{(V-y)}\right)\right] u(y) + \frac{a}{(V+y)} f_X\left[-a \ln\left(\frac{V}{(V+y)}\right)\right] u(-y)$$

where $f_X(x)$ is the probability density of X.

3.4-16. If X in Problem 3.4-15 has the Laplace density of Problem 3.2-11 with $m = 0$, find the density of the output Y. If $a = b$, how is Y distributed?

3.5-1. In a computer simulation it is desired to transform numbers that are values of a random variable uniformly distributed on (0, 1) to numbers that are values of an exponentially distributed random variable, as defined by (2.5-10) with $a = 0$. Find the required transformation.

3.5-2. Work Problem 3.5-1, except to generate a random variable with a *Weibull* distribution as defined by (F-91) in Appendix F.

3.5-3. Work Problem 3.5-1, except to generate a *Cauchy* random variable as defined by (F-31) of Appendix F with $a = 0$.

3.5-4. A random variable Y has the probability density function

$$f_Y(y) = \frac{4a^4 y u(y)}{(a^2 + y^2)^3}$$

where a is a real positive constant and $u(y)$ is the unit-step function of (A-5). Find the mean value, second moment, variance, and cumulative distribution function of Y. Show that the transformation needed to generate Y from a random variable X that is uniformly distributed on (0, 1) is

$$Y = T(X) = a\left\{\left[\frac{1}{1-X}\right]^{1/2} - 1\right\}^{1/2} \qquad 0 < X < 1$$

3.5-5. Extend Example 3.5-3 by repeating the procedures for 1000 random numbers. Compare the accuracy of the results with those of the example.

3.5-6. Work Example 3.5-3 except generate the histogram of a random variable with arcsine distribution (as defined in Example 3.5-2) using 1000 random numbers.

CHAPTER 4

Multiple Random Variables

4.0 INTRODUCTION

In Chapters 2 and 3, various aspects of the theory of a single random variable were studied. The random variable was found to be a powerful concept. It enabled many realistic problems to be described in a probabilistic way such that practical measures could be applied to the problem even though it was random. For example, we have seen that shell impact position along the line of fire from a cannon to a target can be described by a random variable (Problem 2.4-6). From knowledge of the probability distribution or density function of impact position, we can solve for such practical measures as the mean value of impact position, its variance, and skew. These measures are not, however, a complete enough description of the problem in most cases.

Naturally, we may also be interested in how much the impact positions deviate *from* the line of fire in, say, the perpendicular (cross-fire) direction. In other words, we prefer to describe impact position as a point in a plane as opposed to being a point along a line. To handle such situations it is necessary that we extend our theory to include *two* random variables, one for each coordinate axis of the plane in our example. In other problems it may be necessary to extend the theory to include *several* random variables. We accomplish these extensions in this and the next chapter.

Fortunately, many situations of interest in engineering can be handled by the theory of two random variables.† Because of this fact, we emphasize the two-variable case, although the more general theory is also stated in most discussions to follow.

†In particular, it will be found in Chapter 6 that such important concepts as autocorrelation, cross-correlation, and covariance functions, which apply to random processes, are based on two random variables.

4.1 VECTOR RANDOM VARIABLES

Suppose two random variables X and Y are defined on a sample space S, where specific values of X and Y are denoted by x and y, respectively. Then any ordered pair of numbers (x, y) may be conveniently considered to be a *random point* in the xy plane. The point may be taken as a specific value of a *vector random variable* or a *random vector*.† Figure 4.1-1 illustrates the mapping involved in going from S to the xy plane.

The plane of all points (x, y) in the ranges of X and Y may be considered a new sample space. It is in reality a vector space where the components of any vector are the values of the random variables X and Y. The new space is sometimes called the *range sample space* (Davenport, 1970) or the *two-dimensional product space*. In this section and all following work we shall just call it a *joint sample space* and give it the symbol S_J.

As in the case of one random variable, let us define an event A by

$$A = \{X \leq x\} \tag{4.1-1}$$

A similar event B can be defined for Y:

$$B = \{Y \leq y\} \tag{4.1-2}$$

Events A and B refer to the sample space S, while events $\{X \leq x\}$ and $\{Y \leq y\}$ refer to the joint sample space S_J.‡ Figure 4.1-2 illustrates the correspon-

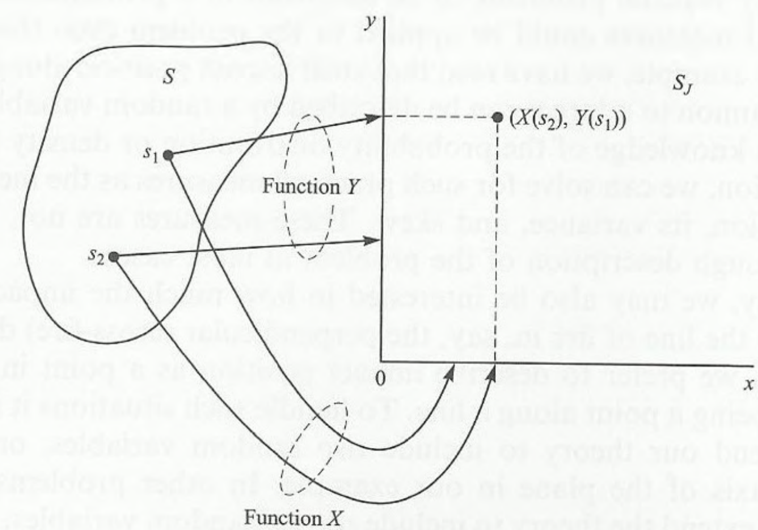


FIGURE 4.1-1
Mapping from the sample space S to the joint sample space S_J (xy plane).

---

†There are some specific conditions that must be satisfied in a complete definition of a random vector (Davenport, 1970, Chapter 5). They are somewhat advanced for our scope and we shall simply assume the validity of our random vectors.

‡Do not forget that elements s of S form the link between the two events since by writing $\{X \leq x\}$ we really refer to the set of those s such that $X(s) \leq x$ for some real number x. A similar statement holds for the event $\{Y \leq y\}$.

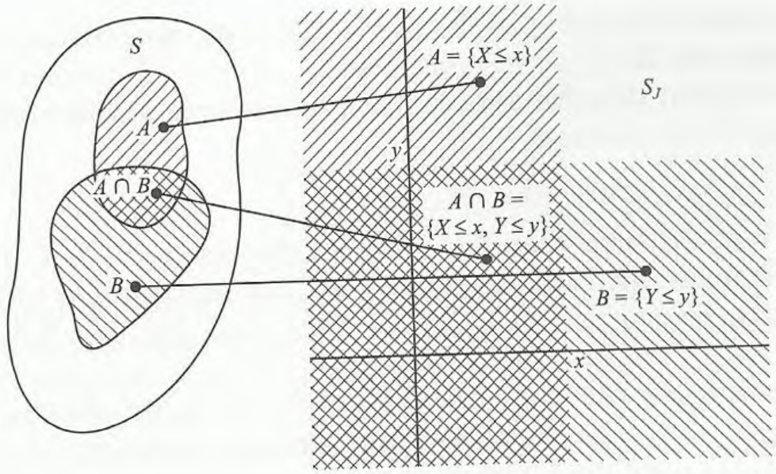


FIGURE 4.1-2
Comparisons of events in S with those in S_J.

dences between events in the two spaces. Event A corresponds to all points in S_J for which the X coordinate values are not greater than x. Similarly, event B corresponds to the Y coordinate values in S_J not exceeding y. Of special interest is to observe that the event $A \cap B$ defined on S corresponds to the *joint event* $\{X \leq x$ and $Y \leq y\}$ defined on S_J, which we write $\{X \leq x, Y \leq y\}$. This joint event is shown crosshatched in figure 4.1-2.

In the more general case where N random variables $X_1, X_2, \ldots, X_N$ are defined on a sample space S, we consider them to be components of an N-*dimensional random vector* or *N-dimensional random variable*. The joint sample space S_J is now N-dimensional.

4.2 JOINT DISTRIBUTION AND ITS PROPERTIES

The probabilities of the two events $A = \{X \leq x\}$ and $B = \{Y \leq y\}$ have already been defined as functions of x and y, respectively, called probability distribution functions:

$$F_X(x) = P\{X \leq x\} \tag{4.2-1}$$

$$F_Y(y) = P\{Y \leq y\} \tag{4.2-2}$$

We must introduce a new concept to include the probability of the joint event $\{X \leq x, Y \leq y\}$.

Joint Distribution Function

We define the probability of the joint event $\{X \leq x, Y \leq y\}$, which is a function of the numbers x and y, by a *joint probability distribution function* and denote it by the symbol $F_{X,Y}(x, y)$. Hence,

$$F_{X,Y}(x, y) = P\{X \leq x, Y \leq y\} \tag{4.2-3}$$

It should be clear that $P\{X \le x, Y \le y\} = P(A \cap B)$, where the joint event $A \cap B$ is defined on S.

To illustrate joint distribution, we take an example where both random variables X and Y are discrete.

EXAMPLE 4.2-1. Assume that the joint sample space S_J has only three possible elements: (1, 1), (2, 1), and (3, 3). The probabilities of these elements are to be $P(1, 1) = 0.2$, $P(2, 1) = 0.3$, and $P(3, 3) = 0.5$. We find $F_{X,Y}(x, y)$.

In constructing the joint distribution function, we observe that the event $\{X \le x, Y \le y\}$ has no elements for any $x < 1$ and/or $y < 1$. Only at the point (1, 1) does the function assume a step value. So long as $x \ge 1$ and $y \ge 1$, this probability is maintained so that $F_{X,Y}(x, y)$ has a stair step holding in the region $x \ge 1$ and $y \ge 1$ as shown in Figure 4.2-1*a*. For larger x and y, the point (2, 1) produces a second stair step of amplitude 0.3 which holds in the region $x \ge 2$ and $y \ge 1$. The second step adds to the first. Finally, a third stair step of amplitude 0.5 is added to the first two when x and y are in the region $x \ge 3$ and $y \ge 3$. The final function is shown in Figure 4.2-1*a*.

