PROBABILITY, RANDOM VARIABLES, AND RANDOM SIGNAL PRINCIPLES

Second Edition

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

Professor of Electrical Engineering
University of Florida

McGraw-Hill, Inc.

New York St. Louis San Francisco Auckland Bogotá Caracas Hamburg
Lisbon London Madrid Mexico Milan Montreal New Delhi Paris
San Juan São Paulo Singapore Sidney Tokyo Toronto

TO MY MOTHER
Maida Erlene Denton Dials

AND STEPFATHER
Ralph Phillip Dials

This book was set in Times Roman by Santype International Limited;
the editor was Sanjeev Rao; the cover was designed by Rafael Hernandez;
the production supervisors were Leroy A. Young and Fred Schulte.
Project supervision was done by Albert Harrison, Harley Editorial Services.
R. R. Donnelley & Sons Company was printer and binder.

PROBABILITY, RANDOM VARIABLES, AND RANDOM SIGNAL PRINCIPLES

567890 DOC/DOC 9987654321

ISBN 0-07-049219-0

Library of Congress Cataloging-in-Publication Data

Peebles, Peyton Z.
Probability, random variables, and random signal principles.

(McGraw-Hill series in electrical engineering. Communications and information theory)
Bibliography: p.
1. Probabilities. 2. Random variables. 3. Signal theory (Telecommunication) I. Title. II. Series.
TA340.P43 1987 519.2 86-10550
ISBN 0-07-049219-0
ISBN 0-07-049220-4 (solutions manual)

CONTENTS

* Star indicates more advanced material.

PREFACE TO THE SECOND EDITION

Because the first edition of this book was well received by the academic and engineering community, a special attempt was made in the second edition to include only those changes that seemed to clearly improve the book's use in the classroom. Most of the modifications were included only after obtaining input from several users of the book.

Except for a few minor corrections and additions, just six significant changes were made. Only two, a new section on the central limit theorem and one on gaussian random processes, represent modification of the original text. A third change, a new chapter (10) added at the end of the book, serves to illustrate a number of the book's theoretical principles by applying them to problems encountered in practice. A fourth change is the addition of Appendix F, which is a convenient list of some useful probability densities that are often encountered.

The remaining two changes are probably the most significant, especially for instructors using the book. First, the number of examples that illustrate the topics discussed has been increased by about 30 percent (over 85 examples are now included). These examples were carefully scattered throughout the text in an effort to include at least one in each section where practical to do so. Second, over 220 new student exercises (problems) have been added at the ends of the chapters (a 54 percent increase).

The book now contains 630 problems and a complete solutions manual is available to instructors from the publisher. This addition was in response to instructors that had used most of the exercises in the first edition. For these instructors' convenience in identifying the new problems, they are listed in each chapter as "Additional Problems."

All other aspects of the book, such as its purpose (a textbook), intended audience (juniors, seniors, first-year graduate students), level, and style of presentation, remain as before.

I would like to thank D. I. Starry for her excellent work in typing the manuscript and the University of Florida for making her services available. Finally, I am again indebted to my wife, Barbara, for her selfless efforts in helping me proofread the book. If the number of in-print errors is small, it is greatly due to her work.

Peyton Z. Peebles, Jr.

PREFACE TO THE FIRST EDITION

This book has been written specifically as a textbook with the purpose of introducing the principles of probability, random variables, and random signals to either junior or senior engineering students.

The *level* of material included in the book has been selected to apply to a typical undergraduate program. However, a small amount of more advanced material is scattered throughout to serve as stimulation for the more advanced student, or to fill out course content in schools where students are at a more advanced level. (Such topics are keyed by a star ★.) The *amount* of material included has been determined by my desire to fit the text to courses of up to one semester in length. (More is said below about course structure.)

The *need* for the book is easily established. The engineering applications of probability concepts have historically been taught at the graduate level, and many excellent texts exist at that level. In recent times, however, many colleges and universities are introducing these concepts into the undergraduate curricula, especially in electrical engineering. This fact is made possible, in part, by refinements and simplifications in the theory such that it can now be grasped by junior or senior engineering students. Thus, there is a definite need for a text that is clearly written in a manner appealing to such students. I have tried to respond to this need by paying careful attention to the organization of the contents, the development of discussions in simple language, and the inclusion of text examples and many problems at the end of each chapter. The book contains over 400 problems and a solutions manual for all problems is available to instructors from the publisher.

Many of the examples and problems have purposely been made very simple in an effort to instill a sense of accomplishment in the student, which, hopefully, will provide the encouragement to go on to the more challenging problems. Although emphasis is placed on examples and problems of electrical engineering, the concepts and theory are applicable to all areas of engineering.

The International System of Units (SI) has been used primarily throughout the text. However, because technology is presently in a transitional stage with regard to measurements, some of the more established customary units (gallons, °F, etc.) are also utilized; in such instances, values in SI units follow in parentheses.

The *student background* required to study the book is only that typical of junior or senior engineering students. Specifically, it is assumed the student has been introduced to multivariable calculus, Fourier series, Fourier transforms, impulse functions, and some linear system theory (transfer function concepts, especially). I recognize, however, that students tend to forget a fair amount of what is initially taught in many of these areas, primarily through lack of opportunity to apply the material in later courses. Therefore, I have inserted short reviews of some of these required topics. These reviews are occasionally included in the text, but, for the most part, exist in appendixes at the end of the book.

The *order of the material* is dictated by the main topic. Chapter 1 introduces probability from the axiomatic definition using set theory. In my opinion this approach is more modern and mathematically correct than other definitions. It also has the advantage of creating a better base for students desiring to go on to graduate work. Chapter 2 introduces the theory of a single random variable. Chapter 3 introduces operations on one random variable that are based on statistical expectation. Chapter 4 extends the theory to several random variables, while Chapter 5 defines operations with several variables. Chapters 6 and 7 introduce random processes. Definitions based on temporal characterizations are developed in Chapter 6. Spectral characterizations are included in Chapter 7.

The remainder of the text is concerned with the response of linear systems with random inputs. Chapter 8 contains the general theory, mainly for linear time-invariant systems; while Chapter 9 considers specific optimum systems that either maximize system output signal-to-noise ratio or minimize a suitably defined average error.

Finally, the book closes with a number of appendixes that contain material helpful to the student in working problems, in reviewing background topics, and in the interpretation of the text.

The book can profitably be used in curricula based on either the quarter or the semester system. At the University of Tennessee, a *one-quarter undergraduate course* at the junior level has been successfully taught that covers Chapters 1 through 8, except for omitting Sections 2.6, 3.4, 4.4, 8.7 through 8.9, and all starred material. The class met three hours per week.

A *one-semester undergraduate course* (three hours per week) can readily be structured to cover Chapters 1 through 9, omitting all starred material except that in Sections 3.3, 5.3, 7.4, and 8.6.

Although the text is mainly developed for the undergraduate, I have also successfully used it in a *one-quarter graduate course* (first-year, three hours per week) that covers Chapters 1 through 7, including all starred material.

It should be possible to cover the entire book, including all starred material, in a *one-semester graduate course* (first-year, three hours per week).

I am indebted to many people who have helped make the book possible. Drs. R. C. Gonzalez and M. O. Pace read portions of the manuscript and suggested a number of improvements. Dr. T. V. Blalock taught from an early version of the manuscript, independently worked a number of the problems, and provided various improvements. I also extend my appreciation to the Advanced Book Program of Addison-Wesley Publishing Company for allowing me to adapt and use several of the figures from my earlier book *Communication System Principles* (1976), and to Dr. J. M. Googe, head of the electrical engineering department of the University of Tennessee, for his support and encouragement of this project. Typing of the bulk of the manuscript was ably done by Ms. Belinda Hudgens; other portions and various corrections were typed by Kymberly Scott, Sandra Wilson, and Denise Smiddy. Finally, I thank my wife, Barbara, for her aid in proofreading the entire book.

Peyton Z. Peebles, Jr.

CHAPTER

ONE

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 between the

† 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.

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 adopt the axiomatic approach in this book.

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

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

‡ 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).

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, 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 explictly 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.

Two sets, A and B, are called *disjoint* or *mutually exclusive* if they have no common elements. (incompatibles) → $A \cap B = \phi$

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

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.

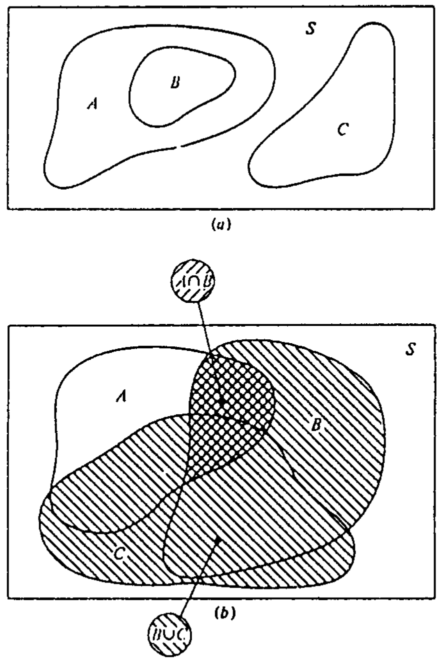


Figure 1.2-1 Venn diagrams. (*a*) Illustration of subsets and mutually exclusive sets, and (*b*) illustration of intersection and union of sets. [*Adapted from Peebles, (1976) with permission of publishers Addison-Wesley, Advanced Book Program.*]

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 1.2-1*a*. 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 $A = B$.

The *difference* of two sets A and B, denoted $A - B$, is the set containing all

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

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:

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

$A \cup C = \{1, 3, 4, 5, 6, 7, 8, 12\}$ $\qquad$ $A \cap C = \{1, 3\}$

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

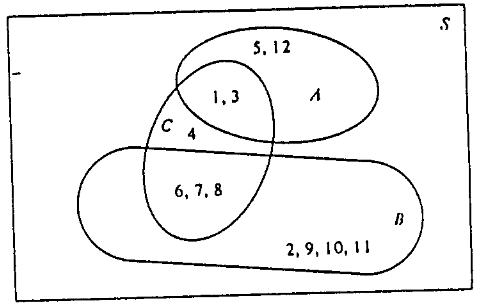


Figure 1.2-2 Venn diagram applicable to Example 1.2-1.

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).

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,

$$\overline{(A \cup B)} = \bar{A} \cap \bar{B} \tag{1.2-12}$$
$$\overline{(A \cap B)} = \bar{A} \cup \bar{B} \tag{1.2-13}$$

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

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 (Papoulis, 1965) 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).

1.3 PROBABILITY INTRODUCED THROUGH SETS

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

† 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.

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 $1/6$ (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 with six elements. Such sample spaces are said to be *discrete* and finite. The sample space can also be 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 coutably 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.

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

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

for all $m \neq n = 1, 2, \ldots, N$, with N possibly infinite. The axiom states that the

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

probability of the event equalto the union of any number of mutually exclusive events is equal to the sum of tle individual event probabilities.

An example should help give a physical pictire 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.

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, *probablities* were assigned to the defined events to mathematically account for the randon 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 $P(A_{ij}) = {}^1\!/_{36}$. 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

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

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

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

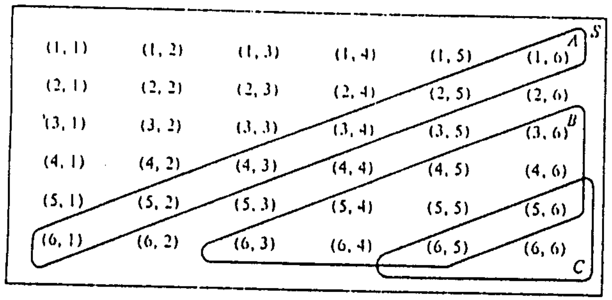


Figure 1.3-1 Sample space applicable to Example 1.3-2.

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

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 from 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

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

The probability $P(A \mid B)$ 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 $P(A \mid B) = 0$.

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). $P(A \mid B)$ 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 $S = A$:

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

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 $P(A \cup C \mid B) = P(A \mid B) + P(C \mid B)$ 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)

$$P[(A \cup C) \mid B] = \frac{P[(A \cup C) \cap B]}{P(B)} = \frac{P(A \cap B)}{P(B)} + \frac{P(C \cap B)}{P(B)}$$
$$= P(A \mid B) + P(C \mid B) \tag{1.4-7}$$

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

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

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

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

† 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.

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

By using (1.4-4) the conditional probabilities become

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

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

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

$P(A \mid B) = P(47\ \Omega \mid 5\%)$ is the probability of drawing a 47-Ω resistor given that the resistor drawn is 5%. $P(A \mid C) = P(47\ \Omega \mid 100\ \Omega)$ 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, $P(B \mid C) = P(5\% \mid 100\ \Omega)$ 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

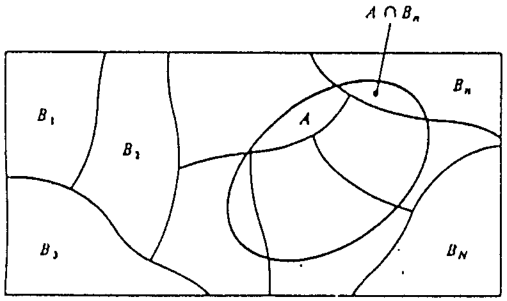
$$B_m \cap B_n = \varnothing \qquad m \neq n = 1, 2, \ldots, N \tag{1.4-8}$$

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

We shall prove that

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

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



$\bigcup_{n=1}^{N} B_n = S, \; B_m \cap B_n = \varnothing$ for all $m \neq n$

Figure 1.4-1 Venn diagram of N mutually exclusive events B_n and another 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

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

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

$$P(B_n \mid A) = \frac{P(B_n \cap A)}{P(A)} \tag{14-13}$$

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

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

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

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

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

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

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

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

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

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

$$P(A_1 \mid B_1) = 0.9$$
$$P(A_2 \mid B_1) = 0.1$$

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

$$P(A_1 \mid B_2) = 0.1$$
$$P(A_2 \mid B_2) = 0.9$$

In either case, $P(A_1 \mid B_i) + P(A_2 \mid B_i) = 1$ 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

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

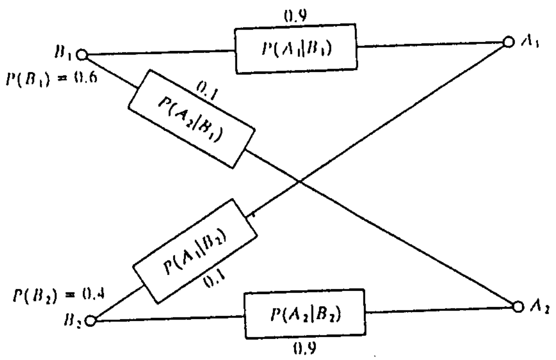


Figure 1.4-2 Binary symmetric communication system diagrammatical model applicable to Example 1.4-2.

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

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

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

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

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

These last two numbers are probabilities of system error while $P(B_1 \mid A_1)$ and $P(B_2 \mid A_2)$ 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 $P(A \mid B_n)$ 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 $P(B_n \mid A)$ are called *a posteriori probabilities*, since they apply after the experiment's performance when some event A is obtained.

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.

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

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

for statistically independent events. We also have

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

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 $P(A) = 4/52$, $P(B) = 8/52$, and $P(C) = 13/52$.

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), $P(A \cap C) = 1/52$, and $P(B \cap C) = 2/52$.

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

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

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

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

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

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 supporting examples are relatively easy to construct. The reader might try this exercise.

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

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

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

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

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

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

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

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

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 pressure 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

$$S = S_1 \times S_2 \qquad (1.6\text{-}1)$$

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

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.

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

$$S = S_1 \times S_2 \times \cdots \times S_N \qquad (1.6\text{-}2)$$

and it is the set of all ordered N-tuples

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

*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

$$C = A \times B \qquad (1.6\text{-}4)$$

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

$$s_1 \in A \quad \text{and} \quad s_2 \in B \qquad (1.6\text{-}5)$$

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

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

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

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

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).

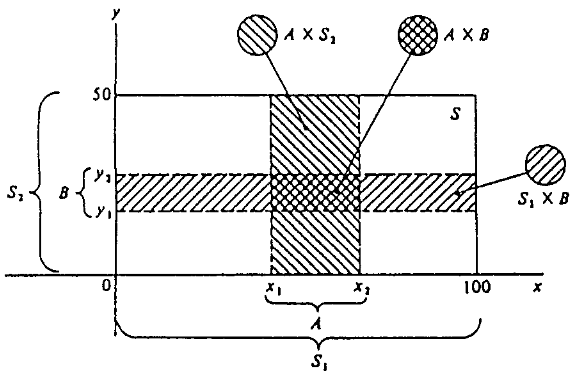


Figure 1.6-1 A combined sample space for two subexperiments.

*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.

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. From combinatorial analysis, the number is known to be

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

The quantity $\binom{N}{k}$ is called the *binomial coefficient*. It is sometimes given the symbol C_k^N.

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?

† 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.

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

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

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

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

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

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 bateria. 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 $p = P\{\text{good culture}\} = {}^{94}/_{100} = 0.94$. Here

$$\begin{aligned} P\{\text{valid datum point}\} &= P\{3 \text{ good cultures}\} + P\{4 \text{ good cultures}\} \\ &= \binom{4}{3}(0.94)^3(1-0.94)^1 + \binom{4}{4}(0.94)^4(1-0.94)^0 \approx 0.98 \end{aligned}$$

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.

PROBLEMS

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

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

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

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

1-4 For each set of Problem 1-3, determine if it is equal to, or a subset of, any of the other sets.

1-5 State every possible subset of the set of letters $\{a, b, c, d\}$.

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).

***1-7** Prove that a set with N elements has 2^N subsets.

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.

1-9 Show that $C \subset A$ if $C \subset B$ and $B \subset A$.

1-10 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*) $A \cap B$

1-11 A univeral 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*) $\bar{A} \cap B$

1-12 Using Venn diagrams for three sets A, B, and 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*) $\overline{(A \cup B)} \cap C$

1-13 Sketch a Venn diagram for three events where $A \cap B \ne \varnothing$, $B \cap C \ne \varnothing$, $C \cap A \ne \varnothing$, but $A \cap B \cap C = \varnothing$.

1-14 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*) $\overline{(A \cap B \cap C)} = \bar{A} \cup \bar{B} \cup \bar{C}$

1-15 Use Venn diagrams to prove De Morgan's laws $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$ and $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$.

1-16 A universal set is $S = \{-20 < s \le -4\}$. If $A = \{-10 \le s \le -5\}$ 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 $A \cap B \cap C$?

1-17 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.

1-18 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 $A \cap B$.

1-19 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?

1-20 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?

1-21 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*) $(A \cup B) \cap C$

1-22 Let A be an arbitrary event. Show that $P(\bar{A}) = 1 - P(A)$.

1-23 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 $P(C) = 1 - P(A) - P(B)$.

1-24 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?

1-25 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?

1-26 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?

1-27 An ordinary 52-card deck is thoroughly shuffled. You are dealt four cards up. What is the probability that all four cards are sevens?

1-28 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 $P(D)$, $P(E)$, $P(D \cap E)$, $P(D \mid E)$, and $P(E \mid D)$.

1-29 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.

1-30 In three boxes there are capacitors as shown in Table P1-30. 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 P1-30 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

1-31 For Problem 1-30, list the nine conditional probabilities of capacitor selection, given certain box selections.

1-32 Rework Example 1.4-2 if $P(B_1) = 0.6$, $P(B_2) = 0.4$, $P(A_1 \mid B_1) = P(A_2 \mid B_2) = 0.95$, and $P(A_2 \mid B_1) = P(A_1 \mid B_2) = 0.05$.

1-33 Rework Example 1.4-2 if $P(B_1) = 0.7$, $P(B_2) = 0.3$, $P(A_1 \mid B_1) = P(A_2 \mid B_2) = 1.0$, and $P(A_2 \mid B_1) = P(A_1 \mid B_2) = 0$. What type of channel does this system have?

1-34 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?

1-35 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?

1-36 Determine whether the three events A, B, and C of Example 1.4-1 are statistically independent.

1-37 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.

*1-38 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 $\bar{A}_2$

*1-39 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.

*1-40 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 $B = \{5 < y \le 40\}$

1-41 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?

1-42 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 that three or more out of six spacecraft will land in the prescribed zone. Is the program successful?

1-43 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.

ADDITIONAL PROBLEMS

1-44 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 if $B \subset A$.

1-45 A set A has three elements a_1, a_2, and a_3. Determine all possible subsets of A.

1-46 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*) $(A \cap B \cap \bar{C}) \cup (\bar{B} \cap C \cap D)$.

1-47 A universal set S is comprised 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*) $C \cap B \cap \bar{A}$.

1-48 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 $A \cup B$?