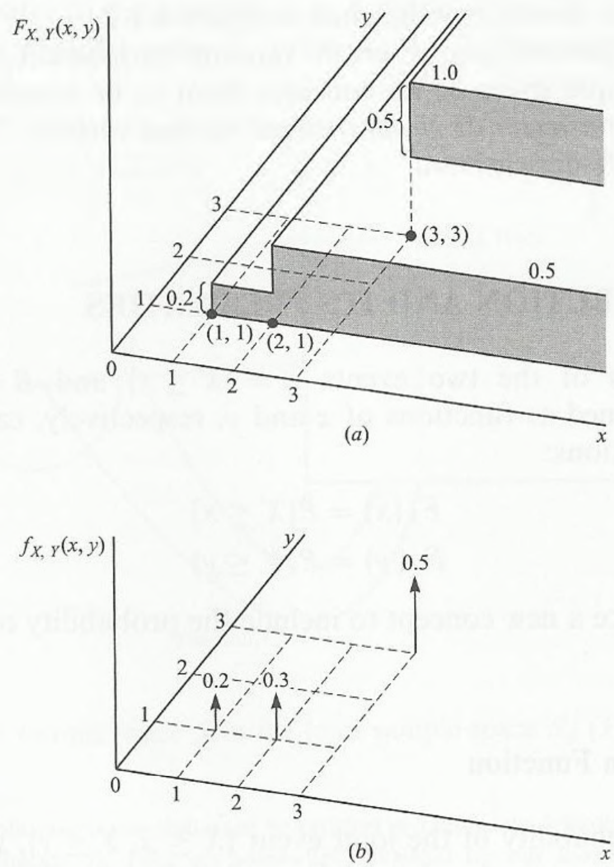


FIGURE 4.2-1
A joint distribution function (*a*), and its corresponding joint density function (*b*), that apply to Examples 4.2-1 and 4.2-2.

The preceding example can be used to identify the form of the joint distribution function for two general discrete random variables. Let X have N possible values x_n and Y have M possible values y_m, then

$$F_{X,Y}(x, y) = \sum_{n=1}^{N}\sum_{m=1}^{M} P(x_n, y_m)u(x - x_n)u(y - y_m) \tag{4.2-4}$$

where $P(x_n, y_m)$ is the probability of the joint event $\{X = x_n, Y = y_m\}$ and $u(\cdot)$ is the unit-step function. As seen in Example 4.2-1, some couples (x_n, y_m) may have zero probability. In some cases N or M, or both, may be infinite.

If $F_{X,Y}(x, y)$ is plotted for continuous random variables X and Y, the same general behavior as shown in Figure 4.2-1a is obtained except the surface becomes smooth and has no stairstep discontinuities.

For N random variables X_n, $n = 1, 2, \ldots, N$, the generalization of (4.2-3) is direct. The joint distribution function, denoted by $F_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N)$, is defined as the probability of the joint event $\{X_1 \le x_1, X_2 \le x_2, \ldots, X_N \le x_N\}$:

$$F_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) = P\{X_1 \le x_1, X_2 \le x_2, \ldots, X_N \le x_N\} \tag{4.2-5}$$

For a single random variable X, we found in Chapter 2 that $F_X(x)$ could be expressed in general as the sum of a function of stairstep form (due to the discrete portion of a mixed random variable X) and a function that was continuous (due to the continuous portion of X). Such a simple decomposition of the joint distribution when $N > 1$ is not generally true [Cramér, 1946, Section 8.4]. However, it is true that joint density functions in practice often correspond to all random variables being either discrete or continuous. Therefore, we shall limit our consideration in this book almost entirely to these two cases when $N > 1$.

Properties of the Joint Distribution

A joint distribution function for two random variables X and Y has several properties that follow readily from its definition. We list them:

(1) $F_{X,Y}(-\infty, -\infty) = 0 \qquad F_{X,Y}(-\infty, y) = 0 \qquad F_{X,Y}(x, -\infty) = 0$ (4.2-6a)

(2) $F_{X,Y}(\infty, \infty) = 1$ (4.2-6b)

(3) $0 \le F_{X,Y}(x, y) \le 1$ (4.2-6c)

(4) $F_{X,Y}(x, y)$ is a nondecreasing function of both x and y (4.2-6d)

(5) $F_{X,Y}(x_2, y_2) + F_{X,Y}(x_1, y_1) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1)$
$= P\{x_1 < X \le x_2, y_1 < Y \le y_2\} \ge 0$ (4.2-6e)

(6) $F_{X,Y}(x, \infty) = F_X(x) \qquad F_{X,Y}(\infty, y) = F_Y(y)$ (4.2-6f)

The first five of these properties are just the two-dimensional extensions of the properties of one random variable given in (2.2-2). Properties 1, 2, and 5 may be used as tests to determine whether some function can be a valid distribution function for two random variables X and Y (Papoulis, 1965, p. 169). Property 6 deserves a few special comments.

Marginal Distribution Functions

Property 6 above states that the distribution function of one random variable can be obtained by setting the value of the other variable to infinity in $F_{X,Y}(x, y)$. The functions $F_X(x)$ or $F_Y(y)$ obtained in this manner are called *marginal distribution functions*.

To justify property 6, it is easiest to return to the basic events A and B defined by $A = \{X \le x\}$ and $B = \{Y \le y\}$, and observe that $F_{X,Y}(x, y) = P\{X \le x, Y \le y\} = P(A \cap B)$. Now if we set y to ∞, this is equivalent to making B the certain event; that is, $B = \{Y \le \infty\} = S$. Furthermore, since $A \cap B = A \cap S = A$, then we have $F_{X,Y}(x, \infty) = P(A \cap S) = P(A) = P\{X \le x\} = F_X(x)$. A similar proof can be stated for obtaining $F_Y(y)$.

EXAMPLE 4.2-2. We find explicit expressions for $F_{X,Y}(x, y)$, and the marginal distributions $F_X(x)$ and $F_Y(y)$ for the joint sample space of Example 4.2-1.

The joint distribution derives from (4.2-4) if we recognize that only three probabilities are nonzero:

$$\begin{aligned} F_{X,Y}(x, y) = {} & P(1, 1)u(x-1)u(y-1) \\ & + P(2, 1)u(x-2)u(y-1) \\ & + P(3, 3)u(x-3)u(y-3) \end{aligned}$$

where $P(1, 1) = 0.2$, $P(2, 1) = 0.3$, and $P(3, 3) = 0.5$. If we set $y = \infty$:

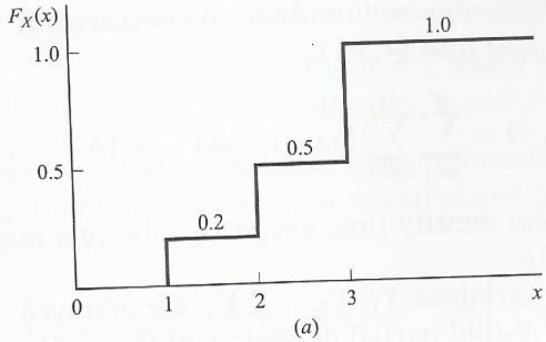
$$\begin{aligned} F_X(x) &= F_{X,Y}(x, \infty) \\ &= P(1, 1)u(x-1) + P(2, 1)u(x-2) + P(3, 3)u(x-3) \\ &= 0.2u(x-1) + 0.3u(x-2) + 0.5u(x-3) \end{aligned}$$

If we set $x = \infty$:

$$\begin{aligned} F_Y(y) &= F_{X,Y}(\infty, y) \\ &= 0.2u(y-1) + 0.3u(y-1) + 0.5u(y-3) \\ &= 0.5u(y-1) + 0.5u(y-3) \end{aligned}$$

Plots of these marginal distributions are shown in Figure 4.2-2.

From an N-dimensional joint distribution function we may obtain a *k-dimensional marginal distribution function*, for any selected group of k of the N random variables, by setting the values of the other $N - k$ random variables to infinity. Here k can be any integer $1, 2, 3, \ldots, N-1$.



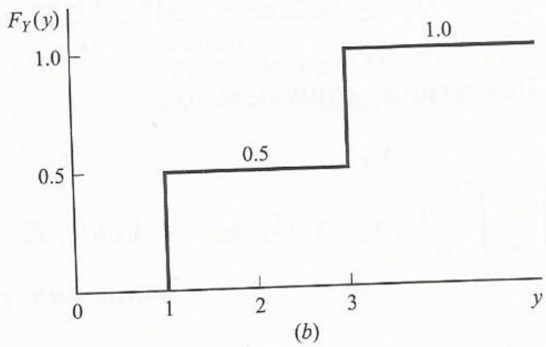


FIGURE 4.2-2
Marginal distributions applicable to Figure 4.2-1 and Example 4.2-2: (*a*) $F_X(x)$ and (*b*) $F_Y(y)$.

4.3 JOINT DENSITY AND ITS PROPERTIES

In this section the concept of a probability density function is extended to include multiple random variables.

Joint Density Function

For two random variables X and Y, the *joint probability density function*, denoted $f_{X,Y}(x, y)$, is defined by the second derivative of the joint distribution function wherever it exists:

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x\, \partial y} \tag{4.3-1}$$

We shall refer often to $f_{X,Y}(x, y)$ as the *joint density function*.

If X and Y are discrete random variables, $F_{X,Y}(x, y)$ will possess step discontinuities (see Example 4.2-1 and Figure 4.2-1). Derivatives at these discontinuities are normally undefined. However, by admitting impulse functions (see Appendix A), we are able to define $f_{X,Y}(x, y)$ at these points. Therefore, the joint density function may be found for any two discrete random variables by substitution of (4.2-4) into (4.3-1):

$$f_{X,Y}(x, y) = \sum_{n=1}^{N}\sum_{m=1}^{M} P(x_n, y_m)\,\delta(x - x_n)\,\delta(y - y_m) \tag{4.3-2}$$

An example of the joint density function of two discrete random variables is shown in Figure 4.2-1b.

When N random variables $X_1, X_2, \ldots, X_N$ are involved, the joint density function becomes the N-fold partial derivative of the N-dimensional distribution function:

$$f_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) = \frac{\partial^N F_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N)}{\partial x_1\,\partial x_2 \ldots \partial x_N} \tag{4.3-3}$$

By direct integration this result is equivalent to

$$F_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) = \int_{-\infty}^{x_N}\cdots\int_{-\infty}^{x_2}\int_{-\infty}^{x_1} f_{X_1,X_2,\ldots,X_N}(\xi_1, \xi_2, \ldots, \xi_N)d\xi_1 d\xi_2 \ldots d\xi_N \tag{4.3-4}$$

Properties of the Joint Density

Several properties of a joint density function may be listed that derive from its definition (4.3-1) and the properties (4.2-6) of the joint distribution function:

(1) $f_{X,Y}(x, y) \ge 0$ (4.3-5a)

(2) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx\,dy = 1$ (4.3-5b)

(3) $\displaystyle F_{X,Y}(x, y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(\xi_1, \xi_2)\,d\xi_1\,d\xi_2$ (4.3-5c)

(4) $\displaystyle F_X(x) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(\xi_1, \xi_2)\,d\xi_2\,d\xi_1$ (4.3-5d)

$\displaystyle F_Y(y) = \int_{-\infty}^{y}\int_{-\infty}^{\infty} f_{X,Y}(\xi_1, \xi_2)\,d\xi_1\,d\xi_2$ (4.3-5e)

(5) $\displaystyle P\{x_1 < X \le x_2, y_1 < Y \le y_2\} = \int_{y_1}^{y_2}\int_{x_1}^{x_2} f_{X,Y}(x, y)\,dx\,dy$ (4.3-5f)

(6) $\displaystyle f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy$ (4.3-5g)

$\displaystyle f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx$ (4.3-5h)

Property 1 and 2 may be used as sufficient tests to determine if some function can be a valid density function. Both tests must be satisfied (Papoulis, 1965, p. 169).

EXAMPLE 4.3-1. Suppose b is a positive constant and we test the function

$$g(x, y) = \begin{cases} b\,e^{-x}\cos(y) & 0 \le x \le 2 \text{ and } 0 \le y \le \pi/2 \\ 0 & \text{all other } x \text{ and } y \end{cases}$$

to see if it can be valid probability density function. For the allowed values of x and y the function is not negative and satisfies (4.3-5a). The final test is (4.3-5b)

$$\int_0^{\pi/2}\int_0^2 b\,e^{-x}\cos(y)\,dx\,dy = b\int_0^2 e^{-x}\,dx\int_0^{\pi/2}\cos(y)\,dy$$
$$= b(1 - e^{-2}) = 1$$

Thus, to be valid $b = 1/[1 - \exp(-2)]$ is necessary.

The first five of the above properties are readily verified from earlier work, and the reader should go through the necessary logic as an exercise. Property 6 introduces a new concept.

Marginal Density Functions

The functions $f_X(x)$ and $f_Y(y)$ of property 6 are called *marginal probability density functions* or just *marginal density functions*. They are the density functions of the single variables X and Y and are defined as the derivatives of the marginal distribution functions:

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{4.3-6}$$

$$f_Y(y) = \frac{dF_Y(y)}{dy} \tag{4.3-7}$$

By substituting (4.3-5d) and (4.3-5e) into (4.3-6) and (4.3-7), respectively, we are able to verify the equations of property 6.

We shall illustrate the calculation of marginal density functions from a given joint density function with an example.

EXAMPLE 4.3-2. We find $f_X(x)$ and $f_Y(y)$ when the joint density function is given by (Clarke and Disney, 1970, p. 108):

$$f_{X,Y}(x, y) = u(x)u(y)xe^{-x(y+1)}$$

From (4.3-5g) and the above equation:

$$f_X(x) = \int_0^\infty u(x)xe^{-x(y+1)}\,dy = u(x)xe^{-x}\int_0^\infty e^{-xy}\,dy$$
$$= u(x)xe^{-x}(1/x) = u(x)e^{-x}$$

after using an integral from Appendix C.

From (4.3-5h):

$$f_Y(y) = \int_0^\infty u(y)xe^{-x(y+1)}\,dx = \frac{u(y)}{(y+1)^2}$$

after using another integral from Appendix C.

For N random variables $X_1, X_2, \ldots, X_N$, the *k-dimensional marginal density function* is defined as the k-fold partial derivative of the k-dimensional marginal distribution function. It can also be found from the joint density function by integrating out all variables except the k variables of interest $X_1, X_2, \ldots, X_k$:

$$f_{X_1,X_2,\ldots,X_k}(x_1, x_2, \ldots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N)dx_{k+1}dx_{k+2}\ldots dx_N \quad (4.3\text{-}8)$$

4.4 CONDITIONAL DISTRIBUTION AND DENSITY

In Section 2.6, the conditional distribution function of a random variable X, given some event B, was defined as

$$F_X(x|B) = P\{X \le x|B\} = \frac{P\{X \le x \cap B\}}{P(B)} \quad (4.4\text{-}1)$$

for any event B with nonzero probability. The corresponding conditional density function was defined through the derivative

$$f_X(x|B) = \frac{dF_X(x|B)}{dx} \quad (4.4\text{-}2)$$

In this section these two functions are extended to include a second random variable through suitable definitions of event B.

Conditional Distribution and Density—Point Conditioning

Often in practical problems we are interested in the distribution function of one random variable X conditioned by the fact that a second random variable Y has some specific value y. This is called *point conditioning*, and we can handle such problems by defining event B by

$$B = \{y - \Delta y < Y \le y + \Delta y\} \quad (4.4\text{-}3)$$

where Δy is a small quantity that we eventually let approach 0. For this event, (4.4-1) can be written

$$F_X(x|y-\Delta y < Y \le y+\Delta y) = \frac{\displaystyle\int_{y-\Delta y}^{y+\Delta y}\int_{-\infty}^{x} f_{X,Y}(\xi_1,\xi_2)\,d\xi_1\,d\xi_2}{\displaystyle\int_{y-\Delta y}^{y+\Delta y} f_Y(\xi)\,d\xi} \tag{4.4-4}$$

where we have used (4.3-5f) and (2.3-6d).

Consider two cases of (4.4-4). In the first case, assume X and Y are both discrete random variables with values x_i, $i = 1, 2, \ldots, N$, and y_j, $j = 1, 2, \ldots, M$, respectively, while the probabilities of these values are denoted $P(x_i)$ and $P(y_j)$, respectively. The probability of the joint occurrence of x_i and y_j is denoted $P(x_i, y_j)$. Thus,

$$f_Y(y) = \sum_{j=1}^{M} P(y_j)\,\delta(y-y_j) \tag{4.4-5}$$

$$f_{X,Y}(x,y) = \sum_{i=1}^{N}\sum_{j=1}^{M} P(x_i,y_j)\,\delta(x-x_i)\,\delta(y-y_j) \tag{4.4-6}$$

Now suppose that the specific value of y of interest is y_k. With substitution of (4.4-5) and (4.4-6) into (4.4-4) and allowing $\Delta y \to 0$, we obtain

$$F_X(x|Y=y_k) = \sum_{i=1}^{N} \frac{P(x_i,y_k)}{P(y_k)}\,u(x-x_i) \tag{4.4-7}$$

After differentiation we have

$$f_X(x|Y=y_k) = \sum_{i=1}^{N} \frac{P(x_i,y_k)}{P(y_k)}\,\delta(x-x_i) \tag{4.4-8}$$

EXAMPLE 4.4-1. To illustrate the use of (4.4-8) assume a joint density function as given in Figure 4.4-1a. Here $P(x_1, y_1) = \frac{2}{15}$, $P(x_2, y_1) = \frac{3}{15}$, etc. Since $P(y_3) = (\frac{4}{15}) + (\frac{5}{15}) = \frac{9}{15}$, use of (4.4-8) will give $f_X(x|Y = y_3)$ as shown in Figure 4.4-1b.

The second case of (4.4-4) that is of interest corresponds to X and Y both continuous random variables. As $\Delta y \to 0$ the denominator in (4.4-4) becomes 0. However, we can still show that the conditional density $f_X(x|Y = y)$ may exist. If Δy is very small, (4.4-4) can be written as

$$F_X(x|\,y-\Delta y < Y \le y+\Delta y) = \frac{\displaystyle\int_{-\infty}^{x} f_{X,Y}(\xi_1,y)\,d\xi_1\,2\Delta y}{f_Y(y)2\Delta y} \tag{4.4-9}$$

and, in the limit as $\Delta y \to 0$

$$F_X(x|\,Y=y) = \frac{\displaystyle\int_{-\infty}^{x} f_{X,Y}(\xi,y)\,d\xi}{f_Y(y)} \tag{4.4-10}$$

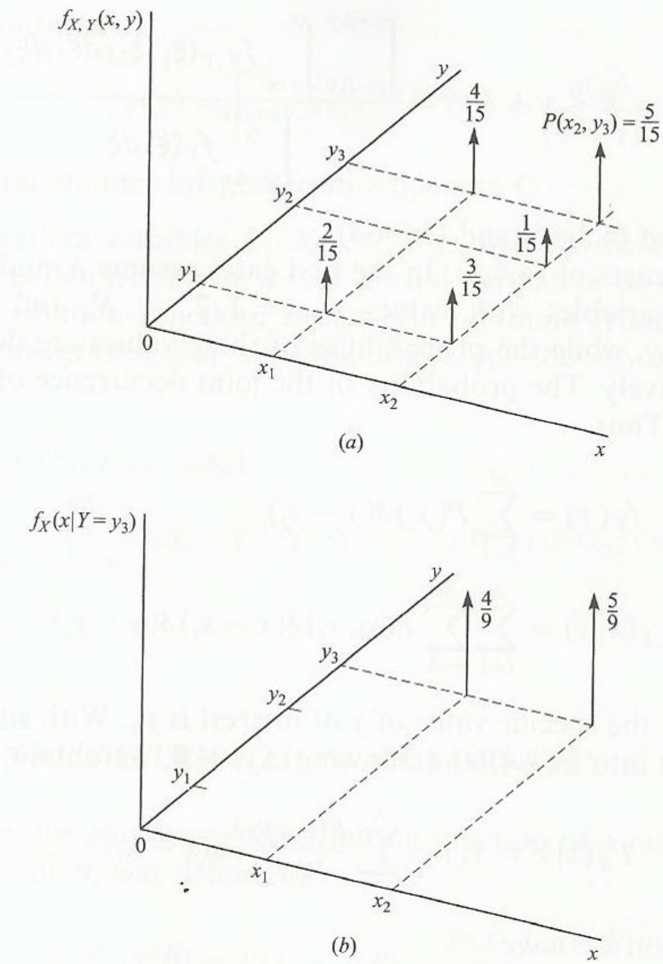


FIGURE 4.4-1
A joint density function (*a*) and a conditional density function (*b*) applicable to Example 4.4-1.

for every y such that $f_Y(y) \neq 0$. After differentiation of both sides of (4.4-10) with respect to x:

$$f_X(x|Y = y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \tag{4.4-11}$$

When there is no confusion as to meaning, we shall often write (4.4-11) as

$$f_X(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \tag{4.4-12}$$

It can also be shown that

$$f_Y(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \tag{4.4-13}$$

EXAMPLE 4.4-2. We find $f_Y(y|x)$ for the density functions defined in Example 4.3-2. Since

$$f_{X,Y}(x, y) = u(x)u(y)xe^{-x(y+1)}$$

and

$$f_X(x) = u(x)e^{-x}$$

are nonzero only for $0 < y$ and $0 < x$, $f_Y(y|x)$ is nonzero only for $0 < y$ and $0 < x$. It is

$$f_Y(y|x) = u(x)u(y)xe^{-xy}$$

from (4.4-13).