1-49 Prove that DeMorgan's law (1.2-13) can be extended to N events A_i, $i = 1, 2, \ldots, N$ as follows

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

1-50 Work Problem 1-49 for (1.2-12) to prove

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

1-51 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.

1-52 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?

***1-53** 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 $P(A_i | B_j) = 0.1$, $i \ne j$, 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.

1-54 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.)

1-55 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.

1-56 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.

1-57 The local manager of the airline of Problem 1-56 desires to make sure that 90% of flights leave on time. 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?

1-58 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?

***1-59** 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.

CHAPTER

TWO

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

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

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.

† Complex random variables are considered in Chapter 5.

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 $\{0 < x \leq 144\}$.

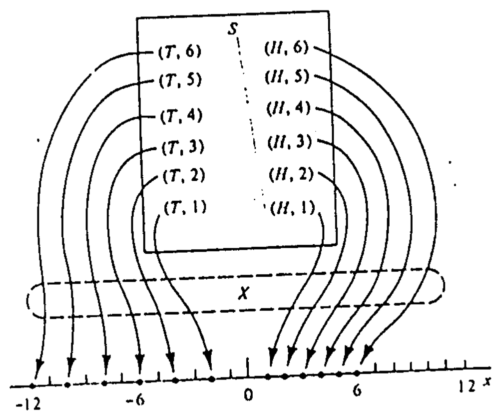


Figure 2.1-1 A random variable mapping of a sample space.

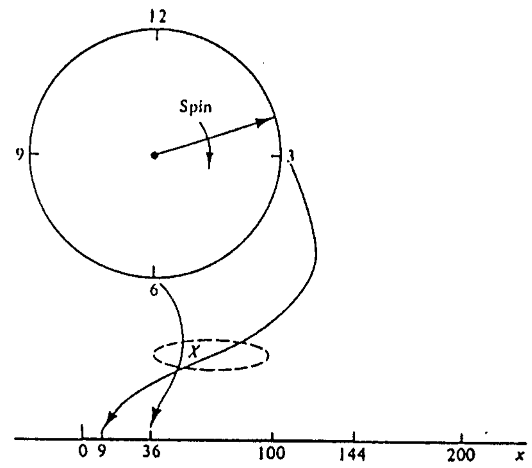


Figure 2.1-2 Mapping applicable to Example 2.1-2.

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 $X = 2$.

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.

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.

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:†

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

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

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

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

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

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

† We use the notation x^+ to imply $x + \varepsilon$ where $\varepsilon > 0$ is infinitesimally small; that is, $\varepsilon \to 0$.

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 $\{X \le x_1\}$ and $\{x_1 < X \le x_2\}$ are mutually exclusive, so the probability of the event $\{X \le x_2\} = \{X \le x_1\} \cup \{x_1 < X \le x_2\}$ is the sum of the probabilities $P\{X \le x_1\}$ and $P\{x_1 < X \le x_2\}$. 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.

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 2.2-1a. 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†

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

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

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

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 $F_X(x)$ for this discrete random variable.

---

† This definition differs slightly from (A-5) by including the equality so that $u(x)$ satisfies (2.2-2f).

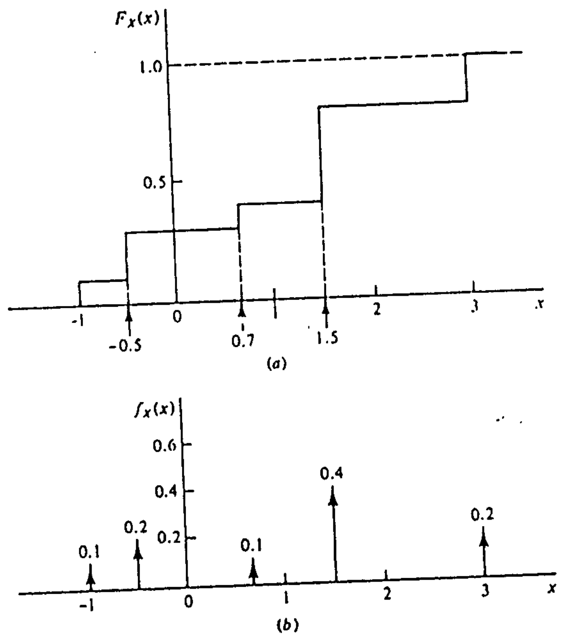


Figure 2.2-1 Distribution function (a) and density function (b) applicable to the discrete random variable of Example 2.2-1. [*Adapted from Peebles (1976) with permission of publishers Addison-Wesley, Advanced Book Program.*]

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.

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 $\{X \le 0\}$ is 0 because there are no sample space points in this set. For $0 < x \le 12$ the probability of $\{0 < X \le x\}$ will increase linearly with x for a fair wheel. Thus, $F_X(x)$ will behave as shown in Figure 2.2-2a.

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

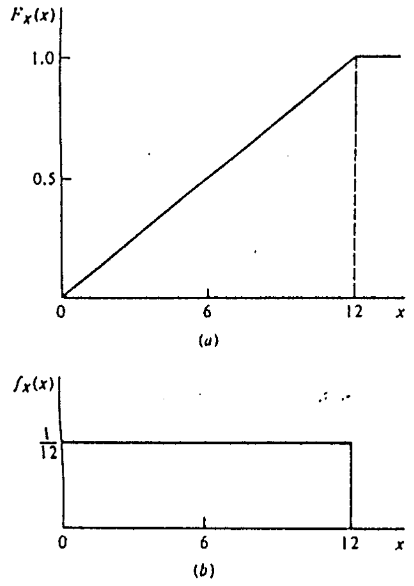


Figure 2.2-2 Distribution function (*a*) and density function (*b*) applicable to the continuous random variable of Example 2.2-2. [*Adapted from Peebles (1976) with permission of publishers Addison-Wesley, Advanced Book Program.*]

2.3 DENSITY FUNCTION

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

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

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*). We shall assume that the number of points where $F_X(x)$ is not differentiable is countable.

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

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

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. 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 2.3-1*a* 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 2.3-1*b*.

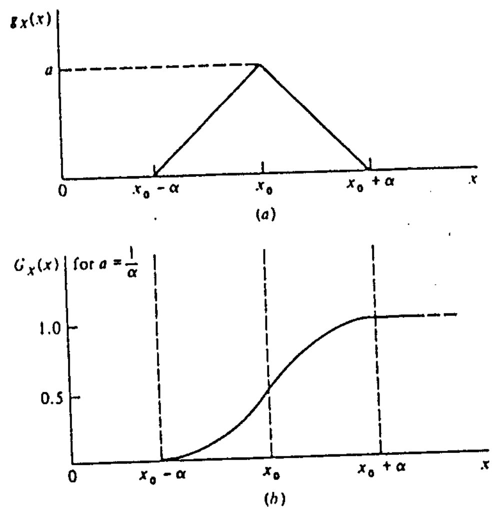


Figure 2.3-1 A possible probability density function (*a*) and a distribution function (*b*) applicable to Example 2.3-1.

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 $a = 1/\alpha = 1/5$. From the earlier work

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

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

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

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

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

2.4 THE GAUSSIAN RANDOM VARIABLE

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

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

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

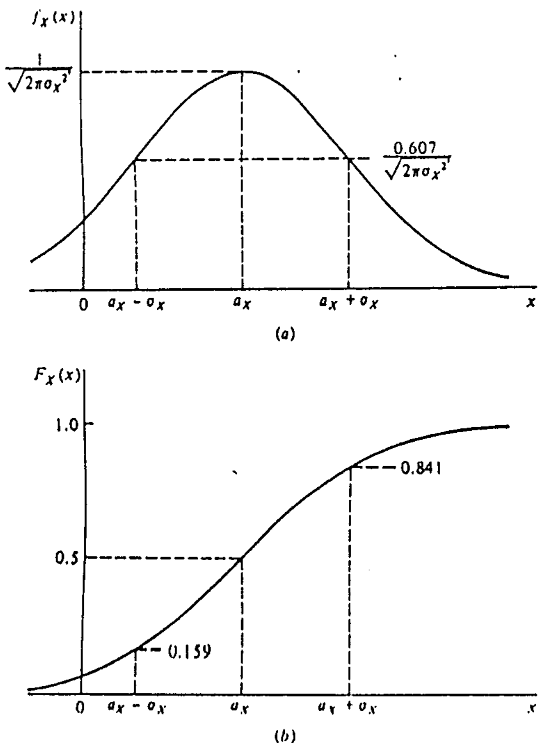


Figure 2.4-1 Density (a) and distribution (b) functions of a gaussian random variable.

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 $x = a_X - \sigma_X$.

The gaussian density is the most important ot all densities. It enters into nearly all areas of engineering and science. We shall encounter the gaussian random variable frequently in later work when we discuss some important types of noise.

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

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

This integral has no known closed-form solution and must be evaluated by numerical 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 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

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

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

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

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

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

in (2.4-2) to obtain

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

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

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

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 $\{X \leq 5.5\}$ for a gaussian random variable having $a_X = 3$ and $\sigma_X = 2$.

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

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

By using the table in Appendix B

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

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:

$$P\{X > 2750\} = 1 - P\{X \leq 2750\} = 1 - F_X(2750)$$

$$= 1 - F\left(\frac{2750 - 1830}{460}\right) = 1 - F(2.0)$$

$$= 1 - 0.9772 = 0.0228$$

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

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 $p = 0.25$.

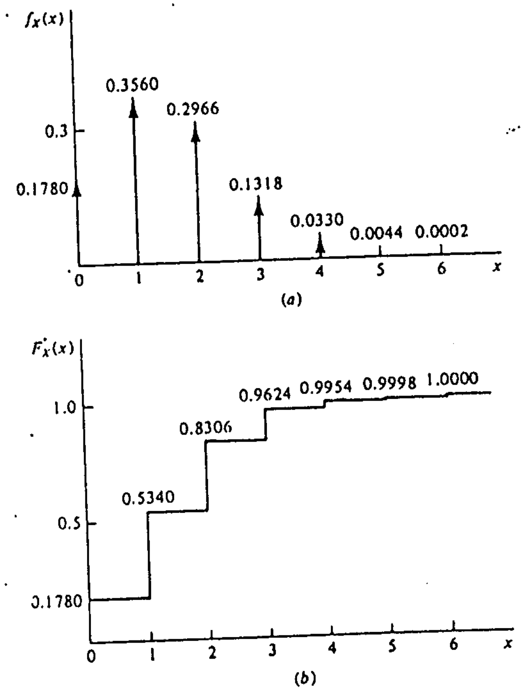


Figure 2.5-1 Binomial density (*a*) and distribution (*b*) functions for the case $N = 6$ and $p = 0.25$.

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

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

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 $\lambda = {}^{50}/_{60}$ cars/minute and $T = 1$ minute, we have $b = {}^{5}/_{6}$. On using (2.5-5)

$$\text{Probability of a waiting line} = 1 - F_X(1) - F_X(0)$$
$$= 1 - e^{-5/6}\left[1 + \frac{5}{6}\right] = 0.2032$$

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:

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

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

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.

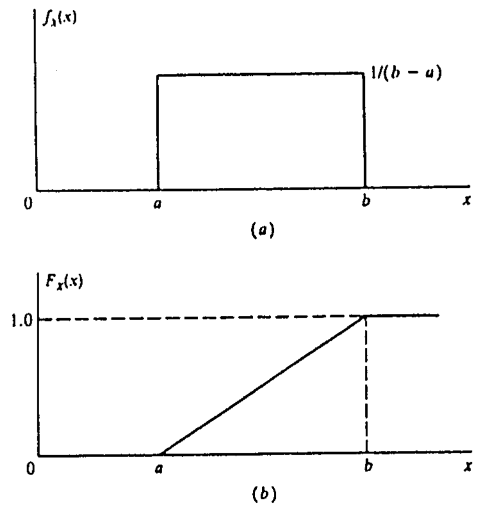


Figure 2.5-2 Uniform probability density function (*a*) and its distribution function (*b*).

Exponential

The *exponential* density and distribution functions are:

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

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

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 approximately 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

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

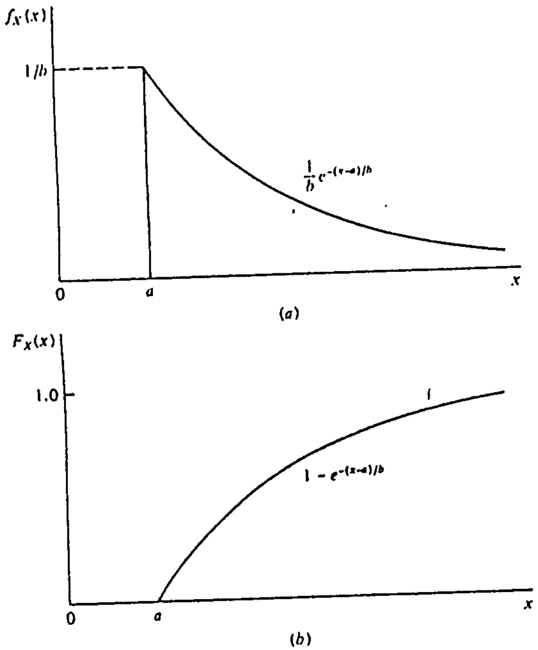


Figure 2.5-3 Exponential density (a) and distribution (b) functions.

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 per cent of the time.

Rayleigh

The *Rayleigh*† density and distribution functions are:

$$f_X(x) = \begin{cases} \frac{2}{b}(x-a)e^{-(x-a)^2/b} & x \ge a \\ 0 & x < a \end{cases} \qquad (2.5\text{-}11)$$

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

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

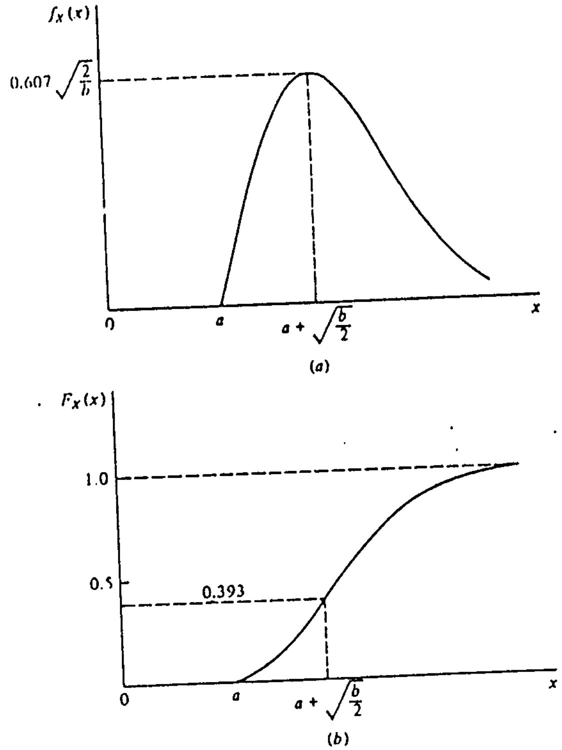


Figure 2.5-4 Rayleigh density (a) and distribution (b) functions.

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

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.

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

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad (2.6\text{-}1)$$

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 $\{X \le x\}$ for the random variable X. The resulting probability $P\{X \le x \mid B\}$ is defined as the *conditional distribution function* of X, which we denote $F_X(x \mid B)$. Thus

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

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

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

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 $F_X(x \mid B)$. In other words, it has the following characteristics:

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

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

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

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

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

(6) $F_X(x^+ \mid B) = F_X(x \mid B)$ (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 $f_X(x \mid B)$, then

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

If $F_X(x \mid B)$ contains step discontinuities, as when X is a discrete or mixed random variable, we assume that impulse functions are present in $f_X(x \mid B)$ 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) $f_X(x \mid B) \ge 0$ (2.6-6a)

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

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

(4) $\displaystyle P\{x_1 < X \le x_2 \mid B\} = \int_{x_1}^{x_2} f_X(x \mid B)\, dx$ (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 $P(B_1) = 2/10$ and $P(B_2) = 8/10$. (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.)

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:

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

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

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

Table 2.6-1 Numbers of colored balls in two boxes

		Box		
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

The conditional probability density $f_X(x \mid B_1)$ becomes

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

By direct integration of $f_X(x \mid B_1)$:

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

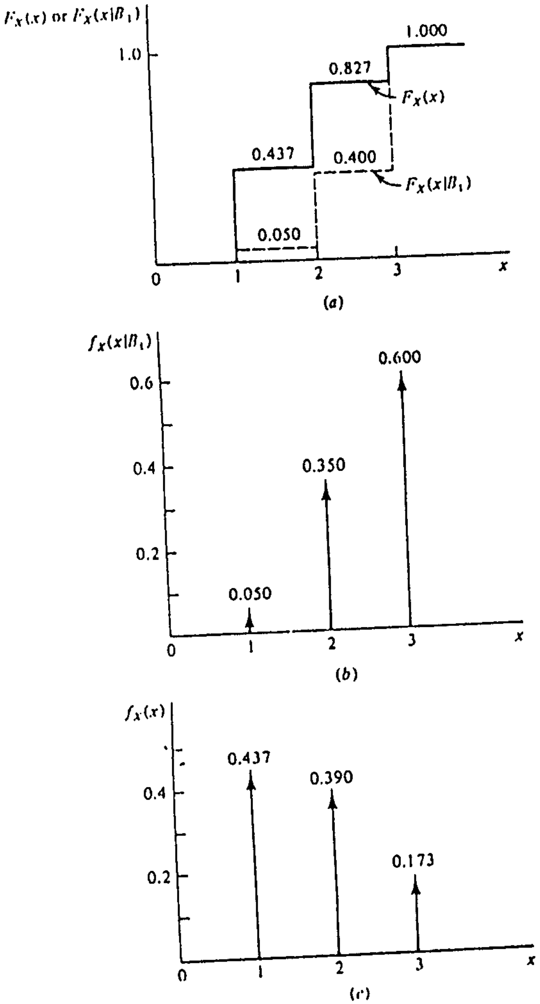


Figure 2.6-1 Distributions (a) and densities (b) and (c) applicable to Example 2.6-1.

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):

$$P(X=1) = P(X=1 \mid B_1)P(B_1) + P(X=1 \mid B_2)P(B_2)$$

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

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

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

Thus

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

and

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

These distributions and densities are plotted in Figure 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

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

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

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

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

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

$$F_X(x \mid X \le b) = \frac{P\{X \le x \cap X \le b\}}{P\{X \le b\}} = \frac{P\{X \le b\}}{P\{X \le b\}} = 1 \qquad b \le x \qquad (2.6\text{-}9)$$

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

$$F_X(x \mid X \le b) = \frac{P\{X \le x \cap X \le b\}}{P\{X \le b\}} = \frac{P\{X \le x\}}{P\{X \le b\}} = \frac{F_X(x)}{F_X(b)} \qquad x < b \qquad (2.6\text{-}10)$$

By combining the last two expressions, we obtain

$$F_X(x \mid X \le b) = \begin{cases} \dfrac{F_X(x)}{F_X(b)} & x < b \\ 1 & b \le x \end{cases} \qquad (2.6\text{-}11)$$

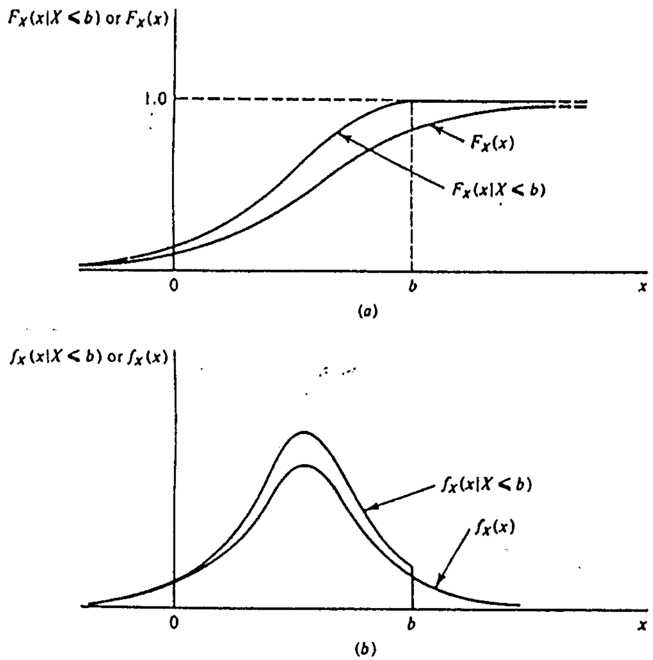


Figure 2.6-2 Possible distribution functions (*a*) and density functions (*b*) applicable to a conditioning event $B = \{X \le b\}$.