*Conditional Distribution and Density—Interval Conditioning

It is sometimes convenient to define event B in (4.4-1) and (4.4-2) in terms of a random variable Y by

$$B = \{y_a < Y \leq y_b\} \tag{4.4-14}$$

where y_a and y_b are real numbers and we assume $P(B) = P\{y_a < Y \leq y_b\} \neq 0$. With this definition it is readily shown that (4.4-1) and (4.4-2) become

$$\begin{aligned} F_X(x|y_a < Y \leq y_b) &= \frac{F_{X,Y}(x, y_b) - F_{X,Y}(x, y_a)}{F_Y(y_b) - F_Y(y_a)} \\ &= \frac{\displaystyle\int_{y_a}^{y_b}\int_{-\infty}^{x} f_{X,Y}(\xi, y)\, d\xi\, dy}{\displaystyle\int_{y_a}^{y_b}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy} \end{aligned} \tag{4.4-15}$$

and

$$f_X(x|\, y_a < Y \leq y_b) = \frac{\displaystyle\int_{y_a}^{y_b} f_{X,Y}(x, y)\, dy}{\displaystyle\int_{y_a}^{y_b}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy} \tag{4.4-16}$$

These last two expressions hold for X and Y either continuous or discrete random variables. In the discrete case, the joint density is given by (4.3-2). The resulting distribution and density will be defined, however, only for y_a and y_b such that the denominators of (4.4-15) and (4.4-16) are nonzero. This requirement is satisfied so long as the interval $y_a < y \leq y_b$ spans at least one possible value of Y having a nonzero probability of occurrence.

An example will serve to illustrate the application of (4.4-16) when X and Y are continuous random variables.

EXAMPLE 4.4-3. We use (4.4-16) to find $f_X(x|Y \leq y)$ for the joint density function of Example 4.3-2. Since we have here defined $B = \{Y \leq y\}$, then $y_a = -\infty$ and $y_b = y$. Furthermore, since $f_{X,Y}(x, y)$ is nonzero only for $0 < x$ and $0 < y$, we need only consider this region of x and y in finding the conditional density function. The denominator of (4.4-16) can be written as $\int_{-\infty}^{y} f_Y(\xi)\, d\xi$. By using results from Example 4.3-1:

$$\int_{-\infty}^{y} f_Y(\xi)\, d\xi = \int_{-\infty}^{y} \frac{u(\xi)\, d\xi}{(\xi+1)^2} = \int_{0}^{y} \frac{d\xi}{(\xi+1)^2} = \frac{y}{y+1} \qquad y > 0$$

and zero for $y < 0$, after using an integral from Appendix C. The numerator of (4.4-16) becomes

$$\begin{aligned} \int_{-\infty}^{y} f_{X,Y}(x, \xi)\, d\xi &= \int_{0}^{y} u(x) x e^{-x(\xi+1)}\, d\xi \\ &= u(x) x e^{-x} \int_{0}^{y} e^{-x\xi}\, d\xi \\ &= u(x) e^{-x}(1 - e^{-xy}) \qquad y > 0 \end{aligned}$$

and zero for $y < 0$, after using another integral from Appendix C. Thus

$$f_X(x|Y \leq y) = u(x)u(y)\left(\frac{y+1}{y}\right) e^{-x}(1 - e^{-xy})$$

This function is plotted in Figure 4.4-2 for several values of y.

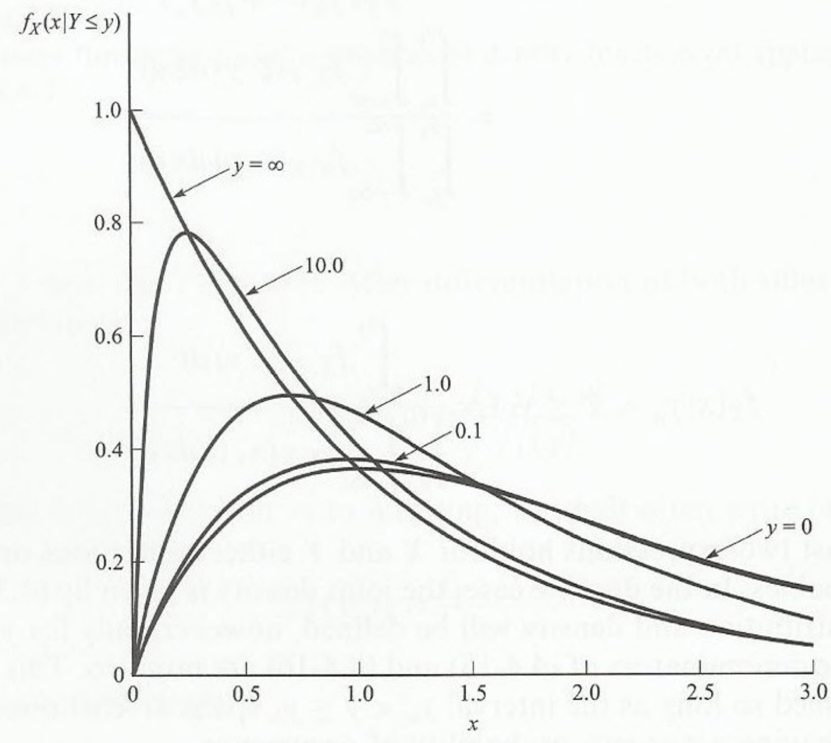


FIGURE 4.4-2
Conditional probability density functions applicable to Example 4.4-3.

4.5 STATISTICAL INDEPENDENCE

It will be recalled from (1.5-3) that two events A and B are statistically independent if (and only if)

$$P(A \cap B) = P(A)P(B) \tag{4.5-1}$$

This condition can be used to apply to two random variables X and Y by defining the events $A = \{X \leq x\}$ and $B = \{Y \leq y\}$ for two real numbers x and y. Thus, X and Y are said to be *statistically independent random variables* if (and only if)

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\} \tag{4.5-2}$$

From this expression and the definitions of distribution functions, it follows that

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \tag{4.5-3}$$

if X and Y are independent. From the definitions of density functions, (4.5-3) gives

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \tag{4.5-4}$$

by differentiation, if X and Y are independent. Either (4.5-3) or (4.5-4) may serve as a sufficient definition of, or test for, independence of two random variables.

The form of the conditional distribution function for independent events is found by use of (4.4-1) with $B = \{Y \leq y\}$:

$$F_X(x|Y \leq y) = \frac{P\{X \leq x, Y \leq y\}}{P\{Y \leq y\}} = \frac{F_{X,Y}(x, y)}{F_Y(y)} \tag{4.5-5}$$

By substituting (4.5-3) into (4.5-5), we have

$$F_X(x|Y \leq y) = F_X(x) \tag{4.5-6}$$

In other words, the conditional distribution ceases to be conditional and simply equals the marginal distribution for independent random variables. It can also be shown that

$$F_Y(y|X \leq x) = F_Y(y) \tag{4.5-7}$$

Conditional density function forms, for independent X and Y, are found by differentiation of (4.5-6) and (4.5-7):

$$f_X(x|Y \leq y) = f_X(x) \tag{4.5-8}$$

$$f_Y(y|X \leq x) = f_Y(y) \tag{4.5-9}$$

EXAMPLE 4.5-1. For the densities of Example 4.3-1:

$$f_{X,Y}(x, y) = u(x)u(y)xe^{-x(y+1)}$$

$$f_X(x)f_Y(y) = u(x)u(y)\frac{e^{-x}}{(y+1)^2} \neq f_{X,Y}(x, y)$$

Therefore, the random variables X and Y are not independent.

EXAMPLE 4.5-2. The joint density of two random variables X and Y is

$$f_{X,Y}(x, y) = \frac{1}{12}u(x)u(y)e^{-(x/4)-(y/3)}$$

We determine if X and Y are statistically independent. From (4.3-5g) and (4.3-5h)

$$f_X(x) = \int_0^{\infty} (1/12)u(x)e^{-x/4}e^{-y/3}\,dy = (1/4)u(x)e^{-x/4}$$

$$f_Y(y) = \int_0^{\infty} (1/12)u(y)e^{-y/3}e^{-x/4}\,dx = (1/3)u(y)e^{-y/3}$$

Since $f_X(x)f_Y(y) = f_{X,Y}(x, y)$, then X and Y are independent.

In the more general study of the statistical independence of N random variables $X_1, X_2, \ldots, X_N$, we define events A_i by

$$A_i = \{X_i \leq x_i\} \qquad i = 1, 2, \ldots, N \tag{4.5-10}$$

where the x_i are real numbers. With these definitions, the random variables X_i are said to be statistically independent if (1.5-6) is satisfied.

It can be shown that if $X_1, X_2, \ldots, X_N$ are statistically independent then any group of these random variables is independent of any other group. Furthermore, a function of any group is independent of any function of any other group of the random variables. For example, with $N = 4$ random variables: X_4 is independent of $X_3 + X_2 + X_1$; X_3 is independent of $X_2 + X_1$, etc. (see Papoulis, 1965, p. 238).

4.6 DISTRIBUTION AND DENSITY OF A SUM OF RANDOM VARIABLES

The problem of finding the distribution and density functions for a sum of *statistically independent* random variables is considered in this section.

Sum of Two Random Variables

Let W be a random variable equal to the sum of two independent random variables X and Y:

$$W = X + Y \tag{4.6-1}$$

This is a very practical problem because X might represent a random signal voltage and Y could represent random noise at some instant in time. The sum W would represent a signal-plus-noise voltage available to some receiver.

The probability distribution function we seek is defined by

$$F_W(w) = P\{W \leq w\} = P\{X + Y \leq w\} \tag{4.6-2}$$

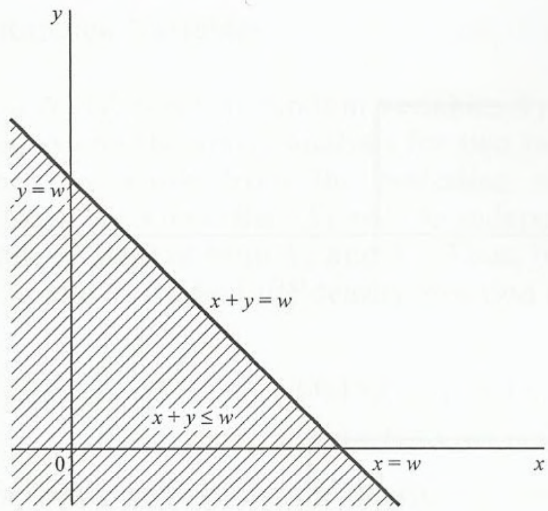


FIGURE 4.6-1
Region in xy plane where $x + y \leq w$.