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

$$f_X(x \mid X \le b) = \begin{cases} \dfrac{f_X(x)}{F_X(b)} = \dfrac{f_X(x)}{\int_{-\infty}^{b} f_X(x)\,dx} & x < b \\ 0 & x \ge b \end{cases} \qquad (2.6\text{-}12)$$

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

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

$$F_X(x \mid X \le b) \ge F_X(x) \qquad (2.6\text{-}13)$$

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

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

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 2-39).

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 $b = 50$:

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

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

PROBLEMS

2-1 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*) $X = (1 - 3s)^{-1}$

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

2-3 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) $\{4, 3.1, 1, -2\}$

2-4 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.

2-5 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."

2-6 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."

2-7 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.

***2-8** 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.

2-9 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.

2-10 Work Problem 2-9 for a biased coin for which $P\{\text{head}\} = 0.6$.

2-11 Resistor R_2 in Figure P2-11 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.

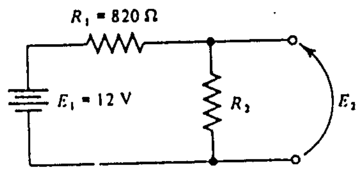


Figure P2-11

2-12 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 $\pm 5\%$ tolerance about the nominal length. What fraction of the production line's output is he purchasing?

2-13 Find and sketch the density and distribution functions for the random variables of parts (a), (b), and (c) in Problem 2-1 if the sample space elements have equal likelihoods of occurrence.

2-14 If temperature in Problem 2-6 is uniformly distributed, sketch the density and distribution functions of the random variable X.

2-15 For the uniform random variable defined by (2.5-7) find:

(a) $P\{0.9a + 0.1b < X \leq 0.7a + 0.3b\}$
(b) $P\{(a + b)/2 < X \leq b\}$

2-16 Determine which of the following are valid distribution functions:

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

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

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

2-17 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*).

† After the French mathematician Marquis Pierre Simon de Laplace (1749–1827).

2-18 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?

***2-19** The amplifier in the master station of Problem 2-18 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?

2-20 Justify that a distribution function $F_X(x)$ satisfies (2.2-2*a*, *b*, *c*).

2-21 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*) $\int_{-\infty}^{\infty} \dfrac{24\delta(x - 2)\, dx}{x^4 + 3x^2 + 2}$

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

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

2-22 Show that the properties of a density function $f_X(x)$, as given by (2.3-6), are valid.

2-23 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*) $P\{X = x_0\}$

2-24 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 $X > 2$?

2-25 Work Problem 2-24 if $a_X = 4$ and $\sigma_X = 2$.

2-26 For the gaussian density function of (2.4-1), show that

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

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

2-27 For the gaussian density function of (2.4-1), show that

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

2-28 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 $a_X = 1000$ Ω and $\sigma_X = 40$ Ω, 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 $a_X = 1050$ Ω (5% shift). What fraction is now rejected?

2-29 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?

2-30 (*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 $I_2 - I_1^2 = b^2$.

2-31 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 $\sqrt{2/b}\exp(-1/2) \approx 0.607\sqrt{2/b}$. 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.)

2-32 Find the value $x = x_0$ of a Rayleigh random variable for which $P\{X \le x_0\} = P\{x_0 < X\}$. This value of x is called the *median* of the random variable.

2-33 The lifetime of a system expressed in weeks is a Rayleigh random variable X for which

$$f_X(x) = \begin{cases} (x/200)e^{-x^2/400} & 0 \le x \\ 0 & x < 0 \end{cases}$$

(*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?

2-34 The *Cauchy*† random variable has the probability density function

$$f_X(x) = \frac{b/\pi}{b^2 + (x - a)^2}$$

† After the French mathematician Augustin Louis Cauchy (1789–1857).

for real numbers $0 < b$ and $-\infty < a < \infty$. Show that the distribution function of X is

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x-a}{b}\right)$$

2-35 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).

2-36 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 $\{0 \leq X \leq 5\}$?

2-37 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.

2-38 Rework Example 2.6-1 to find $f_X(x|B_2)$ and $F_X(x|B_2)$. Sketch the two functions.

*2-39 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

$$f_X(x|a < X \leq b) = \begin{cases} 0 & x < a \\ \dfrac{f_X(x)}{F_X(b)-F_X(a)} = \dfrac{f_X(x)}{\int_a^b f_X(x)\,dx} & a \leq x < b \\ 0 & b \leq x \end{cases}$$

*2-40 Consider the system having a lifetime defined by the random variable X in Problem 2-33. Given that the system will survive beyond 20 weeks, find the probability that it will survive beyond 26 weeks.

ADDITIONAL PROBLEMS

2-41 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.

2-42 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?

2-43 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 $\text{rect}(\cdot)$ 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.

2-44 (*a*) Generalize Problem 2-16(*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?

2-45 Find a constant $b > 0$ so that the function

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

is a valid probability density.

2-46 Given the function

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

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

2-47 A random variable X has the density function

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

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

*2-48 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.)

2-49 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.

2-50 A gaussian random variable X has $a_X = 2$, and $\sigma_X = 2$.

(*a*) Find $P\{X > 1.0\}$.

(*b*) Find $P\{X \le -1.0\}$.

2-51 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?

2-52 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 $A = \{1000 \text{ m} < X \le 3300 \text{ m}\}$ while a second person bets that height will be satisfied by $B = \{2000 \text{ m} < X \le 4200 \text{ m}\}$. A third person bets they are both correct. Find the probabilities that each person will win the bet.

2-53 Let X be a Rayleigh random variable with $a = 0$. Find the probability that X will have values larger than its mode (see Problem 2-31).

2-54 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?

2-55 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 the probability that the radar is still in operation one week later when picked up for maintenance and repairs.

2-56 If the radar of Problem 2-55 is permanently located at the remote site, find the probability that it will be operational as a function of time since its set up.

2-57 A computer undergoes down-time if a certain critical component fails. This component is known to fail at an average rate of once per four weeks. No significant down-time 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 down-time occurring before the ordered components arrive?

(*b*) If the shipment is delayed two weeks what is the probability of significant down-time occurring before the shipment arrives?

*2-58 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?

*2-59 Suppose the depth of water, measured in meters, behind a dam is described by an exponential random variable having a density

$$f_X(x) = (1/13.5) \exp(-x/13.5)$$

There is an emergency overflow at the top of the dam that prevents the depth from exceeding 40.6 m. There is a pipe placed 32.0 m below the overflow (ignore the pipe's finite diameter) that feeds water to a hydroelectric generator.

(*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 the generator will have water to drive it?

(*c*) What is the probability that water will be too low to produce power?

*2-60 In Problem 2-59 find and sketch the distribution and density functions of water depth given that water will be deep enough to generate power but no water is wasted by emergency overflow. Also sketch for comparisons the distribution and density of water depth without any conditions?

*2-61 In Example 2.6-2 a parachuter is an "expert" if he hits the bull's eye. If he falls 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 target what is the probability of being "qualified?"

CHAPTER

THREE

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 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:

$$E[X] = \int_a^\infty \frac{x}{b} e^{-(x-a)/b}\,dx = \frac{e^{a/b}}{b}\int_a^\infty x e^{-x/b}\,dx = a + b$$

If a random variable's density is symmetrical about a line $x = a$, then $E[X] = a$; that is,

$$E[X] = a \qquad \text{if} \qquad f_X(x + a) = f_X(-x + a) \tag{3.1-5}$$

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 (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).

*Conditional Expected Value

If, in (3.1-2), $f_X(x)$ is replaced by the conditional density $f_X(x \mid B)$, where B is any event defined on the sample space, we have the *conditional expected value* of X, denoted $E[X \mid B]$:

$$E[X \mid B] = \int_{-\infty}^{\infty} x f_X(x \mid B)\,dx \tag{3.1-8}$$

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

$$f_X(x \mid X \le b) = \begin{cases} \dfrac{f_X(x)}{\int_{-\infty}^{b} f_X(x)\,dx} & x < b \\ 0 & x \ge b \end{cases} \tag{3.1-10}$$

Thus, by substituting (3.1-10) into (3.1-8):

$$E[X \mid X \le b] = \frac{\int_{-\infty}^{b} x f_X(x)\,dx}{\int_{-\infty}^{b} f_X(x)\,dx} \tag{3.1-11}$$

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

$$\sigma_X^2 = \frac{e^{-(\bar{X}-a)/b}}{b} \int_{a-\bar{X}}^{\infty} \xi^2 e^{-\xi/b}\, d\xi = (a + b - \bar{X})^2 + b^2$$

† 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 the reader can readily verify as an exercise.

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

$$\begin{aligned}\mu_3 &= E[(X - \bar{X})^3] = E[X^3 - 3\bar{X}X^2 + 3\bar{X}^2X - \bar{X}^3] \\ &= \overline{X^3} - 3\bar{X}\overline{X^2} + 2\bar{X}^3 = \overline{X^3} - 3\bar{X}(\sigma_X^2 + \bar{X}^2) + 2\bar{X}^3 \\ &= \overline{X^3} - 3\bar{X}\sigma_X^2 - \bar{X}^3\end{aligned}$$

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.

---

*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 number $-\infty < \omega < \infty$. 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)

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_X(\omega)e^{-j\omega x}\, d\omega \tag{3.3-3}$$

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, 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 3-24.)

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.

---

† 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.

*Moment Generating Function

Another statistical average closely related to the characteristic function is the *moment generating function*, defined by

$$M_X(v) = E[e^{vX}] \tag{3.3-6}$$

where v is a real number $-\infty < v < \infty$. Thus, $M_X(v)$ is given by

$$M_X(v) = \int_{-\infty}^{\infty} f_X(x)e^{vx}\, dx \tag{3.3-7}$$

The main advantage of the moment generating function derives from its ability to give the moments. Moments are related to $M_X(v)$ by the expression:

$$m_n = \frac{d^n M_X(v)}{dv^n}\bigg|_{v=0} \tag{3.3-8}$$

The main disadvantage of the moment generating function, as opposed to the characteristic function, is that it may not exist for all random variables. In fact, $M_X(v)$ exists only if all the moments exist (Davenport and Root, 1958, p. 52).

Example 3.3-2 To illustrate the calculation and use of the moment gener-ating function, let us reconsider the exponential density of the earlier exam-ples. On use of (3.3-7) we have

$$M_X(v) = \int_a^{\infty} \frac{1}{b} e^{-(x-a)/b} e^{vx}\, dx$$
$$= \frac{e^{a/b}}{b} \int_a^{\infty} e^{[v-(1/b)]x}\, dx$$
$$= \frac{e^{av}}{1-bv}$$

In evaluating $M_X(v)$ we have used an integral from Appendix C.

By differentiation we have the first moment

$$m_1 = \frac{dM_X(v)}{dv}\bigg|_{v=0}$$
$$= \frac{e^{av}[a(1-bv)+b]}{(1-bv)^2}\bigg|_{v=0} = a + b$$

which, of course, is the same as previously found.

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 a 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 contin-
uous and T continuous and either monotonically increasing or decreasing with
X; (2) X continuous and T continuous but nonmonotonic; (3) X discrete and T

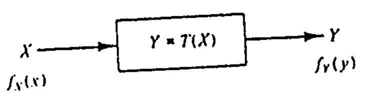


Figure 3.4-1 Transformation of a random variable X to a new random variable Y.

continuous. Note that the transformation in all three cases is assumed contin-
uous. 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 3-32).

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 particular
value y_0 corresponding to the particular value x_0 of X as shown in Figure 3.4-2*a*.
The two numbers are related by

$$y_0 = T(x_0) \quad \text{or} \quad x_0 = T^{-1}(y_0) \tag{3.4-2}$$

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,

$$F_Y(y_0) = P\{Y \le y_0\} = P\{X \le x_0\} = F_X(x_0) \tag{3.4-3}$$

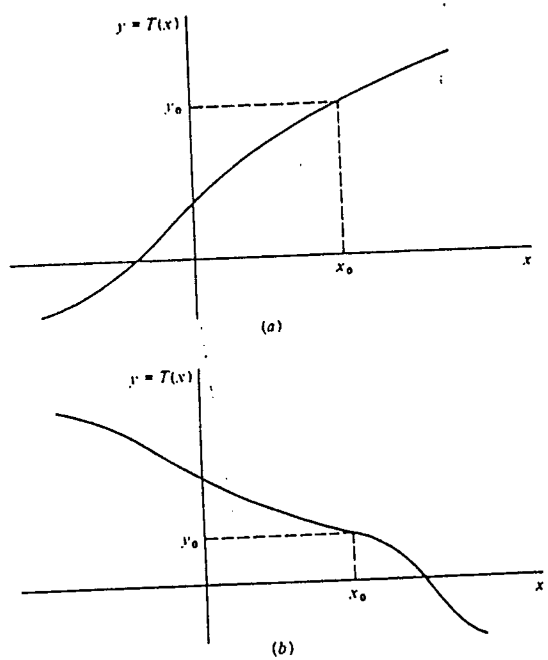


Figure 3.4-2 Monotonic transformations: (*a*) increasing, and (*b*) decreasing. [*Adapted from Peebles (1976) with permission of publishers Addison-Wesley, Advanced Book Program.*]

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

$$f_Y(y_0) = f_X[T^{-1}(y_0)]\,\frac{dT^{-1}(y_0)}{dy_0} \tag{3.4-5}$$

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

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| \tag{3.4-10}$$

† Leibniz's rule, after the great German mathematician Gottfried Wilhelm von Leibniz (1646–1716), states that, if $H(x, u)$ is continuous in x and u and

$$G(u) = \int_{\alpha(u)}^{\beta(u)} H(x, u)\,dx$$

then the derivative of the integral with respect to the parameter u is

$$\frac{dG(u)}{du} = H[\beta(u), u]\,\frac{d\beta(u)}{du} - H[\alpha(u), u]\,\frac{d\alpha(u)}{du} + \int_{\alpha(u)}^{\beta(u)} \frac{\partial H(x, u)}{\partial u}\,dx$$

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

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_X^2}}\, e^{-[(y-b)/a - a_X]^2/2\sigma_X^2}\left|\frac{1}{a}\right|$$

$$= \frac{1}{\sqrt{2\pi a^2\sigma_X^2}}\, e^{-[y-(aa_X+b)]^2/2a^2\sigma_X^2}$$

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 $\{Y \le y_0\}$. For the value of y_0 shown in the figure, the event $\{Y \le y_0\}$ corresponds to the event $\{X \le x_1$ and $x_2 \le X \le x_3\}$. Thus, the probability of the event $\{Y \le y_0\}$ now equals the probability

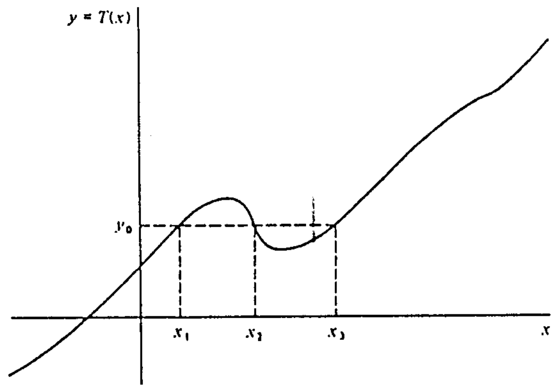


Figure 3.4-3 A nonmonotonic transformation. [*Adapted from Peebles (1976) with permission of publishers Addison-Wesley, Advanced Book Program.*]

of the event $\{x$ values yielding $Y \le y_0\}$, which we shall write as $\{x \mid Y \le y_0\}$. In other words

$$F_Y(y_0) = P\{Y \le y_0\} = P\{x \mid Y \le y_0\} = \int_{\{x \mid Y \le y_0\}} f_X(x)\, dx \tag{3.4-11}$$

Formally, one may differentiate to obtain the density function of Y:

$$f_Y(y_0) = \frac{d}{dy_0} \int_{\{x \mid Y \le y_0\}} f_X(x)\, dx \tag{3.4-12}$$

Although we shall not give a proof, the density function is also given by (Papoulis, 1965, p. 126)

$$f_Y(y) = \sum_n \frac{f_X(x_n)}{\left| \left. \frac{dT(x)}{dx} \right|_{x=x_n} \right|} \tag{3.4-13}$$

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†

$$y = T(x) \tag{3.4-14}$$

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 $\{Y \le y\}$ occurs when $\{-\sqrt{y/c} \le x \le \sqrt{y/c}\} = \{x \mid Y \le y\}$, so (3.4-12) becomes

$$f_Y(y) = \frac{d}{dy} \int_{-\sqrt{y/c}}^{\sqrt{y/c}} f_X(x)\, dx \qquad y \ge 0$$

Upon use of Leibniz's rule we obtain

$$f_Y(y) = f_X(\sqrt{y/c})\, \frac{d(\sqrt{y/c})}{dy} - f_X(-\sqrt{y/c})\, \frac{d(-\sqrt{y/c})}{dy}$$

$$= \frac{f_X(\sqrt{y/c}) + f_X(-\sqrt{y/c})}{2\sqrt{cy}} \qquad y \ge 0$$

† If $y = T(x)$ has no real roots for a given value of y, then $f_Y(y) = 0$.

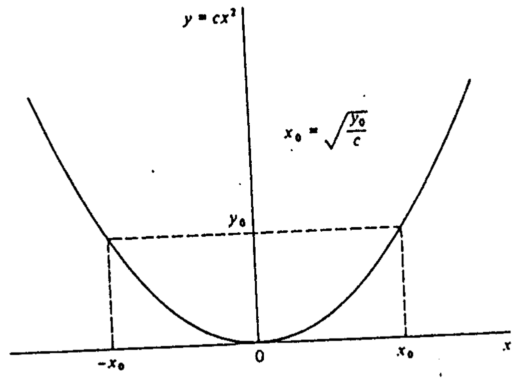


Figure 3.4-4 A square-law transformation. [*Adapted from Peebles (1976) with permission of publishers Addison-Wesley, Advanced Book Program.*]

In the latter case where we use (3.4-13), we have $X = \pm\sqrt{Y/c}$, $Y \ge 0$, so $x_1 = -\sqrt{y/c}$ and $x_2 = \sqrt{y/c}$. Furthermore, $dT(x)/dx = 2cx$ so

$$\left. \frac{dT(x)}{dx} \right|_{x=x_1} = 2cx_1 = -2c\sqrt{\frac{y}{c}} = -2\sqrt{cy}$$

$$\left. \frac{dT(x)}{dx} \right|_{x=x_2} = 2\sqrt{cy}$$

From (3.4-13) we again have

$$f_Y(y) = \frac{f_X(\sqrt{y/c}) + f_X(-\sqrt{y/c})}{2\sqrt{cy}} \qquad y \ge 0$$

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 $y_n = T(x_n)$.

PROBLEMS

3-1 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.

3-2 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.

3-3 If the probabilities in Problem 3-2 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.)

3-4 Give an example of a random variable where its mean value might not equal any of its possible values.

3-5 Find:

(*a*) the expected value, and

(*b*) the variance of the random variable with the triangular density of Figure 2.3-1*a* if $a = 1/\alpha$.

3-6 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

$$\sigma_X^2 = (b - a)^2/12$$

3-7 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 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 3-6.)

3-8 Find:

(*a*) the mean value, and

(*b*) the variance of the random variable X defined by Problems 2-6 and 2-14 of Chapter 2.

***3-9** For the *binomial density* of (2.5-1), show that

$$E[X] = \bar{X} = Np$$

and

$$\sigma_X^2 = Np(1 - p)$$

3-10 (*a*) Let resistance be a random variable in Problem 2-11 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.

3-11 (*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.)

3-12 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

$$\sigma_X^2 = b(4 - \pi)/4$$

3-13 What is the expected lifetime of the system defined in Problem 2-33 of Chapter 2?

3-14 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 $-\infty < m < \infty$.

3-15 Determine the mean value of the *Cauchy* random variable in Problem 2-34 of Chapter 2. What can you say about the variance of this random variable?

***3-16** 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.

3-17 (*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*).

3-18 Find expressions for all the moments about the origin and central moments for the uniform density of (2.5-7).

3-19 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

$$E[g(X)] = 1 - F_X(x_0)$$

3-20 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 $a = \bar{X}$.

3-21 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.

3-22 Prove that central moments μ_n are related to moments m_k about the origin by

$$\mu_n = \sum_{k=0}^{n} \binom{n}{k} (-\bar{X})^{n-k} m_k$$

3-23 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

$$m_n' = \sum_{k=0}^{n} \binom{n}{k} \alpha^{n-k} m_k$$

*3-24 Show that any characteristic function $\Phi_X(\omega)$ satisfies

$$|\Phi_X(\omega)| \leq \Phi_X(0) = 1$$

3-25 A random variable X is uniformly distributed on the interval $(-5, 15)$. Another random variable $Y = e^{-X/5}$ is formed. Find $E[Y]$.

3-26 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.

*3-27 For the system having a lifetime specified in Problem 2-33 of Chapter 2, determine the expected lifetime of the system given that the system has survived 20 weeks.