Figure 4.6-1 illustrates the region in the xy plane where $x + y \leq w$. Now from (4.3-5f), the probability corresponding to an elemental area $dx\,dy$ in the xy plane located at the point (x, y) is $f_{X,Y}(x, y)\,dx\,dy$. If we sum all such probabilities over the region where $x + y \leq w$ we will obtain $F_W(w)$. Thus

$$F_W(w) = \int_{-\infty}^{\infty} \int_{x=-\infty}^{w-y} f_{X,Y}(x, y)\,dx\,dy \tag{4.6-3}$$

and, after using (4.5-4):

$$F_W(w) = \int_{-\infty}^{\infty} f_Y(y) \int_{x=-\infty}^{w-y} f_X(x)\,dx\,dy \tag{4.6-4}$$

By differentiating (4.6-4), using Leibniz's rule, we get the desired density function

$$f_W(w) = \int_{-\infty}^{\infty} f_Y(y) f_X(w-y)\,dy \tag{4.6-5}$$

This expression is recognized as a convolution integral. Consequently, we have shown that *the density function of the sum of two statistically independent random variables is the convolution of their individual density functions.*

EXAMPLE 4.6-1. We use (4.6-5) to find the density of $W = X + Y$ where the densities of X and Y are assumed to be

$$f_X(x) = \frac{1}{a}[u(x) - u(x-a)]$$

$$f_Y(y) = \frac{1}{b}[u(y) - u(y-b)]$$

with $0 < a < b$, as shown in Figure 4.6-2a and b. Now because $0 < X$ and $0 < Y$, we only need examine the case $W = X + Y > 0$. From (4.6-5) we write

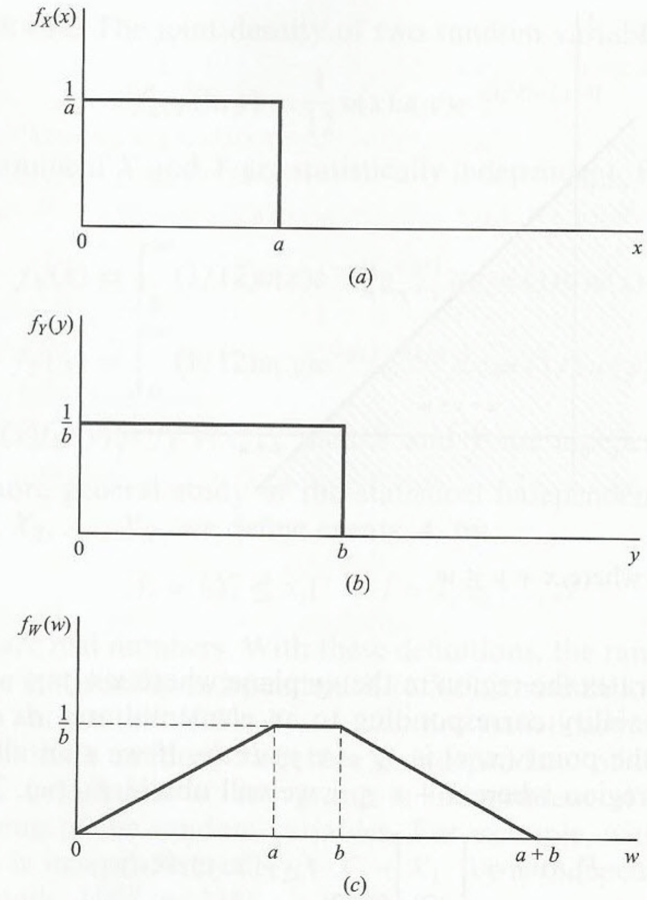


FIGURE 4.6-2
Two density functions (*a*) and (*b*) and their convolution (*c*).

$$\begin{aligned} f_W(w) &= \int_{-\infty}^{\infty} \frac{1}{ab}[u(y) - u(y-b)][u(w-y) - u(w-y-a)]\,dy \\ &= \frac{1}{ab}\int_0^{\infty} [1 - u(y-b)][u(w-y) - u(w-y-a)]\,dy \\ &= \frac{1}{ab}\left[\int_0^{\infty} u(w-y)\,dy - \int_0^{\infty} u(w-y-a)\,dy \right. \\ &\qquad \left. - \int_0^{\infty} u(y-b)u(w-y)\,dy + \int_0^{\infty} u(y-b)u(w-y-a)\,dy\right] \end{aligned}$$

All these integrands are unity; the values of the integrals are determined by the unit-step functions through their control over limits of integration. After straightforward evaluation we get

$$f_W(w) = \begin{cases} w/ab & 0 \le w < a \\ 1/b & a \le w < b \\ (a+b-w)/ab & b \le w < a+b \\ 0 & w \ge a+b \end{cases}$$

which is sketched in Figure 4.6-2*c*.

*Sum of Several Random Variables

When the sum Y of N independent random variables $X_1, X_2, \ldots, X_N$ is to be considered, we may extend the above analysis for two random variables. Let $Y_1 = X_1 + X_2$. Then we know from the preceding work that $f_{Y_1}(y_1) = f_{X_2}(x_2) * f_{X_1}(x_1)$.† Next, we know that X_3 will be independent of $Y_1 = X_1 + X_2$ because X_3 is independent of both X_1 and X_2. Thus, by applying (4.6-5) to the two variables X_3 and Y_1 to find the density function of $Y_2 = X_3 + Y_1$, we get

$$\begin{aligned} f_{Y_2=X_1+X_2+X_3}(y_2) &= f_{X_3}(x_3) * f_{Y_1=X_1+X_2}(y_1) \\ &= f_{X_3}(x_3) * f_{X_2}(x_2) * f_{X_1}(x_1) \end{aligned} \tag{4.6-6}$$

By continuing the process we find that the density function of $Y = X_1 + X_2 + \cdots + X_N$ is the $(N-1)$-fold convolution of the N individual density functions:

$$f_Y(y) = f_{X_N}(x_N) * f_{X_{N-1}}(x_{N-1}) * \cdots * f_{X_1}(x_1) \tag{4.6-7}$$

The distribution function of Y is found from the integral of $f_Y(y)$ using (2.3-6*c*).

Another method using characteristic functions can also be employed to find the density function of a sum of random variables. A discussion of the method is given in Section 5.2 for statistically independent random variables.

*4.7 CENTRAL LIMIT THEOREM

Broadly defined, the *central limit theorem* says that the probability distribution function of the sum of a large number of random variables approaches a gaussian distribution. Although the theorem is known to apply to some cases of statistically *dependent* random variables (Cramér, 1946, p. 219), most applications, and the largest body of knowledge, are directed toward statistically independent random variables. Thus, in all succeeding discussions we assume statistically independent random variables.

*Unequal Distributions

Let $\bar{X}_i$ and $\sigma^2_{X_i}$ be the means and variances, respectively, of N random variables X_i, $i = 1, 2, \ldots, N$, which may have arbitrary probability densities. The central limit theorem states that the sum $Y_N = X_1 + X_2 + \cdots + X_N$, which has mean $\bar{Y}_N = \bar{X}_1 + \bar{X}_2 + \cdots \bar{X}_N$ and variance $\sigma^2_{Y_N} = \sigma^2_{X_1} + \sigma^2_{X_2} + \cdots + \sigma^2_{X_N}$, has a probability distribution that asymptotically approaches gaussian as $N \to \infty$. Necessary conditions for the theorem's validity are difficult to state, but sufficient conditions are known to be (Cramér, 1946; Thomas, 1969)

†The asterisk denotes convolution.

$$\sigma_{X_i}^2 > B_1 > 0 \qquad i = 1, 2, \ldots, N \tag{4.7-1a}$$

$$E[|X_i - \bar{X}_i|^3] < B_2 \qquad i = 1, 2, \ldots, N \tag{4.7-1b}$$

where B_1 and B_2 are positive numbers. These conditions guarantee that no one random variable in the sum dominates.

The reader should observe that the central limit theorem guarantees only that the *distribution* of the sum of random variables becomes gaussian. It does not follow that the probability *density* is always gaussian. For continuous random variables there is usually no problem, but certain conditions imposed on the individual random variables (Cramér, 1946; Papoulis, 1965 and 1984) will guarantee that the density is gaussian.

For discrete random variables X_i the sum Y_N will also be discrete so its density will contain impulses and is, therefore, not gaussian, even though the distribution approaches gaussian. When the possible discrete values of each random variable are kb, $k = 0, \pm 1, \pm 2, \ldots$, with b a constant,† the envelope of the impulses in the density of the sum will be gaussian (with mean $\bar{Y}_N$ and variance $\sigma_{Y_N}^2$). This case is discussed in some detail by Papoulis (1965).

The practical usefulness of the central limit theorem does not reside so much in the exactness of the gaussian distribution for $N \to \infty$ because the variance of Y_N becomes infinite from (4.7-1a). Usefulness derives more from the fact that Y_N for *finite* N may have a distribution that is closely approximated as gaussian. The approximation can be quite accurate, even for relatively small values of N, in the central region of the gaussian curve near the mean. However, the approximation can be very inaccurate in the tail regions away from the mean, even for large values of N (Davenport, 1970; Melsa and Sage, 1973). Of course, the approximation is made more accurate by increasing N.

*Equal Distributions

If all of the statistically independent random variables being summed are continuous and have the same distribution function, and therefore the same density, the proof of the central limit theorem is relatively straightforward and is next developed.

Because the sum $Y_N = X_1 + X_2 + \cdots + X_N$ has an infinite variance as $N \to \infty$, we shall work with the zero-mean, unit-variance random variable

$$W_N = (Y_N - \bar{Y}_N)/\sigma_{Y_N} = \sum_{i=1}^{N}(X_i - \bar{X}_i) \Bigg/ \left[\sum_{i=1}^{N} \sigma_{X_i}^2\right]^{1/2}$$

$$= \frac{1}{\sqrt{N}\sigma_X} \sum_{i=1}^{N}(X_i - \bar{X}) \tag{4.7-2}$$

†These are called *lattice-type* discrete random variables (Papoulis, 1965).

instead. Here we define $\bar{X}$ and σ_X^2 by

$$\bar{X}_i = \bar{X} \qquad \text{all } i \tag{4.7-3}$$

$$\sigma_{X_i}^2 = \sigma_X^2 \qquad \text{all } i \tag{4.7-4}$$

since all the X_i have the same distribution.