*3-28 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 $\Phi_X(\omega)$.

*3-29 Work Problem 3-28 using the moment generating function

$$M_X(v) = \exp(\sigma_X^2 v^2/2)$$

for the zero-mean gaussian random variable.

*3-30 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

$$\Phi_X(\omega) = \frac{1}{N+1} \frac{\sin[(N+1)\omega\Delta/2]}{\sin(\omega\Delta/2)} e^{jN\omega\Delta/2}$$

3-31 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.

3-32 Work Problem 3-31 if X is uniform on the interval $(-\pi, \pi)$.

3-33 A random variable X undergoes the transformation $Y = a/X$, where a is a real number. Find the density function of Y.

3-34 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.

3-35 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.

3-36 If the transformation of Problem 3-35 is applied to a Rayleigh random variable with $a \geq 0$, what is its effect?

*3-37 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 $Y = \cos(\Theta)$.

3-38 A random variable X can have values -4, -1, 2, 3, and 4, each with probability $1/5$. Find:

(a) the density function,
(b) the mean, and
(c) the variance of the random variable $Y = 3X^3$.

ADDITIONAL PROBLEMS

3-39 (*a*) Find the average amount the gambler in Problem 2-42 can expect to win. (*b*) What is his probability of winning on any given playing of the game?

3-40 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.

*3-41 For the animal described in Problem 2-58 find its expected lifetime given that it will not live beyond 20 weeks.

3-42 Find the expected value of the function $g(X) = X^3$ where X is a random variable defined by the density

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

3-43 Continue Problem 3-25 by finding all moments of Y. (*Hint:* Treat Y^n as a function of Y, not as a transformation.)

3-44 Reconsider the production line that manufactures bolts in Problem 2-12.

(*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?

3-45 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.

3-46 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?

3-47 Find the skew and coefficient of skewness for a Rayleigh random variable for which $a = 0$ in (2.5-11).

3-48 A random variable X has the density

$$f_X(x) = \begin{cases} (\tfrac{3}{32})(-x^2 + 8x - 12) & 2 \le x \le 6 \\ 0 & \text{elsewhere} \end{cases}$$

Find the following moments: (*a*) m_0, (*b*) m_1, (*c*) m_2, and (*d*) μ_2.

3-49 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.

3-50 For the density of Problem 3-49 find its arbitrary moment $\overline{X^n}$, $n = 0, 1, 2, \ldots$.

3-51 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 3-49 to find (*a*) the mean value, (*b*) the second moment, and (*c*) the variance of X.

*3-52 Show that the characteristic function of a random variable having the binomial density of (2.5-1) is

$$\Phi_X(\omega) = [1 - p + pe^{j\omega}]^N$$

*3-53 Show that the characteristic function of a Poisson random variable defined by (2.5-4) is

$$\Phi_X(\omega) = \exp[-b(1 - e^{j\omega})]$$

*3-54 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 $\sigma_X^2 = N/a^2$.

3-55 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.

*3-56 For any real random variable X with mean $\bar{X}$ and variance σ_X^2, *Chebychev's Inequality*§ is

$$P\{|X - \bar{X}| \ge \lambda\sigma_X\} \le 1/\lambda^2$$

where $\lambda > 0$ is a real constant. Prove the inequality. (*Hint:* Define a new random variable $Y = 0$ for $|X - \bar{X}| < \lambda\sigma_X$ and $Y = \lambda^2\sigma_X^2$ for $|X - \bar{X}| > \lambda\sigma_X$, observe that $Y \le (X - \bar{X})^2$ and find $E[Y]$.)

† After Ernst Hjalmar Waloddi Weibull (1887– ·), a Swedish applied physicist.
‡ A. K. Erlang (1878-1929) was a Danish engineer.
§ After the Russian mathematician Pafnuty Lvovich Chebychev (1821–1894).

3-57 A gaussian random variable, for which

$$f_X(x) = (2/\sqrt{\pi}) \exp(-4x^2)$$

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 3-49.)

3-58 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.

3-59 Work Problem 3-31 except assume a transformation $Y = T(X) = a \sin(X)$ with $a > 0$.

3-60 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 2-35. This transformation allows log-normal random numbers to be generated from gaussian random numbers by a digital computer.

3-61 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.

CHAPTER
FOUR
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 2-29). 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 has been called the *range sample space* (Davenport, 1970) or the *two-dimensional product space*. 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 correspondences 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.

† 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 $\{Y \leq y\}$.

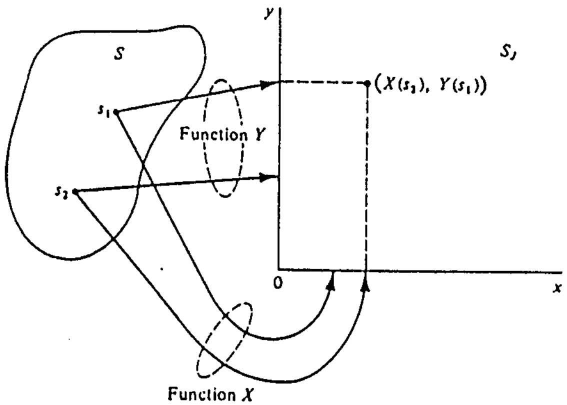


Figure 4.1-1 Mapping from the sample space S to the joint sample space S_J (xy plane).

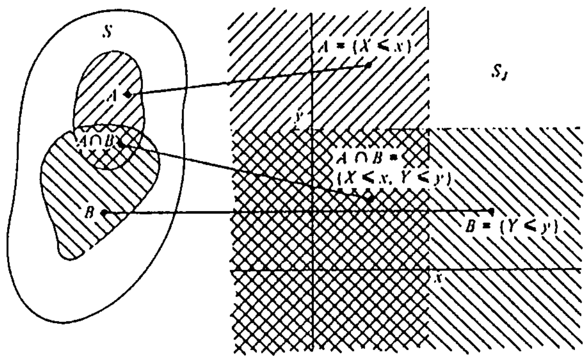


Figure 4.1-2 Comparisons of events in S with those in S_J.

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 assumed 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 do 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*.

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 4.2-1*a* 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,

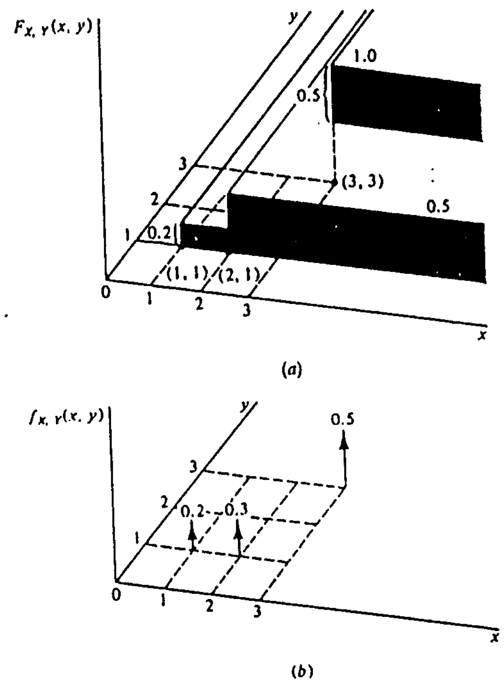


Figure 4.2-1 A joint distribution function (*a*), and its corresponding joint density function (*b*), that apply to Examples 4.2-1 and 4.2-2.

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) $F_{X,Y}(-\infty, -\infty) = 0 \qquad F_{X,Y}(-\infty, y) = 0 \qquad F_{X,Y}(x, -\infty) = 0$ (4.2-6*a*)

(2) $F_{X,Y}(\infty, \infty) = 1$ (4.2-6*b*)

(3) $0 \le F_{X,Y}(x, y) \le 1$ (4.2-6*c*)

(4) $F_{X,Y}(x, y)$ is a nondecreasing function of both x and y (4.2-6*d*)

(5) $F_{X,Y}(x_2, y_2) + F_{X,Y}(x_1, y_1) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1)$
$= P\{x_1 < X \le x_2, y_1 < Y \le y_2\} \ge 0$ (4.2-6*e*)

(6) $F_{X,Y}(x, \infty) = F_X(x) \qquad F_{X,Y}(\infty, y) = F_Y(y)$ (4.2-6*f*)

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.

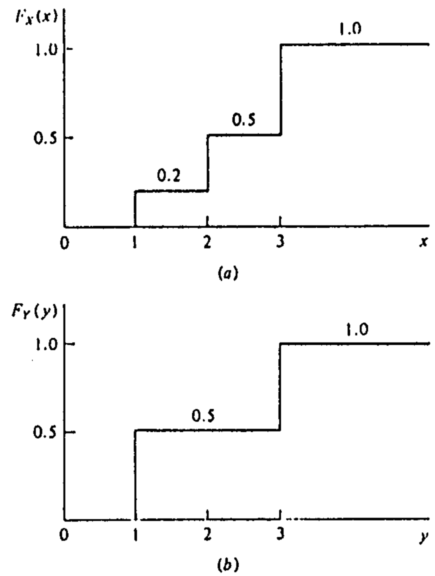


Figure 4.2-2 Marginal distributions applicable to Figure 4.2-1 and Example 4.2-2: (a) $F_X(x)$ and (b) $F_Y(y)$.

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 $1, 2, 3, \ldots, N-1$.

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 4.2-1*b*.

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:

$$f_{X_1, X_2, \ldots, X_N}(x_1, x_2, \ldots, x_N) = \frac{\partial^N F_{X_1, X_2, \ldots, X_N}(x_1, x_2, \ldots, x_N)}{\partial x_1\, \partial x_2 \cdots \partial x_N} \tag{4.3-3}$$

By direct integration this result is equivalent to

$$F_{X_1, X_2, \ldots, X_N}(x_1, x_2, \ldots, x_N) = \int_{-\infty}^{x_N} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_{X_1, X_2, \ldots, X_N}(\xi_1, \xi_2, \ldots, \xi_N)\, d\xi_1\, d\xi_2 \cdots d\xi_N \tag{4.3-4}$$

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) $f_{X,Y}(x, y) \geq 0$ (4.3-5*a*)

(2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy = 1$ (4.3-5*b*)

(3) $F_{X,Y}(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(\xi_1, \xi_2)\, d\xi_1\, d\xi_2$ (4.3-5*c*)

(4) $F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(\xi_1, \xi_2)\, d\xi_2\, d\xi_1$ (4.3-5*d*)

$F_Y(y) = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f_{X,Y}(\xi_1, \xi_2)\, d\xi_1\, d\xi_2$ (4.3-5*e*)

(5) $P\{x_1 < X \leq x_2, y_1 < Y \leq y_2\} = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{X,Y}(x, y)\, dx\, dy$ (4.3-5*f*)

(6) $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$ (4.3-5*g*)

$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$ (4.3-5*h*)

Properties 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).

The first five of these 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 (4.3-5*d*) and (4.3-5*e*) 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 4.3-1 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 (4.3-5*g*) 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 (4.3-5*h*):

$$f_Y(y) = \int_0^{\infty} u(y)xe^{-x(y+1)}\, dx = \frac{u(y)}{(y+1)^2}$$

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 $X_1, X_2, \ldots, X_k$:

$$f_{X_1, X_2, \ldots, X_k}(x_1, x_2, \ldots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, X_2, \ldots, X_N}(x_1, x_2, \ldots, x_N)\, dx_{k+1}\, dx_{k+2} \cdots dx_N \quad (4.3\text{-}8)$$

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

$$F_X(x \mid B) = P\{X \le x \mid B\} = \frac{P\{X \le x \cap B\}}{P(B)} \quad (4.4\text{-}1)$$

for any event B with nonzero probability. The corresponding conditional density function was defined through the derivative

$$f_X(x \mid B) = \frac{dF_X(x \mid B)}{dx} \quad (4.4\text{-}2)$$

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

$$F_X(x \mid y - \Delta y < Y \le y + \Delta y) = \frac{\int_{y-\Delta y}^{y+\Delta y} \int_{-\infty}^{x} f_{X,Y}(\xi_1, \xi_2)\, d\xi_1\, d\xi_2}{\int_{y-\Delta y}^{y+\Delta y} f_Y(\xi)\, d\xi} \quad (4.4\text{-}4)$$

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,

$$f_Y(y) = \sum_{j=1}^{M} P(y_j)\, \delta(y - y_j) \quad (4.4\text{-}5)$$

$$f_{X,Y}(x, y) = \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i, y_j)\, \delta(x - x_i)\, \delta(y - y_j) \quad (4.4\text{-}6)$$

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

$$F_X(x \mid Y = y_k) = \sum_{i=1}^{N} \frac{P(x_i, y_k)}{P(y_k)} u(x - x_i) \quad (4.4\text{-}7)$$

After differentiation we have

$$f_X(x \mid Y = y_k) = \sum_{i=1}^{N} \frac{P(x_i, y_k)}{P(y_k)} \delta(x - x_i) \quad (4.4\text{-}8)$$

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 $P(x_1, y_1) = 2/15$, $P(x_2, y_1) = 3/15$, etc. Since $P(y_3) = (4/15) + (5/15) = 9/15$, use of (4.4-8) will give $f_X(x \mid Y = y_3)$ 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 $f_X(x \mid Y = y)$ may exist. If Δy is very small, (4.4-4) can be written as

$$F_X(x \mid y - \Delta y < Y \le y + \Delta y) = \frac{\int_{-\infty}^{x} f_{X,Y}(\xi_1, y)\, d\xi_1\, 2\Delta y}{f_Y(y) 2\Delta y} \quad (4.4\text{-}9)$$

and, in the limit as $\Delta y \to 0$

$$F_X(x \mid Y = y) = \frac{\int_{-\infty}^{x} f_{X,Y}(\xi, y)\, d\xi}{f_Y(y)} \quad (4.4\text{-}10)$$

for every y such that $f_Y(y) \ne 0$. After differentiation of both sides of (4.4-10) with respect to x:

$$f_X(x \mid Y = y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \quad (4.4\text{-}11)$$

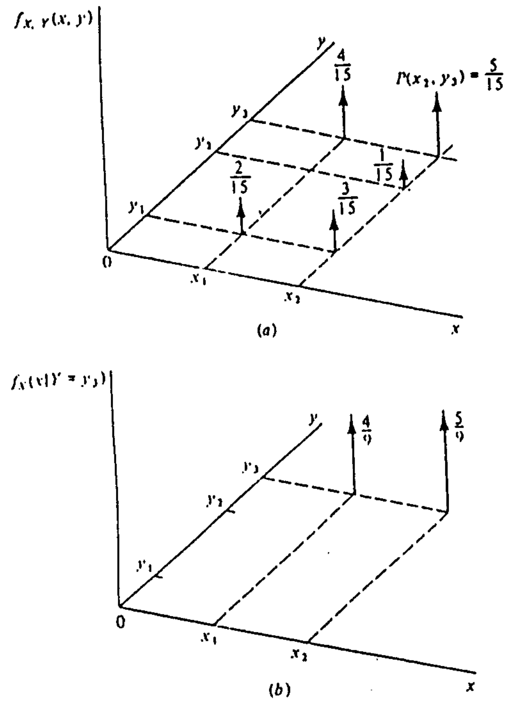


Figure 4.4-1 A joint density function (a) and a conditional density function (b) applicable to Example 4.4-1.

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 4.3-1. 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

$$B = \{y_a < Y \le y_b\} \tag{4.4-14}$$

where y_a and y_b are real numbers and we assume $P(B) = P\{y_a < Y \le y_b\} \ne 0$. With this definition it is readily shown that (4.4-1) and (4.4-2) become

$$F_X(x\,|\,y_a < Y \le y_b) = \frac{F_{X,Y}(x, y_b) - F_{X,Y}(x, y_a)}{F_Y(y_b) - F_Y(y_a)}$$
$$= \frac{\int_{y_a}^{y_b}\int_{-\infty}^{x} f_{X,Y}(\xi, y)\,d\xi\,dy}{\int_{y_a}^{y_b}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx\,dy} \tag{4.4-15}$$

and

$$f_X(x\,|\,y_a < Y \le y_b) = \frac{\int_{y_a}^{y_b} f_{X,Y}(x, y)\,dy}{\int_{y_a}^{y_b}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx\,dy} \tag{4.4-16}$$

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 $y_a < y \le y_b$ 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 $f_X(x\,|\,Y \le y)$ for the joint density function of Example 4.3-1. Since we have here defined $B = \{Y \le y\}$, 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:

$$\int_{-\infty}^{y} f_Y(\xi)\,d\xi = \int_{-\infty}^{y} \frac{u(\xi)\,d\xi}{(\xi+1)^2} = \int_{0}^{y} \frac{d\xi}{(\xi+1)^2} = \frac{y}{y+1} \qquad y > 0$$

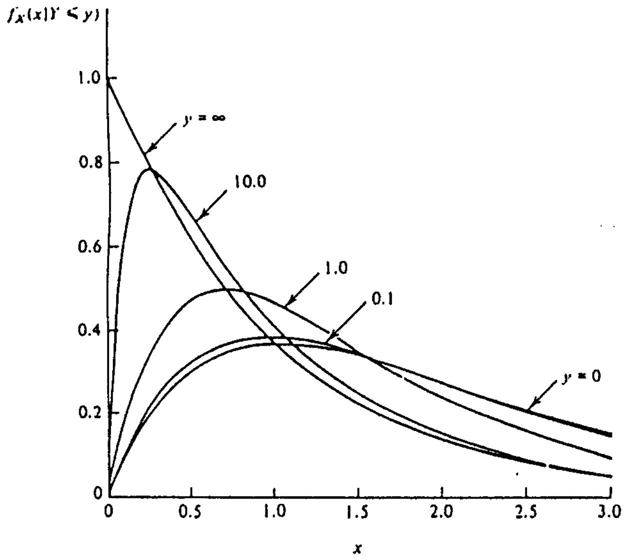


Figure 4.4-2 Conditional probability density functions applicable to Example 4.4-3.

and zero for $y < 0$, after using an integral from Appendix C. The numerator of (4.4-16) becomes

$$\int_{-\infty}^{y} f_{X,Y}(x, \xi)\, d\xi = \int_0^y u(x)xe^{-x(\xi+1)}\, d\xi$$

$$= u(x)xe^{-x}\int_0^y e^{-x\xi}\, d\xi$$

$$= u(x)e^{-x}(1 - e^{-xy}) \qquad y > 0$$

and zero for $y < 0$, after using another integral from Appendix C. Thus

$$f_X(x|Y \le y) = u(x)u(y)\left(\frac{y+1}{y}\right)e^{-x}(1 - e^{-xy})$$

This function is plotted in Figure 4.4-2 for several values of y.

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 $A = \{X \le x\}$ and $B = \{Y \le y\}$ for two real numbers x and y. Thus, X and Y are said to be *statistically independent random variables* if (and only if)

$$P\{X \le x, Y \le y\} = P\{X \le x\}P\{Y \le y\} \tag{4.5-2}$$

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 $B = \{Y \le y\}$:

$$F_X(x|Y \le y) = \frac{P\{X \le x, Y \le y\}}{P\{Y \le y\}} = \frac{F_{X,Y}(x, y)}{F_Y(y)} \tag{4.5-5}$$

By substituting (4.5-3) into (4.5-5), we have

$$F_X(x|Y \le y) = F_X(x) \tag{4.5-6}$$

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

$$F_Y(y|X \le x) = F_Y(y) \tag{4.5-7}$$

Conditional density function forms, for independent X and Y, are found by differentiation of (4.5-6) and (4.5-7):

$$f_X(x|Y \le y) = f_X(x) \tag{4.5-8}$$

$$f_Y(y|X \le x) = f_Y(y) \tag{4.5-9}$$

Example 4.5-1 For the densities of Example 4.3-1:

$$f_{X,Y}(x, y) = u(x)u(y)xe^{-x(y+1)}$$

$$f_X(x)f_Y(y) = u(x)u(y)\frac{e^{-x}}{(y+1)^2} \ne f_{X,Y}(x, y)$$

Therefore the random variables X and Y are not 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 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):

$$F_W(w) = \int_{-\infty}^{\infty} f_Y(y) \int_{x=-\infty}^{w-y} f_X(x)\,dx\,dy \tag{4.6-4}$$

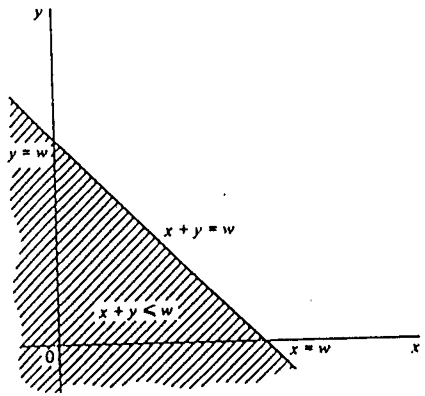


Figure 4.6-1 Region in xy plane where $x + y \leq w$.

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

$$\begin{aligned}
f_W(w) &= \int_{-\infty}^{\infty} \frac{1}{ab}[u(y) - u(y - b)][u(w - y) - u(w - y - a)]\,dy \\
&= \frac{1}{ab}\int_0^{\infty} [1 - u(y - b)][u(w - y) - u(w - y - a)]\,dy \\
&= \frac{1}{ab}\left[\int_0^{\infty} u(w - y)\,dy - \int_0^{\infty} u(w - y - a)\,dy \right.\\
&\qquad \left. - \int_0^{\infty} u(y - b)u(w - y)\,dy + \int_0^{\infty} u(y - b)u(w - y - a)\,dy\right]
\end{aligned}$$

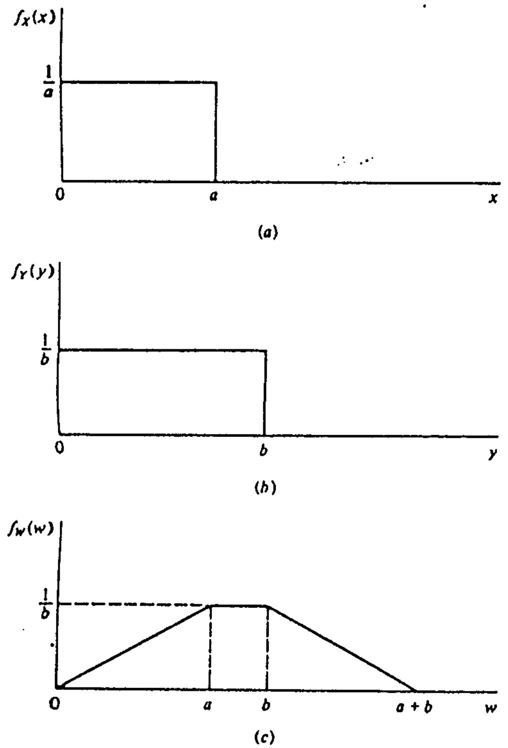


Figure 4.6-2 Two density functions (*a*) and (*b*) and their convolution (*c*).

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*).

*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)

$$\sigma^2_{X_i} > B_1 > 0 \qquad i = 1, 2, \ldots, N \tag{4.7-1a}$$

$$E[|X_i - \bar{X}_i|^3] < B_2 \qquad i = 1, 2, \ldots, N \tag{4.7-1b}$$

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

† The asterisk denotes convolution.

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 (4.7-1*a*). 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

$$W_N = (Y_N - \bar{Y}_N)/\sigma_{Y_N} = \sum_{i=1}^{N}(X_i - \bar{X}_i) \Bigg/ \left[\sum_{i=1}^{N}\sigma_{X_i}^2\right]^{1/2}$$

$$= \frac{1}{\sqrt{N}\,\sigma_X}\sum_{i=1}^{N}(X_i - \bar{X}) \tag{4.7-2}$$

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

$$\Phi_{W_N}(\omega) = \exp(-\omega^2/2) \tag{4.7-5}$$

† These are called *lattice-type* discrete random variables (Papoulis, 1965).

from Problem 3-28. 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

$$\Phi_{W_N}(\omega) = E[e^{j\omega W_N}] = E\left[\exp\left\{\frac{j\omega}{\sqrt{N}\,\sigma_X}\sum_{i=1}^{N}(X_i - \bar{X})\right\}\right]$$

$$= \left\langle E\left\{\exp\left[\frac{j\omega}{\sqrt{N}\,\sigma_X}(X_i - \bar{X})\right]\right\}\right\rangle^N \tag{4.7-6}$$

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:

$$E\left\{\exp\left[\frac{j\omega}{\sqrt{N}\,\sigma_X}(X_i - \bar{X})\right]\right\}$$

$$= E\left\{1 + \left(\frac{j\omega}{\sqrt{N}\,\sigma_X}\right)(X_i - \bar{X}) + \left(\frac{j\omega}{\sqrt{N}\,\sigma_X}\right)^2\frac{(X_i - \bar{X})^2}{2} + \frac{R_N}{N}\right\}$$

$$= 1 - (\omega^2/2N) + E[R_N]/N \tag{4.7-7}$$

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 $b = a$):

$$f_W(w) = \frac{1}{a}\,\text{tri}\left(\frac{w}{a}\right)$$

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 4-63).

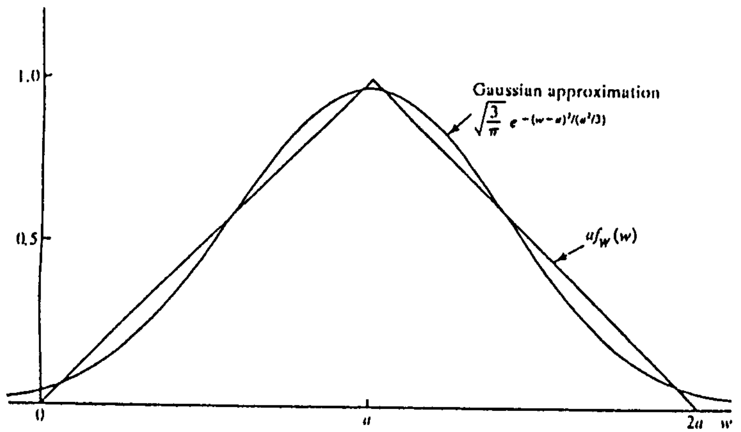


Figure 4.7-1 The triangular density function of Example 4.7-1 and its gaussian approximation.

PROBLEMS

4-1 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 $A = \{X \le x\}$ and $B = \{y_1 < Y \le y_2\}$. Make a sketch of the two sample spaces showing areas corresponding to both events and the event $A \cap B = \{X \le x, y_1 < Y \le y_2\}$.

4-2 Work Problem 4-1 for the two events $A = \{x_1 < X \le x_2\}$ and $B = \{y_1 < Y \le y_2\}$.

4-3 Work Problem 4-1 for the two events $A = \{x_1 < X \le x_2 \text{ or } x_3 < X \le x_4\}$ and $B = \{y_1 < Y \le y_2\}$.

4-4 Three events A, B, and C satisfy $C \subset B \subset A$ and are defined by $A = \{X \le x_a, \ Y \le y_a\}$, $B = \{X \le x_b, \ Y \le y_b\}$, and $C = \{X \le x_c, \ Y \le y_c\}$ 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 $A \cap B \cap C$?

4-5 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 $\{X \le 3\}$.

4-6 Write a mathematical equation for $F_{X,Y}(x, y)$ of Problem 4-5.

4-7 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 $F_{X,Y}(x, y)$.

4-8 By use of the joint distribution function in Problem 4-7, 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*) $P\{-1.5 < X \le 2, \ 1 < Y \le 3\}$.

4-9 Find and sketch the marginal distribution functions for the joint distribution function of Problem 4-5.

4-10 Find and sketch the marginal distribution functions for the joint distribution function of Problem 4-7.

4-11 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.

4-12 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 $F_Y(y)$.

4-13 Show that the function

$$G_{X,Y}(x, y) = \begin{cases} 0 & x < y \\ 1 & x \ge y \end{cases}$$

cannot be a valid joint distribution function. [*Hint:* Use (4.2-6e).]

4-14 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.

4-15 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 $F_{X,Y}(x, y)$.

4-16 If $a < b$ in Problem 4-15, find:

(*a*) $P\{X + Y \le 3a/4\}$ (*b*) $P\{Y \le 2bX/a\}$.

4-17 Find the joint distribution function applicable to Example 4.3-1.

4-18 Sketch the joint density function $f_{X,Y}(x, y)$ applicable to Problem 4-5. Write an equation for $f_{X,Y}(x, y)$.

4-19 Determine the joint density and both marginal density functions for Problem 4-7.

4-20 Find and sketch the joint density function for the distribution function in Problem 4-12.

4-21 (*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.

4-22 (*a*) By use of the joint density function of Problem 4-21, find the marginal density functions.
(*b*) What is $P\{0.5a < X \le 0.75a\}$ in terms of a and b?

4-23 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*) $f_{X,Y}(x, y) = \begin{cases} b(x^2 + 4y^2) & 0 \le |x| < 1 \quad \text{and} \quad 0 \le y < 2 \\ 0 & \text{elsewhere} \end{cases}$

***4-24** 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.)

***4-25** 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.)

4-26 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 $f_Y(y)$.

4-27 Find the conditional density functions $f_X(x \mid y_1)$, $f_X(x \mid y_2)$, $f_Y(y \mid x_1)$, and $f_Y(y \mid x_2)$ for the joint density defined in Example 4.4-1.

4-28 Find the conditional density function $f_X(x \mid y)$ applicable to Example 4.4-2.

4-29 By using the results of Example 4.4-2, calculate the probability of the event $\{Y \le 2 \mid X = 1\}$.

4-30 Random variables X and Y are *jointly 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?

4-31 By use of the joint density of Problem 4-30, show that

$$f_X(x \mid Y = y) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{(x-\rho y)^2}{2(1-\rho^2)}\right]$$

4-32 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 $f_X(x \mid Y = y)$ and $f_Y(y \mid X = x)$.

(*b*) Are the random variables X and Y statistically independent?

4-33 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

$$P\{Y \le X\} = 1 - \int_{-\infty}^{\infty} F_X(y) f_Y(y)\, dy$$

4-34 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?

4-35 Show, by use of (4.4-13), that the area under $f_Y(y \mid x)$ is unity.

***4-36** 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 $f_R(r \mid \Theta = \pi)$.

(*c*) Find $f_R(r \mid \Theta \le \pi)$ and compare to the result found in part (*b*), and explain the comparison.

4-37 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.

4-38 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.

4-39 Find and sketch the density function of $W = X + Y$, where the random variable X is that of Problem 4-37 with $a = 5$ and Y is that of Problem 4-38. Assume X and Y are independent.

4-40 Find the density function of $W = X + Y$, where the random variable X is that of Problem 4-38 and Y is that of Problem 4-37. Assume X and Y are independent. Sketch the density function for $b = 1$ and $b = 4$.

***4-41** 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.

ADDITIONAL PROBLEMS

4-42 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.

4-43 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 $P\{-1 < X \le 4,\ -3 < Y \le 3\}$.

4-44 Random variables X and Y have the joint distribution

$$
F_{X,Y}(x, y) = \begin{cases} \dfrac{5}{4}\left(\dfrac{x + e^{-(x+1)y^2}}{x+1} - e^{-y^2}\right)u(y) & 0 \le x < 4 \\ 0 & x < 0 \text{ or } y < 0 \\ 1 + \dfrac{1}{4}e^{-5y^2} - \dfrac{5}{4}e^{-y^2} & 4 \le x \text{ and any } y \ge 0 \end{cases}
$$

Find: (*a*) The marginal distribution functions of X and Y, and (*b*) the probability $P\{3 < X \le 5,\ 1 < Y \le 2\}$.

4-45 Find the joint distribution function of the random variables having the joint density of Problem 4-48.

4-46 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.

4-47 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.

4-48 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.

4-49 Find the marginal densities of X and Y using the joint density

$$f_{X,Y}(x, y) = 2u(x)u(y)\exp\left[-\left(4y + \frac{x}{2}\right)\right]$$

4-50 Random variables X and Y have the joint density of Problem 4-49. Find the probability that the values of Y are not greater than twice the values of X for $x \le 3$.

4-51 Find the conditional densities $f_X(x \mid Y = y)$ and $f_Y(y \mid X = x)$ applicable to the joint density of Problem 4-47.

4-52 For the joint density of Problem 4-48 determine the conditional densities $f_X(x \mid Y = y)$ and $f_Y(y \mid X = x)$.

4-53 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 result for $T_0 = 1$ h.

(*b*) Repeat part (*a*) given that departure time is at 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.)

***4-54** Start with the expressions

$$F_Y(y \mid B) = P\{Y \le y \mid B\} = \frac{P\{Y \le y \cap B\}}{P(B)}$$

$$f_Y(y \mid B) = \frac{dF_Y(y \mid B)}{dy}$$

which are analogous to (4.4-1) and (4.4-2), and derive $F_Y(y \mid x_a < X \le x_b)$ and $f_Y(y \mid x_a < X \le x_b)$ which are analogous to (4.4-15) and (4.4-16).

***4-55** 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

$$
F_{X,Y}(x, y \mid y_a < Y \le y_b) = \begin{cases} 0 & y \le y_a \\ \dfrac{F_{X,Y}(x, y) - F_{X,Y}(x, y_a)}{F_Y(y_b) - F_Y(y_a)} & y_a < y \le y_b \\ \dfrac{F_{X,Y}(x, y_b) - F_{X,Y}(x, y_a)}{F_Y(y_b) - F_Y(y_a)} & y_b < y \end{cases}
$$

and

$$
f_{X,Y}(x, y \mid y_a < Y \le y_b) = \begin{cases} 0 & y \le y_a \quad \text{and} \quad y > y_b \\ \dfrac{f_{X,Y}(x, y)}{F_Y(y_b) - F_Y(y_a)} & y_a < y \le y_b \end{cases}
$$

4-56 Determine if random variables X and Y of Problem 4-53 are statistically independent.

4-57 Determine if X and Y of Problem 4-49 are statistically independent.

4-58 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) $f_{X_1}(x_1 | x_2, x_3, x_4)$.

4-59 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.

4-60 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) $\{Y \leq 8 - (2|X|/3)\}$, and (b) $\{Y \leq 8 + (2|X|/3)\}$.
Compare the two results.

4-61 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 $W = X + Y$.

***4-62** 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.

***4-63** 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?

***4-64** Work Problem 4-63 except assume

$$f_X(x) = (1/2) \cos(x) \operatorname{rect}(x/\pi).$$

CHAPTER
FIVE
OPERATIONS ON MULTIPLE RANDOM VARIABLES

5.0 INTRODUCTION

After establishing some of the basic theory of several random variables in the previous chapter, it is appropriate to now extend the operations described in Chapter 3 to include multiple random variables. This chapter is dedicated to these extensions. Mainly, the concept of expectation is enlarged to include two or more random variables. Other operations involving moments, characteristic functions, and transformations are all special applications of expectation.

5.1 EXPECTED VALUE OF A FUNCTION OF RANDOM VARIABLES

When more than a single random variable is involved, expectation must be taken with respect to all the variables involved. For example, if $g(X, Y)$ is some function of two random variables X and Y the expected value of $g(\cdot,\cdot)$ is given by

$$\bar{g} = E[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y)\, dx\, dy \tag{5.1-1}$$

This expression is the two-variable extension of (3.1-6).

For N random variables $X_1, X_2, \ldots, X_N$ and some function of these variables, denoted $g(X_1, \ldots, X_N)$, the expected value of the function becomes

$$\bar{g} = E[g(X_1, \ldots, X_N)]$$
$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_N) f_{X_1, \ldots, X_N}(x_1, \ldots, x_N)\, dx_1 \cdots dx_N \tag{5.1-2}$$

Thus, expectation in general involves an N-fold integration when N random variables are involved.

We illustrate the application of (5.1-2) with an example that will develop an important point.

Example 5.1-1 We shall find the mean (expected) value of a sum of N weighted random variables. If we let

$$g(X_1, \ldots, X_N) = \sum_{i=1}^{N} \alpha_i X_i$$

where the "weights" are the constants α_i, the mean value of the weighted sum becomes

$$E[g(X_1, \ldots, X_N)] = E\left[\sum_{i=1}^{N} \alpha_i X_i\right]$$
$$= \sum_{i=1}^{N} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \alpha_i x_i f_{X_1, \ldots, X_N}(x_1, \ldots, x_N)\, dx_1 \cdots dx_N$$

from (5.1-2). After using (4.3-8), the terms in the sum all reduce to the form

$$\int_{-\infty}^{\infty} \alpha_i x_i f_{X_i}(x_i)\, dx_i = E[\alpha_i X_i] = \alpha_i E[X_i]$$

so

$$E\left[\sum_{i=1}^{N} \alpha_i X_i\right] = \sum_{i=1}^{N} \alpha_i E[X_i]$$

which says that *the mean value of a weighted sum of random variables equals the weighted sum of mean values.*

The above extensions (5.1-1) and (5.1-2) of expectation do not invalidate any of our single random variable results. For example, let

$$g(X_1, \ldots, X_N) = g(X_1) \tag{5.1-3}$$

and substitute into (5.1-2). After integrating with respect to all random variables except X_1, (5.1-2) becomes

$$\bar{g} = E[g(X_1)] = \int_{-\infty}^{\infty} g(x_1) f_{X_1}(x_1)\, dx_1 \tag{5.1-4}$$

which is the same as previously given in (3.1-6) for one random variable. Some reflection on the reader's part will verify that (5.1-4) also validates such earlier topics as moments, central moments, characteristic function, etc., for a single random variable.

Joint Moments About the Origin

One important application of (5.1-1) is in defining *joint moments* about the origin. They are denoted by m_{nk} and are defined by

$$m_{nk} = E[X^n Y^k] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^n y^k f_{X,Y}(x, y)\, dx\, dy \tag{5.1-5}$$

for the case of two random variables X and Y. Clearly $m_{n0} = E[X^n]$ are the moments m_n of X, while $m_{0k} = E[Y^k]$ are the moments of Y. The sum $n + k$ is called the *order* of the moments. Thus m_{02}, m_{20}, and m_{11} are all second-order moments of X and Y. The first-order moments $m_{01} = E[Y] = \bar{Y}$ and $m_{10} = E[X] = \bar{X}$ are the expected values of Y and X, respectively, and are the coordinates of the "center of gravity" of the function $f_{X,Y}(x, y)$.

The second-order moment $m_{11} = E[XY]$ is called the *correlation* of X and Y. It is so important to later work that we give it the symbol R_{XY}. Hence,

$$R_{XY} = m_{11} = E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y)\, dx\, dy \tag{5.1-6}$$

If correlation can be written in the form

$$R_{XY} = E[X]E[Y] \tag{5.1-7}$$

then X and Y are said to be *uncorrelated.* Statistical independence of X and Y is sufficient to guarantee they are uncorrelated, as is readily proven by (5.1-6) using (4.5-4). The converse of this statement, that is, that X and Y are independent if X and Y are uncorrelated, is *not* necessarily true in general.†

If

$$R_{XY} = 0 \tag{5.1-8}$$

for two random variables X and Y, they are called *orthogonal.*

A simple example is next developed that illustrates the important new topic of correlation.

† Uncorrelated *gaussian* random variables are, however, known to also be independent (see Section 5.3).

Example 5.1-2 Let X be a random variable that has a mean value $\bar{X} = E[X] = 3$ and variance $\sigma_X^2 = 2$. From (3.2-6) we easily determine the second moment of X about the origin: $E[X^2] = m_{20} = \sigma_X^2 + \bar{X}^2 = 11$.

Next, let another random variable Y be defined by

$$Y = -6X + 22$$

The mean value of Y is $\bar{Y} = E[Y] = E[-6X + 22] = -6\bar{X} + 22 = 4$. The correlation of X and Y is found from (5.1-6)

$$R_{XY} = m_{11} = E[XY] = E[-6X^2 + 22X] = -6E[X^2] + 22\bar{X}$$
$$= -6(11) + 22(3) = 0$$

Since $R_{XY} = 0$, X and Y are orthogonal from (5.1-8). On the other hand, $R_{XY} \neq E[X]E[Y] = 12$, so X and Y are *not* uncorrelated [see (5.1-7)].

We note that two random variables can be orthogonal even though correlated when one, Y, is related to the other, X, by the linear function $Y = aX + b$. It can be shown that X and Y are always correlated if $|a| \neq 0$, regardless of the value of b (see Problem 5-9). They are uncorrelated if $a = 0$, but this is not a case of much practical interest. Orthogonality can likewise be shown to occur when a and b are related by $b = -aE[X^2]/E[X]$ whenever $E[X] \neq 0$. If $E[X] = 0$, X and Y cannot be orthogonal for any value of a except $a = 0$, a noninteresting problem. The reader may wish to verify these statements as an exercise.

For N random variables $X_1, X_2, \ldots, X_N$, the $(n_1 + n_2 + \cdots + n_N)$-order moments are defined by

$$m_{n_1 n_2 \cdots n_N} = E[X_1^{n_1} X_2^{n_2} \cdots X_N^{n_N}]$$
$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{n_1} \cdots x_N^{n_N} f_{X_1,\ldots,X_N}(x_1, \ldots, x_N)\, dx_1 \cdots dx_N \quad (5.1\text{-}9)$$

where $n_1, n_2, \ldots, n_N$ are all integers $= 0, 1, 2, \ldots$.