The theorem's proof consists of showing that the characteristic function of W_N is that of a zero-mean, unit-variance gaussian random variable, which is

$$\Phi_{W_N}(\omega) = \exp(-\omega^2/2) \tag{4.7-5}$$

from Problem 3.3-2. If this is proved the density of W_N must be gaussian from (3.3-3) and the fact that Fourier transforms are unique. The characteristic function of W_N is

$$\begin{aligned}\Phi_{W_N}(\omega) = E[e^{j\omega W_N}] &= E\left[\exp\left\{\frac{j\omega}{\sqrt{N}\sigma_X}\sum_{i=1}^{N}(X_i - \bar{X})\right\}\right] \\ &= \left\langle E\left\{\exp\left[\frac{j\omega}{\sqrt{N}\sigma_X}(X_i - \bar{X})\right]\right\}\right\rangle^N\end{aligned} \tag{4.7-6}$$

The last step in (4.7-6) follows from the independence and equal distribution of the X_i. Next, the exponential in (4.7-6) is expanded in a Taylor polynomial with a remainder term R_N/N:

$$\begin{aligned}&E\left\{\exp\left[\frac{j\omega}{\sqrt{N}\sigma_X}(X_i - \bar{X})\right]\right\} \\ &= E\left\{1 + \left(\frac{j\omega}{\sqrt{N}\sigma_X}\right)(X_i - \bar{X}) + \left(\frac{j\omega}{\sqrt{N}\sigma_X}\right)^2\frac{(X_i - \bar{X})^2}{2} + \frac{R_N}{N}\right\} \\ &= 1 - (\omega^2/2N) + E[R_N]/N\end{aligned} \tag{4.7-7}$$

where $E[R_N]$ approaches zero as $N \to \infty$ (Davenport, 1970, p. 442). On substitution of (4.7-7) into (4.7-6) and forming the natural logarithm, we have

$$\ln[\Phi_{W_N}(\omega)] = N\ln\{1 - (\omega^2/2N) + E[R_N]/N\} \tag{4.7-8}$$

Since

$$\ln(1 - z) = -\left[z + \frac{z^2}{2} + \frac{z^3}{3} + \cdots\right] \qquad |z| < 1 \tag{4.7-9}$$

we identify z with $(\omega^2/2N) - E[R_N]/N$ and write (4.7-8) as

$$\ln[\Phi_{W_N}(\omega)] = -(\omega^2/2) + E[R_N] - \frac{N}{2}\left[\frac{\omega^2}{2N} - \frac{E[R_N]}{N}\right]^2 + \cdots \tag{4.7-10}$$

so

$$\lim_{N\to\infty}\{\ln[\Phi_{W_N}(\omega)]\} = \ln\left\{\lim_{N\to\infty}\Phi_{W_N}(\omega)\right\} = -\omega^2/2 \tag{4.7-11}$$

Finally, we have

$$\lim_{N \to \infty} \Phi_{W_N}(\omega) = e^{-\omega^2/2} \tag{4.7-12}$$

which was to be shown.

We illustrate the use of the central limit theorem through an example.

EXAMPLE 4.7-1. Consider the sum of just two independent uniformly distributed random variables X_1 and X_2 having the same density

$$f_X(x) = \frac{1}{a}[u(x) - u(x-a)]$$

where $a > 0$ is a constant. The means and variances of X_1 and X_2 are $\bar{X} = a/2$ and $\sigma_X^2 = a^2/12$, respectively. The density of the sum $W = X_1 + X_2$ is available from Example 4.6-1 (with $b = a$):

$$f_X(w) = \frac{1}{a}\,\mathrm{tri}\left(\frac{w}{a}\right)$$

where the function tri$(\cdot)$ is defined in (E-4). The gaussian approximation to W has variance $\sigma_W^2 = 2\sigma_X^2 = a^2/6$ and mean $\bar{W} = 2(a/2) = a$:

$$\text{Approximation to } f_W(w) = \frac{e^{-(w-a)^2/(a^2/3)}}{\sqrt{\pi(a^2/3)}}$$

Figure 4.7-1 illustrates $f_W(w)$ and its gaussian approximation. Even for the case of only two random variables being summed the gaussian approximation is a fairly good one. For other densities the approximation may be very poor (see Problem 4.7-1).

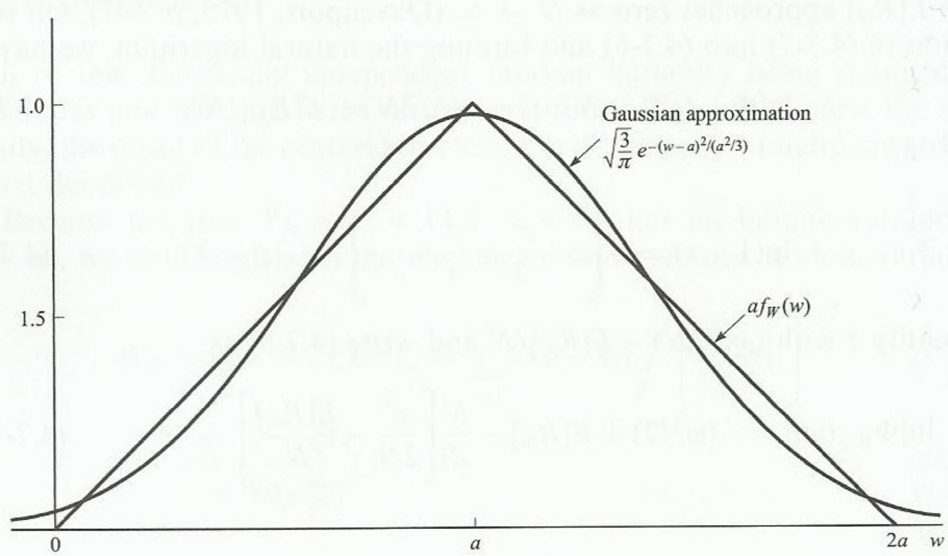


FIGURE 4.7-1
The triangular density function of Example 4.7-1 and its gaussian approximation.

4.8 SUMMARY

One random variable is inadequate to represent many practical problems. The theory of multiple random variables is needed and was developed in this chapter. The main points covered were:

- Multiple (vector) random variables were defined and related through examples to real problems.
- The earlier concepts of joint density and distribution functions, and their properties, were extended to include several random variables.
- Conditional density and distribution functions were developed for several random variables.
- Multiple-variable statistical independence was developed.
- Methods were given to find the distribution and density functions of two or more statistically independent random variables. For some cases the problem can be approximated by the central limit theorem, which is developed in some detail.

This chapter consisted mainly of the extension of the one-variable theory of Chapter 2 to the multiple-random-variable case. The next logical step is to extend the one-variable operations of Chapter 3 to cover several random variables. This extension follows in the next chapter.

PROBLEMS

4.1-1. Two events A and B defined on a sample space S are related to a joint sample space through random variables X and Y and are defined by $A = \{X \leq x\}$ and $B = \{y_1 < Y \leq y_2\}$. Make a sketch of the two sample spaces showing areas corresponding to both events and the event $A \cap B = \{X \leq x, y_1 < Y \leq y_2\}$.

4.1-2. Work Problem 4.1-1 for the two events $A = \{x_1 < X \leq x_2\}$ and $B = \{y_1 < Y \leq y_2\}$.

4.1-3. Work Problem 4.1-1 for the two events $A = \{x_1 < X \leq x_2$ or $x_3 < X \leq x_4\}$ and $B = \{y_1 < Y \leq y_2\}$.

4.1-4. Three events A, B, and C satisfy $C \subset B \subset A$ and are defined by $A = \{X \leq x_a, Y \leq y_a\}$, $B = \{X \leq x_b, Y \leq y_b\}$, and $C = \{X \leq x_c, Y \leq y_c\}$ for two random variables X and Y.
(*a*) Sketch the two sample spaces S and S_J and show the regions corresponding to the three events.
(*b*) What region corresponds to the event $A \cap B \cap C$?

4.1-5. In a gambling game two fair dice are tossed and the sum of the numbers that show up determines who wins among two players. Random variables X and Y represent the winnings of the first and second numbered players, respectively. The first wins \$3 if the sum is 4, 5, or 6, and loses \$2 if the sum is 11 or 12; he neither wins nor loses for all other sums. The second player wins \$2 for a sum of 8 or more, loses \$3 for a sum of 5 or less, and neither wins nor loses for other sums.

(*a*) Draw sample spaces S and S_J and show how elements of S map to elements of S_J.

(*b*) Find the probabilities of all joint outcomes possible in S_J.

4.1-6. Sketch the joint sample space for two random variables X and Y and define the regions that correspond to the events $A = \{Y \le 2X\}$, $B = \{X \le 4\}$, and $C = \{Y > -2\}$. Indicate the region defined by $A \cap B \cap C$.

4.2-1. A joint sample space for two random variables X and Y has four elements (1, 1), (2, 2), (3, 3), and (4, 4). Probabilities of these elements are 0.1, 0.35, 0.05, and 0.5, respectively.

(*a*) Determine through logic and sketch the distribution function $F_{X,Y}(x, y)$.

(*b*) Find the probability of the event $\{X \le 2.5, Y \le 6\}$.

(*c*) Find the probability of the event $\{X \le 3\}$.

4.2-2. Write a mathematical equation for $F_{X,Y}(x, y)$ of Problem 4.2-1.

4.2-3. The joint distribution function for two random variables X and Y is

$$F_{X,Y}(x, y) = u(x)u(y)[1 - e^{-ax} - e^{-ay} + e^{-a(x+y)}]$$

where $u(\cdot)$ is the unit-step function and $a > 0$. Sketch $F_{X,Y}(x, y)$.

4.2-4. By use of the joint distribution function in Problem 4.2-3, and assuming $a =$ 0.5 in each case, find the probabilities:

(a) $P\{X \le 1, Y \le 2\}$

(*b*) $P\{0.5 < X < 1.5\}$

(*c*) $P\{-1.5 < X \le 2, 1 < Y \le 3\}$.

4.2-5. Find and sketch the marginal distribution functions for the joint distribution function of Problem 4.2-1.

4.2-6. Find and sketch the marginal distribution functions for the joint distribution function of Problem 4.2-3.

4.2-7. Given the function

$$G_{X,Y}(x, y) = u(x)u(y)[1 - e^{-(x+y)}]$$

Show that this function satisfies the first four properties of (4.2-6) but fails the fifth one. The function is therefore not a valid joint probability distribution function.

4.2-8. Random variables X and Y are components of a two-dimensional random vector and have a joint distribution

$$F_{X,Y}(x, y) = \begin{cases} 0 & x < 0 \quad \text{or} \quad y < 0 \\ xy & 0 \le x < 1 \quad \text{and} \quad 0 \le y < 1 \\ x & 0 \le x < 1 \quad \text{and} \quad 1 \le y \\ y & 1 \le x \quad \text{and} \quad 0 \le y < 1 \\ 1 & 1 \le x \quad \text{and} \quad 1 \le y \end{cases}$$

(*a*) Sketch $F_{X,Y}(x, y)$.
(*b*) Find and sketch the marginal distribution functions $F_X(x)$ and $F_Y(y)$.

4.2-9. Show the function

$$G_{X,Y}(x, y) = \begin{cases} 0 & x < y \\ 1 & x \geq y \end{cases}$$

cannot be a valid joint distribution function. [*Hint*: Use (4.26*e*).]

4.2-10. Discrete random variables X and Y have a joint distribution function

$$\begin{aligned} F_{X,Y}(x, y) = {} & 0.10u(x+4)u(y-1) + 0.15u(x+3)u(y+5) \\ & + 0.17u(x+1)u(y-3) + 0.05u(x)u(y-1) \\ & + 0.18u(x-2)u(y+2) + 0.23u(x-3)u(y-4) \\ & + 0.12u(x-4)u(y+3) \end{aligned}$$

Find: (*a*) the marginal distributions $F_X(x)$ and $F_Y(y)$ and sketch the two functions, (*b*) $\bar{X}$ and $\bar{Y}$, and (*c*) the probability $P\{-1 < X \leq 4, -3 < Y \leq 3\}$.

4.2-11. Random variables X and Y have the joint distribution

$$F_{X,Y}(x, y) = \begin{cases} \dfrac{5}{4}\left(\dfrac{x + e^{-(x+1)y^2}}{x+1} - e^{-y^2}\right)u(y) & 0 \leq x < 4 \\ 0 & x < 0 \text{ or } y < 0 \\ 1 + \dfrac{1}{4}e^{-5y^2} - \dfrac{5}{4}e^{-y^2} & 4 \leq x \text{ and any } y \geq 0 \end{cases}$$

Find: (*a*) The marginal distribution functions of X and Y, and (*b*) the probability $P\{3 < X \leq 5, 1 < Y \leq 2\}$.

4.2-12. Find the joint distribution function of the random variables having the joint density of Problem 4.3-16.

4.2-13. The function

$$F_{X,Y}(x, y) = a\left[\frac{\pi}{2} + \tan^{-1}\left(\frac{x}{2}\right)\right]\left[\frac{\pi}{2} + \tan^{-1}\left(\frac{y}{3}\right)\right]$$

is a valid joint distribution function for random variables X and Y if the constant a is chosen properly. What should be the value of a?

4.2-14. Work Problem 4.2-13, except assume the function

$$\begin{aligned} F_{X,Y}(x, y) = a & \left[\frac{\pi}{2} + \frac{\sqrt{3}x}{3+x^2} + \tan^{-1}\left(\frac{x}{\sqrt{3}}\right)\right] \\ & \cdot\left[\frac{\pi}{2} + \frac{\sqrt{5}y}{5+y^2} + \tan^{-1}\left(\frac{y}{\sqrt{5}}\right)\right] \end{aligned}$$

4.2-15. Suppose that a pair of random numbers generated by a computer are represented as values of random variables X and Y having the joint distribution function

$$F_{X,Y}(x,y) = \begin{cases} 0 & x < 0 \text{ or } y < 0 \\ \dfrac{27}{26}x\left(1 - \dfrac{x^2}{27}\right) & 0 \le x < 1 \text{ and } 1 \le y \\ \dfrac{27}{26}y\left(1 - \dfrac{y^2}{27}\right) & 1 \le x \text{ and } 0 \le y < 1 \\ \dfrac{27}{26}xy\left(1 - \dfrac{x^2y^2}{27}\right) & 0 \le x < 1 \text{ and } 0 \le y < 1 \\ 1 & 1 \le x \text{ and } 1 \le y \end{cases}$$

(*a*) Determine the marginal distribution functions of X and Y.
(*b*) Find the probability of the event $\{0 < X \le 0.5, 0 < Y \le 0.25\}$.

4.3-1. A fair coin is tossed twice. Define random variables by: $X =$ "number of heads on the first toss" and $Y =$ "number of heads on the second toss" (note that X and Y can have only the values 0 or 1).
(*a*) Find and sketch the joint density function of X and Y.
(*b*) Find and sketch the joint distribution function.

4.3-2. A joint probability density function is

$$f_{X,Y}(x,y) = \begin{cases} 1/ab & 0 < x < a \quad \text{and} \quad 0 < y < b \\ 0 & \text{elsewhere} \end{cases}$$

Find and sketch $F_{X,Y}(x,y)$.

4.3-3. If $a < b$ in Problem 4.3-2, find:
(*a*) $P\{X + Y \le 3a/4\}$ (*b*) $P\{Y \le 2bX/a\}$

4.3-4. Find the joint distribution function applicable to Example 4.3-2.

4.3-5. Sketch the joint density function $f_{X,Y}(x,y)$ applicable to Prolem 4.2-1. Write an equation for $f_{X,Y}(x,y)$.

4.3-6. Determine the joint density and both marginal density functions for Problem 4.2-3.

4.3-7. Find and sketch the joint density function for the distribution function in Problem 4.2-8.

4.3-8. (*a*) Find a constant b (in terms of a) so that the function

$$f_{X,Y}(x,y) = \begin{cases} be^{-(x+y)} & 0 < x < a \quad \text{and} \quad 0 < y < \infty \\ 0 & \text{elsewhere} \end{cases}$$

is a valid joint density function.
(*b*) Find an expression for the joint distribution function.

4.3-9. (*a*) By use of the joint density function of Problem 4.3-8, find the marginal density functions.

(*b*) What is $P\{0.5a < X \le 0.75a\}$ in terms of a and b?

4.3-10. Determine a constant b such that each of the following are valid joint density functions:

(*a*) $f_{X,Y}(x, y) = \begin{cases} 3xy & 0 < x < 1 \quad \text{and} \quad 0 < y < b \\ 0 & \text{elsewhere} \end{cases}$

(*b*) $f_{X,Y}(x, y) = \begin{cases} bx(1-y) & 0 < x < 0.5 \quad \text{and} \quad 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$

(*c*) $f_{X,Y}(x, y) = \begin{cases} b(x^2 + 4y^2) & 0 \le |x| < 1 \quad \text{and} \quad 0 \le y < 2 \\ 0 & \text{elsewhere} \end{cases}$

4.3-11. Given the function

$$f_{X,Y}(x, y) = \begin{cases} (x^2 + y^2)/8\pi & x^2 + y^2 < b \\ 0 & \text{elsewhere} \end{cases}$$

(*a*) Find a constant b so that this is a valid joint density function.
(*b*) Find $P\{0.5b < X^2 + Y^2 \le 0.8b\}$. (*Hint*: Use polar coordinates in both parts.)

4.3-12. On a firing range the coordinates of bullet strikes relative to the target bull's-eye are random variables X and Y having a joint density given by

$$f_{X,Y}(x, y) = \frac{e^{-(x^2+y^2)/2\sigma^2}}{2\pi\sigma^2}$$

Here σ^2 is a constant related to the accuracy of manufacturing a gun's barrel. What value of σ^2 will allow 80% of all bullets to fall inside a circle of diameter 6 cm? (*Hint*: Use polar coordinates.)

4.3-13. Given the function

$$f_{X,Y}(x, y) = \begin{cases} b(x + y)^2 & -2 < x < 2 \quad \text{and} \quad -3 < y < 3 \\ 0 & \text{elsewhere} \end{cases}$$

(*a*) Find the constant b such that this is a valid joint density function.
(*b*) Determine the marginal density functions $f_X(x)$ and $f_Y(y)$.

4.3-14. Find a value of the constant b so that the function

$$f_{X,Y}(x, y) = bxy^2 \exp(-2xy)u(x - 2)u(y - 1)$$

is a valid joint probability density.

4.3-15. The locations of hits of darts thrown at a round dartboard of radius r are determined by a vector random variable with components X and Y. The joint density of X and Y is uniform, that is,

$$f_{X,Y}(x, y) = \begin{cases} 1/\pi r^2 & x^2 + y^2 < r^2 \\ 0 & \text{elsewhere} \end{cases}$$

Find the densities of X and Y.

4.3-16. Two random variables X and Y have a joint density

$$f_{X,Y}(x, y) = \tfrac{10}{4}[u(x) - u(x-4)]u(y)\,y^3 \exp[-(x+1)\,y^2]$$

Find the marginal densities and distributions of X and Y.

4.3-17. Find the marginal densities of X and Y using the joint density

$$f_{X,Y}(x, y) = 2u(x)u(y)\exp\left[-\left(4y + \frac{x}{2}\right)\right]$$

4.3-18. Random variables X and Y have the joint density of Problem 4.3-17. Find the probability that the values of Y are not greater than twice the values of X for $X \le 3$.

4.3-19. The joint density of two random variables X and Y is

$$\begin{aligned} f_{X,Y}(x, y) = {} & 0.1\delta(x)\delta(y) + 0.12\delta(x-4)\delta(y) \\ & + 0.05\delta(x)\delta(y-1) + 0.25\delta(x-2)\delta(y-1) \\ & + 0.3\delta(x-2)\delta(y-3) + 0.18\delta(x-4)\delta(y-3) \end{aligned}$$

Find and plot the marginal distributions of X and Y.

4.3-20. Assume a has the proper value in Problem 4.2-13 and determine the joint density of X and Y. Find the marginal densities of X and Y.

4.3-21. Work Problem 4.3-20 but assume the distribution of Problem 4.2-14.

4.3-22. (*a*) Find the joint probability density function for the computer-generated numbers of Problem 4.2-15.
(*b*) Find the marginal densities of X and Y.
(*c*) Find the probability of the event $\{Y > 1 - X\}$.

4.3-23. The joint density function of random variables X and Y is

$$f_{X,Y}(x, y) = \begin{cases} \dfrac{25}{23ab}\left(\dfrac{y}{a}\right)\left[1 - \left(\dfrac{x}{b}\right)^4\left(\dfrac{y}{a}\right)^3\right] & -b < x < b \text{ and } 0 < y < a \\ 0 & \text{elsewhere} \end{cases}$$

where $a > 0$ and $b > 0$ are constants. Find the marginal densities of X and Y.

4.4-1. Find the conditional density functions $f_X(x|y_1)$, $f_X(x|y_2)$, $f_Y(y|x_1)$, and $f_Y(y|x_2)$ for the joint density defined in Example 4.4-1.

4.4-2. Find the conditional density function $f_X(x|y)$ applicable to Example 4.4-2.

4.4-3. By using the results of Example 4.4-2, calculate the probability of the event $\{Y \le 2|X = 1\}$.

4.4-4. Find the conditional densities $f_X(x|Y = y)$ and $f_Y(y|X = x)$ applicable to the joint density of Problem 4.3-15.

4.4-5. For the joint density of Problem 4.3-16 determine the conditional densities $f_X(x|Y=y)$ and $f_Y(y|X=x)$.

***4.4-6.** The time it takes a person to drive to work is a random variable Y. Because of traffic, driving time depends on the (random) time of departure, denoted X, which occurs in an interval of duration T_0 that begins at 7:30 A.M. each day. There is a minimum driving time T_1 required, regardless of the time of departure. The joint density of X and Y is known to be

$$f_{X,Y}(x,y) = c(y-T_1)^3 u(y-T_1)[u(x)-u(x-T_0)]\exp[-(y-T_1)(x+1)]$$

where

$$c = (1+T_0)^3/2[(1+T_0)^3-1]$$

(*a*) Find the average driving time that results when it is given that departure occurs at 7:30 A.M. Evaluate your results for $T_0 = 1$ h.
(*b*) Repeat part (*a*) given that departure time is 7:30 A.M. plus T_0.
(*c*) What is the average time of departure if $T_0 = 1$ h? (*Hint*: Note that point conditioning applies.)