Joint Central Moments

Another important application of (5.1-1) is in defining *joint central moments*. For two random variables X and Y, these moments, denoted by μ_{nk}, are given by

$$\mu_{nk} = E[(X - \bar{X})^n (Y - \bar{Y})^k]$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{X})^n (y - \bar{Y})^k f_{X,Y}(x, y)\, dx\, dy \quad (5.1\text{-}10)$$

The second-order central moments

$$\mu_{20} = E[(X - \bar{X})^2] = \sigma_X^2 \quad (5.1\text{-}11)$$
$$\mu_{02} = E[(Y - \bar{Y})^2] = \sigma_Y^2 \quad (5.1\text{-}12)$$

are just the variances of X and Y.

The second-order joint moment μ_{11} is very important. It is called the *covariance* of X and Y and is given the symbol C_{XY}. Hence

$$C_{XY} = \mu_{11} = E[(X - \bar{X})(Y - \bar{Y})]$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{X})(y - \bar{Y}) f_{X,Y}(x, y)\, dx\, dy \quad (5.1\text{-}13)$$

By direct expansion of the product $(x - \bar{X})(y - \bar{Y})$, this integral reduces to the form

$$C_{XY} = R_{XY} - \bar{X}\bar{Y} = R_{XY} - E[X]E[Y] \quad (5.1\text{-}14)$$

when (5.1-6) is used. If X and Y are either independent or uncorrelated, then (5.1-7) applies and (5.1-14) shows their covariance is zero:

$$C_{XY} = 0 \qquad X \text{ and } Y \text{ independent or uncorrelated} \quad (5.1\text{-}15)$$

If X and Y are orthogonal random variables, then

$$C_{XY} = -E[X]E[Y] \qquad X \text{ and } Y \text{ orthogonal} \quad (5.1\text{-}16)$$

from use of (5.1-8) with (5.1-14). Clearly, $C_{XY} = 0$ if either X or Y also has zero mean value.

The normalized second-order moment

$$\rho = \mu_{11}/\sqrt{\mu_{20}\mu_{02}} = C_{XY}/\sigma_X \sigma_Y \quad (5.1\text{-}17a)$$

given by

$$\rho = E\left[\frac{(X - \bar{X})}{\sigma_X} \frac{(Y - \bar{Y})}{\sigma_Y}\right] \quad (5.1\text{-}17b)$$

is known as the *correlation coefficient* of X and Y. It can be shown (see Problem 5-10) that

$$-1 \leq \rho \leq 1 \quad (5.1\text{-}18)$$

For N random variables $X_1, X_2, \ldots, X_N$ the $(n_1 + n_2 + \cdots + n_N)$-order joint central moment is defined by

$$\mu_{n_1 n_2 \cdots n_N} = E[(X_1 - \bar{X}_1)^{n_1} (X_2 - \bar{X}_2)^{n_2} \cdots (X_N - \bar{X}_N)^{n_N}]$$
$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_1 - \bar{X}_1)^{n_1} \cdots$$
$$(x_N - \bar{X}_N)^{n_N} f_{X_1,\ldots,X_N}(x_1, \ldots, x_N)\, dx_1 \cdots dx_N \quad (5.1\text{-}19)$$

An example is next developed that involves the use of covariances.

Example 5.1-3 Again let X be a weighted sum of N random variables X_i; that is, let

$$X = \sum_{i=1}^{N} \alpha_i X_i$$

where the α_i are real weighting constants. The variance of X will be found. From Example 5.1-1,

$$E[X] = \sum_{i=1}^{N} \alpha_i E[X_i] = \sum_{i=1}^{N} \alpha_i \bar{X}_i = \bar{X}$$

so we have

$$X - \bar{X} = \sum_{i=1}^{N} \alpha_i(X_i - \bar{X}_i)$$

and

$$\sigma_X^2 = E[(X - \bar{X})^2] = E\left[\sum_{i=1}^{N} \alpha_i(X_i - \bar{X}_i)\sum_{j=1}^{N} \alpha_j(X_j - \bar{X}_j)\right]$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j E[(X_i - \bar{X}_i)(X_j - \bar{X}_j)] = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j C_{X_i X_j}$$

Thus, the variance of a weighted sum of N random variables X_i (weights α_i) equals the weighted sum of all their covariances $C_{X_i X_j}$ (weights $\alpha_i \alpha_j$). For the special case of uncorrelated random variables, where

$$C_{X_i X_j} = \begin{cases} 0 & i \neq j \\ \sigma_{X_i}^2 & i = j \end{cases}$$

is true, we get

$$\sigma_X^2 = \sum_{i=1}^{N} \alpha_i^2 \sigma_{X_i}^2$$

In words: *the variance of a weighted sum of uncorrelated random variables (weights α_i) equals the weighted sum of the variances of the random variables (weights α_i^2).*

*5.2 JOINT CHARACTERISTIC FUNCTIONS

The *joint characteristic function* of two random variables X and Y is defined by

$$\Phi_{X,Y}(\omega_1, \omega_2) = E[e^{j\omega_1 X + j\omega_2 Y}] \tag{5.2-1}$$

where ω_1 and ω_2 are real numbers. An equivalent form is

$$\Phi_{X,Y}(\omega_1, \omega_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y) e^{j\omega_1 x + j\omega_2 y}\, dx\, dy \tag{5.2-2}$$

This expression is recognized as the two-dimensional Fourier transform (with signs of ω_1 and ω_2 reversed) of the joint density junction. From the inverse Fourier transform we also have

$$f_{X,Y}(x, y) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi_{X,Y}(\omega_1, \omega_2) e^{-j\omega_1 x - j\omega_2 y}\, d\omega_1\, d\omega_2 \tag{5.2-3}$$

By setting either $\omega_2 = 0$ or $\omega_1 = 0$ in (5.2-2), the characteristic functions of X or Y are obtained. They are called *marginal characteristic functions*:

$$\Phi_X(\omega_1) = \Phi_{X,Y}(\omega_1, 0) \tag{5.2-4}$$

$$\Phi_Y(\omega_2) = \Phi_{X,Y}(0, \omega_2) \tag{5.2-5}$$

Joint moments m_{nk} can be found from the joint characteristic function as follows:

$$m_{nk} = (-j)^{n+k} \left.\frac{\partial^{n+k}\Phi_{X,Y}(\omega_1, \omega_2)}{\partial\omega_1^n\, \partial\omega_2^k}\right|_{\omega_1=0,\, \omega_2=0} \tag{5.2-6}$$

This expression is the two-dimensional extension of (3.3-4).

Example 5.2-1 Two random variables X and Y have the joint characteristic function

$$\Phi_{X,Y}(\omega_1, \omega_2) = \exp(-2\omega_1^2 - 8\omega_2^2)$$

We show that X and Y are both zero-mean random variables and that they are uncorrelated.

The means derive from (5.2-6):

$$\bar{X} = E[X] = m_{10} = -j \left.\frac{\partial\Phi_{X,Y}(\omega_1, \omega_2)}{\partial\omega_1}\right|_{\omega_1=0,\, \omega_2=0}$$

$$= -j(-4\omega_1)\exp(-2\omega_1^2 - 8\omega_2^2)\Big|_{\omega_1=0,\, \omega_2=0} = 0$$

$$\bar{Y} = E[Y] = m_{01} = -j(-16\omega_2)\exp(-2\omega_1^2 - 8\omega_2^2)\Big|_{\omega_1=0,\, \omega_2=0} = 0$$

Also from (5.2-6):

$$R_{XY} = E[XY] = m_{11} = (-j)^2 \frac{\partial^2}{\partial\omega_1\, \partial\omega_2}[\exp(-2\omega_1^2 - 8\omega_2^2)]\Big|_{\omega_1=0,\, \omega_2=0}$$

$$= -(-4\omega_1)(-16\omega_2)\exp(-2\omega_1^2 - 8\omega_2^2)\Big|_{\omega_1=0,\, \omega_2=0} = 0$$

Since means are zero, $C_{XY} = R_{XY}$ from (5.1-14). Therefore, $C_{XY} = 0$ and X and Y are uncorrelated.

The joint characteristic function for N random variables $X_1, X_2, \ldots, X_N$ is defined by

$$\Phi_{X_1,\ldots,X_N}(\omega_1, \ldots, \omega_N) = E[e^{j\omega_1 X_1 + \cdots + j\omega_N X_N}] \tag{5.2-7}$$

Joint moments are obtained from

$$m_{n_1 n_2 \cdots n_N} = (-j)^R \frac{\partial^R \Phi_{X_1,\ldots,X_N}(\omega_1, \ldots, \omega_N)}{\partial\omega_1^{n_1}\, \partial\omega_2^{n_2} \cdots \partial\omega_N^{n_N}}\Bigg|_{\text{all } \omega_i = 0} \tag{5.2-8}$$

where

$$R = n_1 + n_2 + \cdots + n_N \tag{5.2-9}$$

5.3 JOINTLY GAUSSIAN RANDOM VARIABLES

Gaussian random variables are very important because they show up in nearly every area of science and engineering. In this section, the case of two gaussian random variables is first examined. The more advanced case of N random variables is then introduced.

Two Random Variables

Two random variables X and Y are said to be *jointly gaussian* if their joint density function is of the form

$$\boxed{f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\bar{X})^2}{\sigma_X^2} - \frac{2\rho(x-\bar{X})(y-\bar{Y})}{\sigma_X\sigma_Y} + \frac{(y-\bar{Y})^2}{\sigma_Y^2}\right]\right\}} \tag{5.3-1}$$

which is sometimes called the *bivariate gaussian density*. Here

$$\bar{X} = E[X] \tag{5.3-2}$$

$$\bar{Y} = E[Y] \tag{5.3-3}$$

$$\sigma_X^2 = E[(X-\bar{X})^2] \tag{5.3-4}$$

$$\sigma_Y^2 = E[(Y-\bar{Y})^2] \tag{5.3-5}$$

$$\rho = E[(X-\bar{X})(Y-\bar{Y})]/\sigma_X\sigma_Y \tag{5.3-6}$$

Figure 5.3-1*a* illustrates the appearance of the joint gaussian density function (5.3-1). Its maximum is located at the point $(\bar{X}, \bar{Y})$. The maximum value is obtained from

$$f_{X,Y}(x, y) \le f_{X,Y}(\bar{X}, \bar{Y}) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \tag{5.3-7}$$

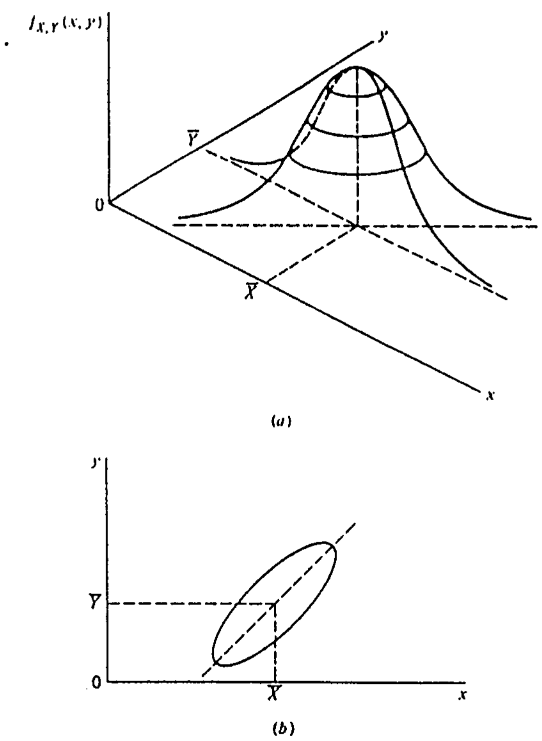


Figure 5.3-1 Sketch of the joint density function of two gaussian random variables.

The locus of constant values of $f_{X,Y}(x, y)$ will be an ellipse† as shown in Figure 5.3-1*b*. This is equivalent to saying that the line of intersection formed by slicing the function $f_{X,Y}(x, y)$ with a plane parallel to the xy plane is an ellipse.

Observe that if $\rho = 0$, corresponding to uncorrelated X and Y, (5.3-1) can be written as

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \tag{5.3-8}$$

where $f_X(x)$ and $f_Y(y)$ are the marginal density functions of X and Y given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left[-\frac{(x-\bar{X})^2}{2\sigma_X^2}\right] \tag{5.3-9}$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left[-\frac{(y-\bar{Y})^2}{2\sigma_Y^2}\right] \tag{5.3-10}$$

† When $\sigma_X = \sigma_Y$ and $\rho = 0$ the ellipse degenerates into a circle; when $\rho = +1$ or -1 the ellipses degenerate into axes rotated by angles $\pi/4$ and $-\pi/4$ respectively that pass through the point $(\bar{X}, \bar{Y})$.

Now the form of (5.3-8) is sufficient to guarantee that X and Y are statistically independent. Therefore we conclude that *any two uncorrelated gaussian random variables are also statistically independent*. It results that a coordinate rotation (linear transformation of X and Y) through an angle

$$\theta = \frac{1}{2}\tan^{-1}\left[\frac{2\rho\sigma_X\sigma_Y}{\sigma_X^2 - \sigma_Y^2}\right] \tag{5.3-11}$$

is sufficient to convert correlated random variables X and Y, having variances σ_X^2 and σ_Y^2, respectively, correlation coefficient ρ, and the joint density of (5.3-1), into two statistically independent gaussian random variables.†

By direct application of (4.4-12) and (4.4-13), the conditional density functions $f_X(x \mid Y = y)$ and $f_Y(y \mid X = x)$ can be found from the above expressions (see Problem 5-29).

Example 5.3-1 We show by example that (5.3-11) applies to arbitrary as well as gaussian random variables. Consider random variables Y_1 and Y_2 related to arbitrary random variables X and Y by the coordinate rotation

$$Y_1 = X\cos(\theta) + Y\sin(\theta)$$
$$Y_2 = -X\sin(\theta) + Y\cos(\theta)$$

If $\bar{X}$ and $\bar{Y}$ are the means of X and Y, respectively, the means of Y_1 and Y_2 are clearly $\bar{Y}_1 = \bar{X}\cos(\theta) + \bar{Y}\sin(\theta)$ and $\bar{Y}_2 = -\bar{X}\sin(\theta) + \bar{Y}\cos(\theta)$, respectively. The covariance of Y_1 and Y_2 is

$$\begin{aligned}
C_{Y_1Y_2} &= E[(Y_1 - \bar{Y}_1)(Y_2 - \bar{Y}_2)] \\
&= E[\{(X - \bar{X})\cos(\theta) + (Y - \bar{Y})\sin(\theta)\} \\
&\quad \cdot \{-(X - \bar{X})\sin(\theta) + (Y - \bar{Y})\cos(\theta)\}] \\
&= (\sigma_Y^2 - \sigma_X^2)\sin(\theta)\cos(\theta) + C_{XY}[\cos^2(\theta) - \sin^2(\theta)] \\
&= (\sigma_Y^2 - \sigma_X^2)(\tfrac{1}{2})\sin(2\theta) + C_{XY}\cos(2\theta)
\end{aligned}$$

Here $C_{XY} = E[(X - \bar{X})(Y - \bar{Y})] = \rho\sigma_X\sigma_Y$. If we require Y_1 and Y_2 to be uncorrelated, we must have $C_{Y_1Y_2} = 0$. By equating the above equation to zero we obtain (5.3-11). Thus, (5.3-11) applies to arbitrary as well as gaussian random variables.

---

† Wozencraft and Jacobs (1965), p. 155.

*N Random Variables

N random variables $X_1, X_2, \ldots, X_N$ are called *jointly gaussian* if their joint density function can be written as†

$$f_{X_1,\ldots,X_N}(x_1,\ldots,x_N) = \frac{\|[C_X]^{-1}\|^{1/2}}{(2\pi)^{N/2}} \exp\left\{-\frac{[x - \bar{X}]^t[C_X]^{-1}[x - \bar{X}]}{2}\right\} \tag{5.3-12}$$

where we define matrices

$$[x - \bar{X}] = \begin{bmatrix} x_1 - \bar{X}_1 \\ x_2 - \bar{X}_2 \\ \vdots \\ x_N - \bar{X}_N \end{bmatrix} \tag{5.3-13}$$

and

$$[C_X] = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1N} \\ C_{21} & C_{22} & \cdots & C_{2N} \\ \vdots & \vdots & & \vdots \\ C_{N1} & C_{N2} & \cdots & C_{NN} \end{bmatrix} \tag{5.3-14}$$

We use the notation $[\cdot]^t$ for the matrix transpose, $[\cdot]^{-1}$ for the matrix inverse, and $\|[\cdot]\|$ for the determinant. Elements of $[C_X]$, called the *covariance matrix* of the N random variables, are given by

$$C_{ij} = E[(X_i - \bar{X}_i)(X_j - \bar{X}_j)] = \begin{cases} \sigma_{X_i}^2 & i = j \\ C_{X_iX_j} & i \neq j \end{cases} \tag{5.3-15}$$

The density (5.3-12) is often called the *N-variate gaussian density* function.

For the special case where $N = 2$, the covariance matrix becomes

$$[C_X] = \begin{bmatrix} \sigma_{X_1}^2 & \rho\sigma_{X_1}\sigma_{X_2} \\ \rho\sigma_{X_1}\sigma_{X_2} & \sigma_{X_2}^2 \end{bmatrix} \tag{5.3-16}$$

so

$$[C_X]^{-1} = \frac{1}{(1 - \rho^2)}\begin{bmatrix} 1/\sigma_{X_1}^2 & -\rho/\sigma_{X_1}\sigma_{X_2} \\ -\rho/\sigma_{X_1}\sigma_{X_2} & 1/\sigma_{X_2}^2 \end{bmatrix} \tag{5.3-17}$$

$$\|[C_X]^{-1}\| = 1/\sigma_{X_1}^2\sigma_{X_2}^2(1 - \rho^2) \tag{5.3-18}$$

On substitution of (5.3-17) and (5.3-18) into (5.3-12), and letting $X_1 = X$ and $X_2 = Y$, it is easy to verify that the bivariate density of (5.3-1) results.

† We denote a matrix symbolically by use of heavy brackets $[\cdot]$.

*Some Properties of Gaussian Random Variables

We state without proof some of the properties exhibited by N jointly gaussian random variables $X_1, \ldots, X_N$.

1. Gaussian random variables are completely defined through only their first- and second-order moments; that is, by their means, variances, and covariances. This fact is readily apparent since only these quantities are needed to completely determine (5.3-12).
2. If the random variables are uncorrelated, they are also statistically independent. This property was given earlier for two variables.
3. Random variables produced by a linear transformation of $X_1, \ldots, X_N$ will also be gaussian, as proven in Section 5.5.
4. Any k-dimensional (k-variate) marginal density function obtained from the N-dimensional density function (5.3-12) by integrating out $N-k$ random variables will be gaussian. If the variables are ordered so that $X_1, \ldots, X_k$ occur in the marginal density and $X_{k+1}, \ldots, X_N$ are integrated out, then the covariance matrix of $X_1, \ldots, X_k$ is equal to the leading $k \times k$ submatrix of the covariance matrix of $X_1, \ldots, X_N$ (Wilks, 1962, p. 168).
5. The conditional density $f_{X_1, \ldots, X_k}(x_1, \ldots, x_k \mid X_{k+1} = x_{k+1}, \ldots, X_N = x_N)$ is gaussian (Papoulis, 1965, p. 257). This holds for any $k < N$.

*5.4 TRANSFORMATIONS OF MULTIPLE RANDOM VARIABLES

The function g in either (5.1-1) or (5.1-2) can be considered a transformation involving more than one random variable. By defining a new variable $Y = g(X_1, X_2, \ldots, X_N)$, we see that (5.1-2) is the expected value of Y. In calculating expected values it was not necessary to determine the density function of the new random variable Y. It may be, however, that the density function of Y is required in some practical problems, and its determination is briefly considered in this section.

In fact, one may be more generally interested in finding the joint density function for a set of new random variables

$$Y_i = T_i(X_1, X_2, \ldots, X_N) \qquad i = 1, 2, \ldots, N \tag{5.4-1}$$

defined by functional transformations T_i. Now all the possible cases described in Chapter 3 for one random variable carry over to the N-dimensional problem. That is, the X_i can be continuous, discrete, or mixed, while the functions T_i can be linear, nonlinear, continuous, segmented, etc. Because so many cases are possible, many of them being beyond our scope, we shall discuss only one representative problem.

We shall assume that the new random variables Y_i, given by (5.4-1), are produced by single-valued continuous functions T_i having continuous partial derivatives everywhere. It is further assumed that a set of inverse continuous functions T_j^{-1} exists such that the old variables may be expressed as single-valued continuous functions of the new variables:

$$X_j = T_j^{-1}(Y_1, Y_2, \ldots, Y_N) \qquad j = 1, 2, \ldots, N \tag{5.4-2}$$

These assumptions mean that a point in the joint sample space of the X_i maps into only one point in the space of the new variables Y_j.

Let R_X be a closed region of points in the space of the X_i and R_Y be the corresponding region of mapped points in the space of the Y_j, then the probability that a point falls in R_X will equal the probability that its mapped point falls in R_Y. These probabilities, in terms of joint densities, are given by

$$\int \cdots \int_{R_X} f_{X_1, \ldots, X_N}(x_1, \ldots, x_N)\, dx_1 \cdots dx_N = \int \cdots \int_{R_Y} f_{Y_1, \ldots, Y_N}(y_1, \ldots, y_N)\, dy_1 \cdots dy_N \tag{5.4-3}$$

This equation may be solved for $f_{Y_1, \ldots, Y_N}(y_1, \ldots, y_N)$ by treating it as simply a multiple integral involving a change of variables.

By working on the left side of (5.4-3) we change the variables x_i to new variables y_j by means of the variable changes (5.4-2). The integrand is changed by direct functional substitution. The limits change from the region R_X to the region R_Y. Finally, the differential hypervolume $dx_1 \cdots dx_N$ will change to the value $|J|\, dy_1 \cdots dy_N$ (Spiegel, 1963, p. 182), where $|J|$ is the magnitude of the jacobian† J of the transformations. The jacobian is the determinant of a matrix of derivatives defined by

$$J = \begin{vmatrix} \dfrac{\partial T_1^{-1}}{\partial Y_1} & \cdots & \dfrac{\partial T_1^{-1}}{\partial Y_N} \\ \vdots & & \vdots \\ \dfrac{\partial T_N^{-1}}{\partial Y_1} & \cdots & \dfrac{\partial T_N^{-1}}{\partial Y_N} \end{vmatrix} \tag{5.4-4}$$

Thus, the left side of (5.4-3) becomes

$$\int \cdots \int_{R_X} f_{X_1, \ldots, X_N}(x_1, \ldots, x_N)\, dx_1 \cdots dx_N = \int \cdots \int_{R_Y} f_{X_1, \ldots, X_N}(x_1 = T_1^{-1}, \ldots, x_N = T_N^{-1})\,|J|\, dy_1 \cdots dy_N \tag{5.4-5}$$

† After the German mathematician Karl Gustav Jakob Jacobi (1804–1851).

Since this result must equal the right side of (5.4-3), we conclude that

$$f_{Y_1,\ldots,Y_N}(y_1,\ldots,y_N)=f_{X_1,\ldots,X_N}(x_1=T_1^{-1},\ldots,x_N=T_N^{-1})|J| \quad (5.4\text{-}6)$$

When $N=1$, (5.4-6) reduces to (3.4-9) previously derived for a single random variable.

The solution (5.4-6) for the joint density of the new variables Y_j is illustrated here with an example.

Example 5.4-1 Let the transformations be linear and given by

$$Y_1=T_1(X_1,X_2)=aX_1+bX_2$$
$$Y_2=T_2(X_1,X_2)=cX_1+dX_2$$

where a, b, c, and d are real constants. The inverse functions are easy to obtain by solving these two equations for the two variables X_1 and X_2:

$$X_1=T_1^{-1}(Y_1,Y_2)=(dY_1-bY_2)/(ad-bc)$$
$$X_2=T_2^{-1}(Y_1,Y_2)=(-cY_1+aY_2)/(ad-bc)$$

where we shall assume $(ad-bc)\neq 0$. From (5.4-4):

$$J=\begin{vmatrix} d/(ad-bc) & -b/(ad-bc)\\ -c/(ad-bc) & a/(ad-bc)\end{vmatrix}=\frac{1}{(ad-bc)}$$

Finally, from (5.4-6),

$$f_{Y_1,Y_2}(y_1,y_2)=\frac{f_{X_1,X_2}\left(\dfrac{dy_1-by_2}{ad-bc},\dfrac{-cy_1+ay_2}{ad-bc}\right)}{|ad-bc|}$$

---

*5.5 LINEAR TRANSFORMATION OF GAUSSIAN RANDOM VARIABLES

Equation (5.4-6) can be readily applied to the problem of linearly transforming a set of gaussian random variables $X_1, X_2, \ldots, X_N$ for which the joint density of (5.3-12) applies. The new variables $Y_1, Y_2, \ldots, Y_N$ are

$$\begin{aligned} Y_1&=a_{11}X_1+a_{12}X_2+\cdots+a_{1N}X_N\\ Y_2&=a_{21}X_1+a_{22}X_2+\cdots+a_{2N}X_N\\ &\vdots\\ Y_N&=a_{N1}X_1+a_{N2}X_2+\cdots+a_{NN}X_N\end{aligned} \quad (5.5\text{-}1)$$

where the coefficients a_{ij}, i and $j=1,2,\ldots,N$, are real numbers. Now if we define the following matrices:

$$[T]=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N}\\ a_{21} & a_{22} & \cdots & a_{2N}\\ \vdots & \vdots & & \vdots\\ a_{N1} & a_{N2} & \cdots & a_{NN}\end{bmatrix} \quad (5.5\text{-}2)$$

$$[Y]=\begin{bmatrix}Y_1\\ \vdots\\ Y_N\end{bmatrix}\quad [\bar{Y}]=\begin{bmatrix}\bar{Y}_1\\ \vdots\\ \bar{Y}_N\end{bmatrix}\quad [X]=\begin{bmatrix}X_1\\ \vdots\\ X_N\end{bmatrix}\quad [\bar{X}]=\begin{bmatrix}\bar{X}_1\\ \vdots\\ \bar{X}_N\end{bmatrix} \quad (5.5\text{-}3)$$

then it is clear from (5.5-1) that

$$[Y]=[T][X] \qquad [Y-\bar{Y}]=[T][X-\bar{X}] \quad (5.5\text{-}4)$$
$$[X]=[T]^{-1}[Y] \qquad [X-\bar{X}]=[T]^{-1}[Y-\bar{Y}] \quad (5.5\text{-}5)$$

so long as $[T]$ is nonsingular. Thus,

$$X_i=T_i^{-1}(Y_1,\ldots,Y_N)=a^{i1}Y_1+a^{i2}Y_2+\cdots+a^{iN}Y_N \quad (5.5\text{-}6)$$

$$\frac{\partial X_i}{\partial Y_j}=\frac{\partial T_i^{-1}}{\partial Y_j}=a^{ij} \quad (5.5\text{-}7)$$

$$X_i-\bar{X}_i=a^{i1}(Y_1-\bar{Y}_1)+\cdots+a^{iN}(Y_N-\bar{Y}_N) \quad (5.5\text{-}8)$$

from (5.5-5). Here a^{ij} represents the ijth element of $[T]^{-1}$.

The density function of the new variables $Y_1,\ldots,Y_N$ is found by solving the right side of (5.4-6) in two steps. The first step is to determine $|J|$. By using (5.5-7) with (5.4-4) we find that J equals the determinant of the matrix $[T]^{-1}$. Hence,†

$$|J|=\|[T]^{-1}\|=\frac{1}{\|[T]\|} \quad (5.5\text{-}9)$$

The second step in solving (5.4-6) proceeds by using (5.5-8) to obtain

$$\begin{aligned} C_{X_iX_j}=E[(X_i-\bar{X}_i)(X_j-\bar{X}_j)]&=\sum_{k=1}^{N}a^{ik}\sum_{m=1}^{N}a^{jm}E[(Y_k-\bar{Y}_k)(Y_m-\bar{Y}_m)]\\ &=\sum_{k=1}^{N}a^{ik}\sum_{m=1}^{N}a^{jm}C_{Y_kY_m}\end{aligned} \quad (5.5\text{-}10)$$

Since $C_{X_iX_j}$ is the ijth element in the covariance matrix $[C_X]$ of (5.3-12) and $C_{Y_kY_m}$

† We represent the magnitude of the determinant of a matrix by $\|[\cdot]\|$.

is the kmth element in the covariance matrix of the new variables Y_i, which we denote $[C_Y]$, (5.5-10) can be written in the form

$$[C_X] = [T]^{-1}[C_Y]([T]^t)^{-1} \tag{5.5-11}$$

Here $[T]^t$ represents the transpose of $[T]$. The inverse of (5.5-11) is

$$[C_X]^{-1} = [T]^t[C_Y]^{-1}[T] \tag{5.5-12}$$

which has a determinant

$$|[C_X]^{-1}| = |[C_Y]^{-1}|\,|[T]|^2 \tag{5.5-13}$$

On substitution of (5.5-13) and (5.5-12) into (5.3-12):

$$f_{X_1,\ldots,X_N}(x_1 = T_1^{-1}, \ldots, x_N = T_N^{-1})$$
$$= \frac{|[T]|\,|[C_Y]^{-1}|^{1/2}}{(2\pi)^{N/2}} \exp\left\{-\frac{[x-\bar{X}]^t[T]^t[C_Y]^{-1}[T][x-\bar{X}]}{2}\right\} \tag{5.5-14}$$

Finally, (5.5-14) and (5.5-9) are substituted into (5.4-6), and (5.5-4) is used to obtain

$$f_{Y_1,\ldots,Y_N}(y_1,\ldots,y_N) = \frac{|[C_Y]^{-1}|^{1/2}}{(2\pi)^{N/2}} \exp\left\{-\frac{[y-\bar{Y}]^t[C_Y]^{-1}[y-\bar{Y}]}{2}\right\} \tag{5.5-15}$$

This result shows that the new random variables $Y_1, Y_2, \ldots, Y_N$ are jointly gaussian because (5.5-15) is of the required form.

In summary, (5.5-15) shows that a linear transformation of gaussian random variables produces gaussian random variables. The new variables have mean values

$$\bar{Y}_j = \sum_{k=1}^{N} a_{jk}\bar{X}_k \tag{5.5-16}$$

from (5.5-1) and covariances given by the elements of the covariance matrix

$$[C_Y] = [T][C_X][T]^t \tag{5.5-17}$$

as found from (5.5-11).

Example 5.5-1 Two gaussian random variables X_1 and X_2 have zero means and variances $\sigma_{X_1}^2 = 4$ and $\sigma_{X_2}^2 = 9$. Their covariance $C_{X_1X_2}$ equals 3. If X_1 and X_2 are linearly transformed to new variables Y_1 and Y_2 according to

$$Y_1 = X_1 - 2X_2$$
$$Y_2 = 3X_1 + 4X_2$$

we use the above results to find the means, variances, and covariance of Y_1 and Y_2.

Here

$$[T] = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad [C_X] = \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}$$

Since X_1 and X_2 are zero-mean and gaussian, Y_1 and Y_2 will also be zero-mean and gaussian, thus $\bar{Y}_1 = 0$ and $\bar{Y}_2 = 0$. From (5.5-17):

$$[C_Y] = [T][C_X][T]^t = \begin{bmatrix} 1 & -2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 28 & -66 \\ -66 & 252 \end{bmatrix}$$

Thus, $\sigma_{Y_1}^2 = 28$, $\sigma_{Y_2}^2 = 252$, and $C_{Y_1Y_2} = -66$.

---

*5.6 COMPLEX RANDOM VARIABLES

A *complex random variable* Z can be defined in terms of real random variables X and Y by

$$Z = X + jY \tag{5.6-1}$$

where $j = \sqrt{-1}$. In considering expected values involving Z, the joint density of X and Y must be used. For instance, if $g(\cdot)$ is some function (real or complex) of Z, the expected value of $g(Z)$ is obtained from

$$E[g(Z)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(z) f_{X,Y}(x, y)\,dx\,dy \tag{5.6-2}$$

Various important quantities such as the mean value and variance are obtained through application of (5.6-2). The mean value of Z is

$$\bar{Z} = E[Z] = E[X] + jE[Y] = \bar{X} + j\bar{Y} \tag{5.6-3}$$

The variance σ_Z^2 of Z is defined as the mean value of the function $g(Z) = |Z - E[Z]|^2$; that is,

$$\sigma_Z^2 = E[|Z - E[Z]|^2] \tag{5.6-4}$$

Equation (5.6-2) can be extended to include functions of two random variables

$$Z_m = X_m + jY_m \tag{5.6-5}$$

and

$$Z_n = X_n + jY_n \tag{5.6-6}$$

$n \neq m$, if expectation is taken with respect to four random variables X_m, Y_m, X_n, Y_n through their joint density function $f_{X_m,Y_m,X_n,Y_n}(x_m, y_m, x_n, y_n)$. If this density satisfies

$$f_{X_m,Y_m,X_n,Y_n}(x_m, y_m, x_n, y_n) = f_{X_m,Y_m}(x_m, y_m) f_{X_n,Y_n}(x_n, y_n) \tag{5.6-7}$$

then Z_m and Z_n are called *statistically independent*. The extension to N random variables is straightforward.

The *correlation* and *covariance* of Z_m and Z_n are defined by

$$R_{Z_m Z_n} = E[Z_m^* Z_n] \qquad n \neq m \tag{5.6-8}$$

and

$$C_{Z_m Z_n} = E[\{Z_m - E[Z_m]\}^*\{Z_n - E[Z_n]\}] \qquad n \neq m \tag{5.6-9}$$

respectively, where the superscripted asterisk* represents the complex conjugate. If the covariance is 0, Z_m and Z_n are said to be *uncorrelated random variables*. By setting (5.6-9) to 0, we find that

$$R_{Z_m Z_n} = E[Z_m^*]E[Z_n] \qquad m \neq n \tag{5.6-10}$$

for uncorrelated random variables. Statistical independence is sufficient to guarantee that Z_m and Z_n are uncorrelated.

Finally, we note that two complex random variables are called *orthogonal* if their correlation, given by (5.6-8), equals 0.

PROBLEMS

5-1 Random variables X and Y have the joint density

$$f_{X,Y}(x, y) = \begin{cases} \dfrac{1}{24} & 0 < x < 6 \quad \text{and} \quad 0 < y < 4 \\ 0 & \text{elsewhere} \end{cases}$$

What is the expected value of the function $g(X, Y) = (XY)^2$?

5-2 Extend Problem 5-1 by finding the expected value of $g(X_1, X_2, X_3, X_4) = X_1^{n_1}X_2^{n_2}X_3^{n_3}X_4^{n_4}$, where n_1, n_2, n_3, and n_4 are integers ≥ 0 and

$$f_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4) = \begin{cases} \dfrac{1}{abcd} & 0 < x_1 < a \text{ and } 0 < x_2 < b \text{ and } 0 < x_3 < c \\ & \text{and } 0 < x_4 < d \\ 0 & \text{elsewhere} \end{cases}$$

5-3 The density function of two random variables X and Y is

$$f_{X,Y}(x, y) = u(x)u(y)16e^{-4(x+y)}$$

Find the mean value of the function

$$g(X, Y) = \begin{cases} 5 & 0 < X \leq \frac{1}{2} \quad \text{and} \quad 0 < Y \leq \frac{1}{2} \\ -1 & \frac{1}{2} < X \quad \text{and/or} \quad \frac{1}{2} < Y \\ 0 & \text{all other } X \text{ and } Y \end{cases}$$

5-4 For the random variables in Problem 5-3, find the mean value of the function

$$g(X, Y) = e^{-2(X^2+Y^2)}$$

5-5 Three statistically independent random variables X_1, X_2, and X_3 have mean values $\bar{X}_1 = 3$, $\bar{X}_2 = 6$, and $\bar{X}_3 = -2$. Find the mean values of the following functions:

(*a*) $g(X_1, X_2, X_3) = X_1 + 3X_2 + 4X_3$
(*b*) $g(X_1, X_2, X_3) = X_1X_2X_3$
(*c*) $g(X_1, X_2, X_3) = -2X_1X_2 - 3X_1X_3 + 4X_2X_3$
(*d*) $g(X_1, X_2, X_3) = X_1 + X_2 + X_3$

5-6 Find the mean value of the function

$$g(X, Y) = X^2 + Y^2$$

where X and Y are random variables defined by the density function

$$f_{X,Y}(x, y) = \frac{e^{-(x^2+y^2)/2\sigma^2}}{2\pi\sigma^2}$$

with σ^2 a constant.

5-7 Two statistically independent random variables X and Y have mean values $\bar{X} = E[X] = 2$ and $\bar{Y} = E[Y] = 4$. They have second moments $\overline{X^2} = E[X^2] = 8$ and $\overline{Y^2} = E[Y^2] = 25$. Find:

(*a*) the mean value (*b*) the second moment and
(*c*) the variance of the random variable $W = 3X - Y$.

5-8 Two random variables X and Y have means $\bar{X} = 1$ and $\bar{Y} = 2$, variances $\sigma_X^2 = 4$ and $\sigma_Y^2 = 1$, and a correlation coefficient $\rho_{XY} = 0.4$. New random variables W and V are defined by

$$V = -X + 2Y \qquad W = X + 3Y$$

Find:

(*a*) the means (*b*) the variances (*c*) the correlation and
(*d*) the correlation coefficient ρ_{VW} of V and W.

5-9 Two random variables X and Y are related by the expression

$$Y = aX + b$$

where a and b are any real numbers.

(*a*) Show that their correlation coefficient is

$$\rho = \begin{cases} 1 & \text{if } a > 0 \text{ for any } b \\ -1 & \text{if } a < 0 \text{ for any } b \end{cases}$$

(*b*) Show that their covariance is

$$C_{XY} = a\sigma_X^2$$

where σ_X^2 is the variance of X.

*5-10 Show that the correlation coefficient satisfies the expression

$$|\rho| = \frac{|\mu_{11}|}{\sqrt{\mu_{02}\mu_{20}}} \leq 1$$

5-11 Find all the second-order moments and central moments for the density function given in Problem 5-3.

5-12 Random variables X and Y have the joint density function

$$f_{X,Y}(x, y) = \begin{cases} (x+y)^2/40 & -1 < x < 1 \quad \text{and} \quad -3 < y < 3 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find all the second-order moments of X and Y.
(b) What are the variances of X and Y?
(c) What is the correlation coefficient?

5-13 Find all the third-order moments by using (5.1-5) for X and Y defined in Problem 5-12.

5-14 For discrete random variables X and Y, show that:
(a) Joint moments are

$$m_{nk} = \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i, y_j) x_i^n y_j^k$$

(b) Joint central moments are

$$\mu_{nk} = \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i, y_j)(x_i - \bar{X})^n (y_j - \bar{Y})^k$$

where $P(x_i, y_j) = P\{X = x_i, Y = y_j\}$, X has N possible values x_i, and Y has M possible values y_j.

5-15 For two random variables X and Y:

$$f_{X,Y}(x, y) = 0.15\delta(x+1)\delta(y) + 0.1\delta(x)\delta(y) + 0.1\delta(x)\delta(y-2) + 0.4\delta(x-1)\delta(y+2) + 0.2\delta(x-1)\delta(y-1) + 0.05\delta(x-1)\delta(y-3)$$

Find: (a) the correlation, (b) the covariance, and (c) the correlation coefficient of X and Y.
(d) Are X and Y either uncorrelated or orthogonal?

5-16 Discrete random variables X and Y have the joint density

$$f_{X,Y}(x, y) = 0.4\delta(x+\alpha)\delta(y-2) + 0.3\delta(x-\alpha)\delta(y-2) + 0.1\delta(x-\alpha)\delta(y-\alpha) + 0.2\delta(x-1)\delta(y-1)$$

Determine the value of α, if any, that minimizes the correlation between X and Y and find the minimum correlation. Are X and Y orthogonal?

5-17 For two discrete random variables X and Y:

$$f_{X,Y}(x, y) = 0.3\delta(x-\alpha)\delta(y-\alpha) + 0.5\delta(x+\alpha)\delta(y-4) + 0.2\delta(x+2)\delta(y+2)$$

Determine the value of α, if any, that minimizes the covariance of X and Y. Find the minimum covariance. Are X and Y uncorrelated?

5-18 The density function

$$f_{X,Y}(x, y) = \begin{cases} \dfrac{xy}{9} & 0 < x < 2 \quad \text{and} \quad 0 < y < 3 \\ 0 & \text{elsewhere} \end{cases}$$

applies to two random variables X and Y.
(a) Show, by use of (5.1-6) and (5.1-7), that X and Y are uncorrelated.
(b) Show that X and Y are also statistically independent.

5-19 Two random variables X and Y have the density function

$$f_{X,Y}(x, y) = \begin{cases} \dfrac{2}{43}(x+0.5y)^2 & 0 < x < 2 \quad \text{and} \quad 0 < y < 3 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Find all the first- and second-order moments.
(b) Find the covariance.
(c) Are X and Y uncorrelated?

5-20 Define random variables V and W by

$$V = X + aY$$
$$W = X - aY$$

where a is a real number and X and Y are random variables. Determine a in terms of moments of X and Y such that V and W are orthogonal.