***4.4-7.** Start with the expressions

$$F_Y(y|B) = P\{Y \le y|B\} = \frac{P\{Y \le y \cap B\}}{P(B)}$$

$$f_Y(y|B) = \frac{dF_Y(y|B)}{dy}$$

which are analogous to (4.4-1) and (4.4-2), and derive $F_Y(y|x_a < X \le x_b)$ and $f_Y(y|x_a < X \le x_b)$ which are analogous to (4.4-15) and (4.4-16).

***4.4-8.** Extend the procedures of the text that lead to (4.4-16) to show that the joint distribution and density of random variables X and Y, conditional on the event $B = \{y_a < Y \le y_b\}$, are

$$F_{X,Y}(x,y|y_a < Y \le y_b) = \begin{cases} 0 & y \le y_a \\ \dfrac{F_{X,Y}(x,y) - F_{X,Y}(x,y_a)}{F_Y(y_b) - F_Y(y_a)} & y_a < y \le y_b \\ \dfrac{F_{X,Y}(x,y_b) - F_{X,Y}(x,y_a)}{F_Y(y_b) - F_Y(y_a)} & y_b < y \end{cases}$$

and

$$f_{X,Y}(x,y|\,y_a < Y \le y_b) = \begin{cases} 0 & y \le y_a \quad \text{and} \quad y > y_b \\ \dfrac{f_{X,Y}(x,y)}{F_Y(y_b) - F_Y(y_a)} & y_a < y \le y_b \end{cases}$$

***4.4-9.** Assume that transoceanic aircraft arrive at a random point x (value of random variable X) within a strip of coastal region of width 10 km centered on a small city. Aircraft altitude at the time of arrival is not more than 25 km and is a random variable Y. If X and Y have the joint density of Problem 4.3-23, find the probability density of arrival altitude, given that aircraft arrive on one side of the city. Repeat for arrivals on the other side of the city.

***4.4-10.** Work Problem 4.4-9 except find the probability density of arrival point X given that arrival altitude is above 10 km.

4.5-1. Random variables X and Y are *joint gaussian and normalized* if

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right] \quad \text{where} \quad -1 \le \rho \le 1$$

(*a*) Show that the marginal density functions are

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \qquad f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp(-y^2/2)$$

(*Hint*: Complete the square and use the fact that the area under a gaussian density is unity.)

(*b*) Are X and Y statistically independent?

4.5-2. By use of the joint density of Problem 4.5-1, show that

$$f_X(x|Y = y) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{(x-\rho y)^2}{2(1-\rho^2)}\right]$$

4.5-3. Given the joint distribution function

$$F_{X,Y}(x, y) = u(x)u(y)[1 - e^{-ax} - e^{-ay} + e^{-a(x+y)}]$$

find:

(*a*) The conditional density functions $f_X(x|Y = y)$ and $f_Y(y|X = x)$.

(*b*) Are the random variables X and Y statistically independent?

4.5-4. For two independent random variables X and Y show that

$$P\{Y \le X\} = \int_{-\infty}^{\infty} F_Y(x) f_X(x)\, dx$$

or

$$P\{Y \le X\} = 1 - \int_{-\infty}^{\infty} F_X(y) f_Y(y)\, dy$$

4.5-5. Two random variables X and Y have a joint probability density function

$$f_{X,Y}(x, y) = \begin{cases} \frac{5}{16}x^2 y & 0 < y < x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

(*a*) Find the marginal density functions of X and Y.

(*b*) Are X and Y statistically independent?

4.5-6. Determine if random variables X and Y of Problem 4.4-6 are statistically independent.

4.5-7. Determine if X and Y of Problem 4.3-17 are statistically independent.

4.5-8. The joint density of four random variables $X_i, i = 1, 2, 3,$ and 4, is

$$f_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4) = \prod_{i=1}^{4} \exp(-2|x_i|)$$

Find densities
(*a*) $f_{X_1,X_2,X_3}(x_1, x_2, x_3|x_4)$
(*b*) $f_{X_1,X_2}(x_1, x_2|x_3, x_4)$, and
(*c*) $f_{X_1}(x_1|x_2, x_3, x_4)$

4.5-9. Assume that random variables X and Y have the joint density

$$f_{X,Y}(x, y) = \begin{cases} k\cos^2\left(\frac{\pi}{2}xy\right) & -1 < x < 1 \text{ and } -1 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

where

$$k = \frac{\pi/2}{\pi + \mathrm{Si}(\pi)} \approx 0.315$$

and the *sine integral* is defined by

$$\mathrm{Si}(x) = \int_0^x \frac{\sin(\xi)}{\xi}\, d\xi$$

(see Abramowitz and Stegun, 1964). By use of (4.5-4), determine whether X and Y are statistically independent.

4.5-10. Random variables X and Y have the joint density

$$f_{X,Y}(x, y) = \tfrac{1}{12}u(x)u(y)e^{-(x/4)-(y/3)}$$

Find:
(*a*) $P\{2 < X \le 4, -1 < Y \le 5\}$ and
(*b*) $P\{0 < X < \infty, -\infty < Y \le -2\}$

4.6-1. Show, by use of (4.4-13), that the area under $f_Y(y|x)$ is unity.

***4.6-2.** Two random variables R and Θ have the joint density function

$$f_{R,\Theta}(r, \theta) = \frac{u(r)[u(\theta) - u(\theta - 2\pi)]r}{2\pi} e^{-r^2/2}$$

(*a*) Find $P\{0 < R \le 1, 0 < \Theta \le \pi/2\}$.
(*b*) Find $f_R(r|\Theta = \pi)$.
(*c*) Find $f_R(r|\Theta \le \pi)$ and compare to the result found in part (*b*), and explain the comparison.

4.6-3. Random variables X and Y have respective density functions

$$f_X(x) = \frac{1}{a}[u(x) - u(x - a)]$$

$$f_Y(y) = bu(y)e^{-by}$$

where $a > 0$ and $b > 0$. Find and sketch the density function of $W = X + Y$ if X and Y are statistically independent.

4.6-4. Random variables X and Y have respective density functions

$$f_X(x) = 0.1\delta(x-1) + 0.2\delta(x-2) + 0.4\delta(x-3) + 0.3\delta(x-4)$$
$$f_Y(y) = 0.4\delta(y-5) + 0.5\delta(y-6) + 0.1\delta(y-7)$$

Find and sketch the density function of $W = X + Y$ if X and Y are independent.

4.6-5. Find and sketch the density function of $W = X + Y$, where the random variable X is that of Problem 4.6-3 with $a = 5$ and Y is that of Problem 4.6-4. Assume X and Y are independent.

4.6-6. Find the density function of $W = X + Y$, where the random variable X is that of Problem 4.6-4 and Y is that of Problem 4.6-3. Assume X and Y are independent. Sketch the density function for $b = 1$ and $b = 4$.

***4.6-7.** Three statistically independent random variables X_1, X_2, and X_3 all have the same density function

$$f_{X_i}(x_i) = \frac{1}{a}[u(x_i) - u(x_i - a)] \qquad i = 1, 2, 3$$

Find and sketch the density function of $Y = X_1 + X_2 + X_3$ if $a > 0$ is constant.

4.6-8. If the difference $W = X - Y$ is formed instead of the sum in (4.6-1), develop the probability density of W. Compare the result with (4.6-5). Is the density still a convolution of the densities of X and Y? Discuss.

4.6-9. Statistically independent random variables X and Y have respective densities

$$f_X(x) = [u(x+12) - u(x-12)][1 - |x/12|]/12$$
$$f_Y(y) = (1/4)u(y)\exp(-y/4)$$

Find the probabilities of the events:
(*a*) $\{Y \le 8 - (2|X|/3)\}$, and (*b*) $\{Y \le 8 + (2|X|/3)\}$.
Compare the two results.

4.6-10. Statistically independent random variables X and Y have respective densities

$$f_X(x) = 5u(x)\exp(-5x)$$
$$f_Y(y) = 2u(y)\exp(-2y)$$

Find the density of the sum $W = X + Y$.

***4.6-11.** N statistically independent random variables X_i, $i = 1, 2, \ldots, N$, all have the same density

$$f_{X_i}(x_i) = au(x_i)\exp(-ax_i)$$

where $a > 0$ is a constant. Find an expression for the density of the sum $W = X_1 + X_2 + \cdots + X_N$ for any N.

*4.6-12. Statistically independent random variables X and Y have probability densities

$$f_X(x) = \frac{3}{2a^3}[u(x+a) - u(x-a)]x^2 \qquad a > \pi/2$$

$$f_Y(y) = \frac{1}{2}\ \text{rect}\left(\frac{y}{\pi}\right)\cos(y)$$

Find the exact probability density of the sum $W = X + Y$.

4.6-13. The probability density functions of two statistically independent random variables X and Y are

$$f_X(x) = \tfrac{1}{2}u(x-1)e^{-(x-1)/2}$$

$$f_Y(y) = \tfrac{1}{4}u(y-3)e^{-(y-3)/4}$$

Find the probability density of the sum $W = X + Y$.

4.6-14. Statistically independent random variables X and Y have probability densities

$$f_X(x) = \begin{cases} \frac{3}{32}(4 - x^2) & -2 \le x \le 2 \\ 0 & \text{elsewhere in } x \end{cases}$$

$$f_Y(y) = \tfrac{1}{2}[u(y+1) - u(y-1)]$$

Find the exact probability density of the sum $W = X + Y$.

*4.7-1. Find the exact probability density for the sum of two statistically independent random variables each having the density

$$f_X(x) = 3[u(x+a) - u(x-a)]x^2/2a^3$$

where $a > 0$ is a constant. Plot the density along with the gaussian approximation (to the density of the sum) that has variance $2\sigma_X^2$ and mean $2\bar{X}$. Is the approximation a good one?

*4.7-2. Work Problem 4.7-1 except assume

$$f_X(x) = (1/2)\cos(x)\,\text{rect}\,(x/\pi)$$

*4.7-3. Three statistically independent random variables X_1, X_2, and X_3 are defined by

$$\bar{X}_1 = -1 \qquad \sigma_{X_1}^2 = 2.0$$

$$\bar{X}_2 = 0.6 \qquad \sigma_{X_2}^2 = 1.5$$

$$\bar{X}_3 = 1.8 \qquad \sigma_{X_3}^2 = 0.8$$

Write the equation describing the gaussian approximation for the density function of the sum $X = X_1 + X_2 + X_3$. (*Hint*: Refer to the text on the central limit theorem.)

*4.7-4. Two statistically independent random variables X_1 and X_2 have the same probability density given by

$$f_{X_i}(x_i) = \begin{cases} 2x_i/a^2 & 0 \le x_i < a \\ 0 & \text{elsewhere in } x_i \end{cases}$$

for $i = 1$ and 2, where $a > 0$ is a constant.