*5-21 If X and Y in Problems 5-20 are gaussian, show that W and V are statistically independent if $a^2 = \sigma_X^2/\sigma_Y^2$, where σ_X^2 and σ_Y^2 are the variances of X and Y, respectively.

5-22 Three uncorrelated random variables X_1, X_2, and X_3 have means $\bar{X}_1 = 1$, $\bar{X}_2 = -3$, and $\bar{X}_3 = 1.5$ and second moments $E[X_1^2] = 2.5$, $E[X_2^2] = 11$, and $E[X_3^2] = 3.5$. Let $Y = X_1 - 2X_2 + 3X_3$ be a new random variable and find:
(a) the mean value, (b) the variance of Y.

5-23 Given $W = (aX + 3Y)^2$ where X and Y are zero-mean random variables with variances $\sigma_X^2 = 4$ and $\sigma_Y^2 = 16$. Their correlation coefficient is $\rho = -0.5$.
(a) Find a value for the parameter a that minimizes the mean value of W.
(b) Find the minimum mean value.

*5-24 Find the joint characteristic function for X and Y defined in Problem 5-3.

*5-25 Show that the joint characteristic function of N independent random variables X_i, having characteristic functions $\Phi_{X_i}(\omega_i)$ is

$$\Phi_{X_1, \ldots, X_N}(\omega_1, \ldots, \omega_N) = \prod_{i=1}^{N} \Phi_{X_i}(\omega_i)$$

*5-26 For N random variables, show that

$$|\Phi_{X_1, \ldots, X_N}(\omega_1, \ldots, \omega_N)| \le \Phi_{X_1, \ldots, X_N}(0, \ldots, 0) = 1$$

*5-27 For two zero-mean gaussian random variables X and Y, show that their joint characteristic function is

$$\Phi_{X,Y}(\omega_1, \omega_2) = \exp\{-\tfrac{1}{2}[\sigma_X^2\omega_1^2 + 2\rho\sigma_X\sigma_Y\omega_1\omega_2 + \sigma_Y^2\omega_2^2]\}$$

*5-28 Zero-mean gaussian random variables X and Y have variances $\sigma_X^2 = 3$ and $\sigma_Y^2 = 4$, respectively, and a correlation coefficient $\rho = -\tfrac{1}{4}$.

(a) Write an expression for the joint density function.

(b) Show that a rotation of coordinates through the angle given by (5.3-11) will produce new statistically independent random variables.

*5-29 Find the conditional density functions $f_X(x \mid Y = y)$ and $f_Y(y \mid X = x)$ applicable to two gaussian random variables X and Y defined by (5.3-1) and show that they are also gaussian.

*5-30 Zero-mean gaussian random variables X_1, X_2, and X_3 having a covariance matrix

$$[C_X] = \begin{bmatrix} 4 & 2.05 & 1.05 \\ 2.05 & 4 & 2.05 \\ 1.05 & 2.05 & 4 \end{bmatrix}$$

are transformed to new variables

$$Y_1 = 5X_1 + 2X_2 - X_3$$
$$Y_2 = -X_1 + 3X_2 + X_3$$
$$Y_3 = 2X_1 - X_2 + 2X_3$$

(a) Find the covariance matrix of Y_1, Y_2, and Y_3.

(b) Write an expression for the joint density function of Y_1, Y_2, and Y_3.

*5-31 A complex random variable Z is defined by

$$Z = \cos(X) + j\sin(Y)$$

where X and Y are independent real random variables uniformly distributed from $-\pi$ to π.

(a) Find the mean value of Z.

(b) Find the variance of Z.

ADDITIONAL PROBLEMS

5-32 Two random variables have a uniform density on a circular region defined by

$$f_{X,Y}(x, y) = \begin{cases} 1/\pi r^2 & x^2 + y^2 \le r^2 \\ 0 & \text{elsewhere} \end{cases}$$

Find the mean value of the function $g(X, Y) = X^2 + Y^2$.

*5-33 Define the conditional expected value of a function $g(X, Y)$ of random variables X and Y as

$$E[g(X, Y) \mid B] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y \mid B)\, dx\, dy$$

(a) If event B is defined as $B = \{y_a < Y \le y_b\}$, where $y_a < y_b$ are constants, evaluate $E[g(X, Y) \mid B]$. (*Hint:* Use results of Problem 4-55.)

(b) If B is defined by $B = \{Y = y\}$ what does the conditional expected value of part (a) become?

5-34 For random variables X and Y having $\bar{X} = 1$, $\bar{Y} = 2$, $\sigma_X^2 = 6$, $\sigma_Y^2 = 9$, and $\rho = -\tfrac{2}{3}$, find (a) the covariance of X and Y, (b) the correlation of X and Y, and (c) the moments m_{20} and m_{02}.

5-35 $\bar{X} = \tfrac{1}{2}$, $\overline{X^2} = \tfrac{5}{2}$, $\bar{Y} = 2$, $\overline{Y^2} = \tfrac{19}{2}$, and $C_{XY} = -1/2\sqrt{3}$ for random variables X and Y.

(a) Find σ_X^2, σ_Y^2, R_{XY}, and ρ.

(b) What is the mean value of the random variable $W = (X + 3Y)^2 + 2X + 3$?

5-36 Let X and Y be statistically independent random variables with $\bar{X} = \tfrac{3}{4}$, $\overline{X^2} = 4$, $\bar{Y} = 1$, and $\overline{Y^2} = 5$. For a random variable $W = X - 2Y + 1$ find (a) R_{XY}, (b) R_{XW}, (c) R_{YW}, and (d) C_{XY}. (e) Are X and Y uncorrelated?

5-37 Statistically independent random variables X and Y have moments $m_{10} = 2$, $m_{20} = 14$, $m_{02} = 12$, and $m_{11} = -6$. Find the moment μ_{22}.

5-38 A joint density is given as

$$f_{X,Y}(x, y) = \begin{cases} x(y + 1.5) & 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find all the joint moments m_{nk}, n and $k = 0, 1, \ldots$.

5-39 Find all the joint central moments μ_{nk}, n and $k = 0, 1, \ldots$, for the density of Problem 5-38.

*5-40 Find the joint characteristic function for random variables X and Y defined by

$$f_{X,Y}(x, y) = (1/2\pi)\,\mathrm{rect}(x/\pi)\,\mathrm{rect}[(x + y)/\pi]\cos(x + y)$$

Use the result to find the marginal characteristic functions of X and Y.

*5-41 Random variables X_1 and X_2 have the joint characteristic function

$$\Phi_{X_1, X_2}(\omega_1, \omega_2) = [(1 - j2\omega_1)(1 - j2\omega_2)]^{-N/2}$$

where $N > 0$ is an integer.

(a) Find the correlation and moments m_{20} and m_{02}.
(b) Determine the means of X_1 and X_2.
(c) What is the correlation coefficient?

*5-42 The joint probability density of two discrete random variables X and Y consists of impulses located at all lattice points (mb, nd), where $m = 0, 1, \ldots, M$ and $n = 1, 2, \ldots, N$ with $b > 0$ and $d > 0$ being constants. All possible points are equally probable. Determine the joint characteristic function.

*5-43 Let X_k, $k = 1, 2, \ldots, K$, be statistically independent Poisson random variables, each with its own variance b_k (Problem 3-16). Show that the sum $X = X_1 + X_2 + \cdots + X_K$ is a Poisson random variable. (*Hint:* Use results of Problems 5-25 and 3-53.)

5-44 Assume $\sigma_X = \sigma_Y = \sigma$ in (5.3-1) and show that the locus of the maximum of the joint density is a line passing through the point $(\bar{X}, \bar{Y})$ with slope $\pi/4$ (or $-\pi/4$) when $\rho = 1$ (or -1).

5-45 Two gaussian random variables X and Y have variances $\sigma_X^2 = 9$ and $\sigma_Y^2 = 4$, respectively, and correlation coefficient ρ. It is known that a coordinate rotation by an angle $-\pi/8$ results in new random variables Y_1 and Y_2 that are uncorrelated. What is ρ?

*5-46 Let X and Y be jointly gaussian random variables where $\sigma_X^2 = \sigma_Y^2$ and $\rho = -1$. Find a transformation matrix such that new random variables Y_1 and Y_2 are statistically independent.

*5-47 Random variables X and Y having the joint density

$$f_{X,Y}(x, y) = (8/3)u(x - 2)u(y - 1)xy^2 \exp(4 - 2xy)$$

undergo a transformation

$$[T] = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

to generate new random variables Y_1 and Y_2.

(a) Find the joint density of Y_1 and Y_2.
(b) Show what points in the $y_1 y_2$ plane correspond to a nonzero value of the new density.

*5-48 Equation (5.4-5) can sometimes be used to find the density of a single function of several random variables if *auxiliary random variables* are used. Apply the idea to finding the density function of $Z = aXY$, where a is a constant and X and Y are random variables, by defining the auxiliary variable $W = X$.

*5-49 Apply the method of Problem 5-48 to finding the density function of $Z = bY/X$, with b a constant, when using the auxiliary variable $W = X$.

*5-50 Two gaussian random variables X_1 and X_2 are defined by the mean and covariance matrices

$$[\bar{X}] = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \qquad [C_X] = \begin{bmatrix} 5 & -2/\sqrt{5} \\ -2/\sqrt{5} & 4 \end{bmatrix}$$

Two new random variables Y_1 and Y_2 are formed using the transformation

$$[T] = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}$$

Find matrices (a) $[\bar{Y}]$ and (b) $[C_Y]$. (c) Also find the correlation coefficient of Y_1 and Y_2.

*5-51 Complex random variables Z_1 and Z_2 have zero means. The correlation of the real parts of Z_1 and Z_2 is 4, while the correlation of the imaginary parts is 6. The real part of Z_1 and the imaginary part of Z_2 are statistically independent as a pair, as are the imaginary part of Z_1 and the real part of Z_2.

(a) What is the correlation of Z_1 and Z_2?
(b) Are Z_1 and Z_2 statistically independent?

CHAPTER
SIX
RANDOM PROCESSES

6.0 INTRODUCTION

In the real world of engineering and science, it is necessary that we be able to deal with time waveforms. Indeed, we frequently encounter *random* time waveforms in practical systems. More often than not, a *desired* signal in some system is random. For example, the bit stream in a binary communication system is a random message because each bit in the stream occurs randomly. On the other hand, a desired signal is often accompanied by an *undesired* random waveform, noise. The noise interferes with the message and ultimately limits the performance of the system. Thus, any hope we have of determining the performance of systems with random waveforms hinges on our ability to describe and deal with such waveforms. In this chapter we introduce concepts that allow the description of random waveforms in a probabilistic sense.

6.1 THE RANDOM PROCESS CONCEPT

The concept of a random process is based on enlarging the random variable concept to include time. Since a random variable X is, by its definition, a function of the possible outcomes s of an experiment, it now becomes a function of both s and time. In other words, we assign, according to some rule, a time function

$$x(t, s) \tag{6.1-1}$$

to every outcome s. The family of all such functions, denoted $X(t, s)$, is called a *random process*. As with random variables where x was denoted as a specific value of the random variable X, we shall often use the convenient short-form

notation $x(t)$ to represent a specific waveform of a random process denoted by $X(t)$.

Clearly, a random process $X(t, s)$ represents a family or *ensemble* of time functions when t and s are variables. Figure 6.1-1 illustrates a few members of an ensemble. Each member time function is called a *sample function*, *ensemble member*, or sometimes a *realization* of the process. Thus, a random process also represents a *single* time function when t is a variable and s is fixed at a specific value (outcome).

A random process also represents a random variable when t is fixed and s is a variable. For example, the random variable $X(t_1, s) = X(t_1)$ is obtained from the process when time is "frozen" at the value t_1. We often use the notation X_1 to denote the random variable associated with the process $X(t)$ at time t_1. X_1 corresponds to a vertical "slice" through the ensemble at time t_1, as illustrated in Figure 6.1-1. The statistical properties of $X_1 = X(t_1)$ describe the statistical properties of the random process at time t_1. The expected value of X_1 is called the *ensemble average* as well as the expected or mean value of the random process (at time t_1). Since t_1 may have various values, the mean value of a process may not be constant; in general, it may be a function of time. We easily visualize any

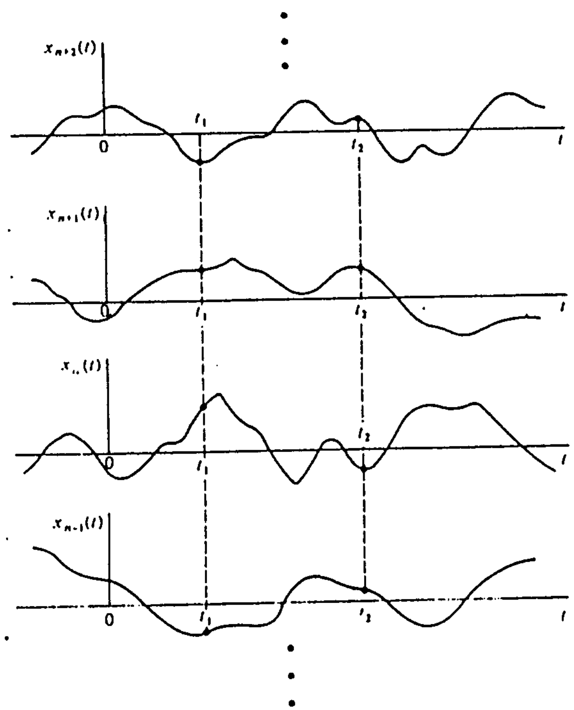


Figure 6.1-1 A continuous random process. [*Reproduced from Peebles (1976) with permission of publishers Addison-Wesley, Advanced Book Program.*]

number of random variables X_i derived from a random process $X(t)$ at times t_i, $i = 1, 2, \ldots$:

$$X_i = X(t_i, s) = X(t_i) \tag{6.1-2}$$

A random process can also represent a mere number when t and s are both fixed.

Classification of Processes

It is convenient to classify random processes according to the characteristics of t and the random variable $X = X(t)$ at time t. We shall consider only four cases based on t and X having values in the ranges $-\infty < t < \infty$ and $-\infty < x < \infty$.†

† Other cases can be defined based on a definition of random processes on a finite time interval (see for example: Rosenblatt (1974), p. 91; Prabhu (1965), p. 1; Miller (1974), p. 31; Parzen (1962), p. 7; Dubes (1968), p. 320; Ross (1972), p. 56). Other recent texts on random processes are Helstrom (1984), and Gray and Davisson (1986).

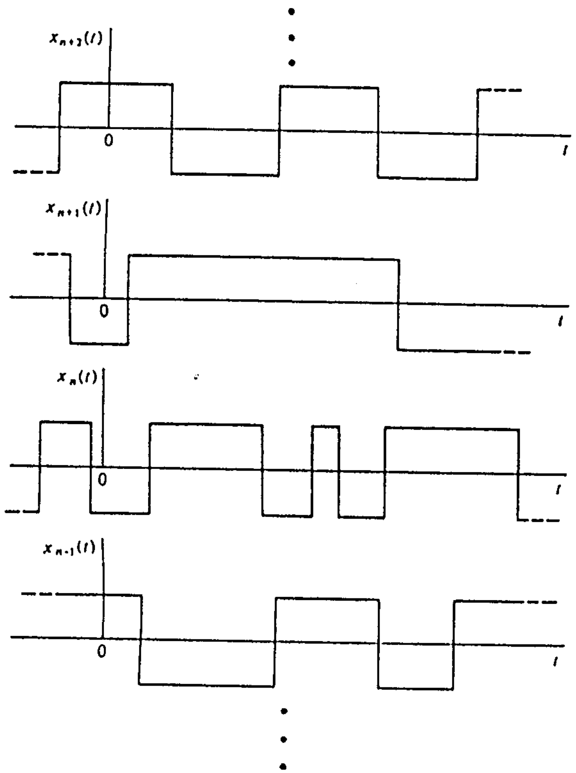


Figure 6.1-2 A discrete random process formed by heavily limiting the waveforms of Figure 6.1-1. [*Reproduced from Peebles (1976) with permission of publishers Addison-Wesley, Advanced Book Program.*]

If X is continuous and t can have any of a continuum of values, then $X(t)$ is called a *continuous random process*. Figure 6.1-1 is an illustration of this class of process. Thermal noise generated by any realizable network is a practical example of a waveform that is modeled as a sample function of a continuous random process. In this example, the network is the outcome in the underlying random experiment of selecting a network. (The presumption is that many networks are available from which to choose; this may not be the case in the real world, but it should not prevent us from imagining a production line producing any number of similar networks.) Each network establishes a sample function, and all sample functions form the process.†

A second class of random process, called a *discrete random process*, corresponds to the random variable X having only discrete values while t is continuous. Figure 6.1-2 illustrates such a process derived by heavily limiting the sample functions shown in Figure 6.1-1. The sample functions have only two dis-

† Note that finding the mean value of the process at any time t is equivalent to finding the average voltage that would be produced by all the various networks at time t.

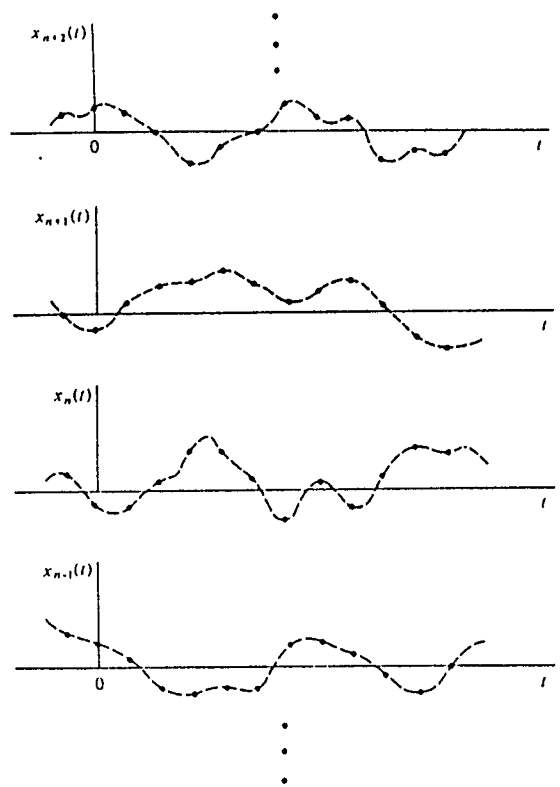


Figure 6.1-3 A continuous random sequence formed by sampling the waveforms of Figure 6.1-1. [*Reproduced from Peebles (1976), with permission of publishers Addison-Wesley, Advanced Book Program.*]

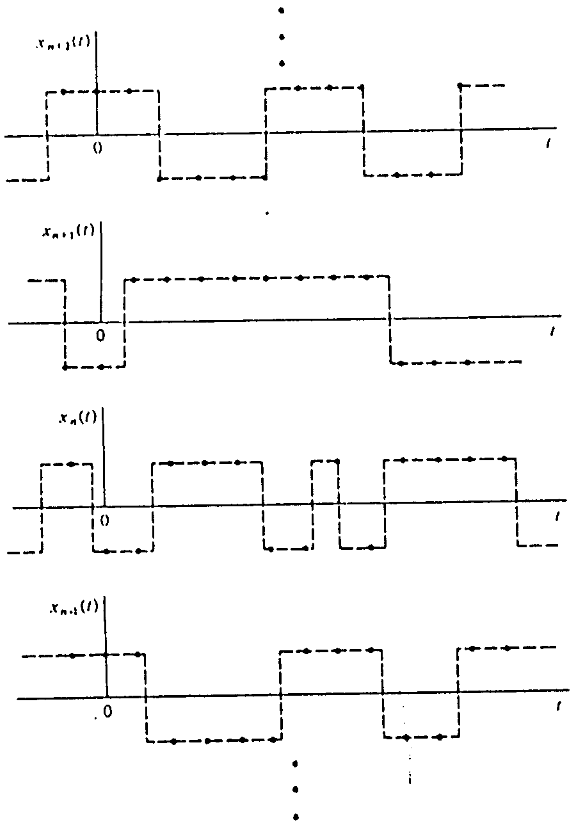


Figure 6.1-4 A discrete random sequence formed by sampling the waveforms of Figure 6.1-2. [*Adapted from Peebles (1976) with permission of publishers Addison-Wesley, Advanced Book Program.*]

crete values: the positive level is generated whenever a sample function in Figure 6.1-1 is positive and the negative level occurs for other times.

A random process for which X is continuous but time has only discrete values is called a *continuous random sequence* (Thomas, 1969, p. 80). Such a sequence can be formed by periodically sampling the ensemble members of Figure 6.1-1. The result is illustrated in Figure 6.1-3.

A fourth class of random process, called a *discrete random sequence*, corresponds to both time and the random variable being discrete. Figure 6.1-4 illustrates a discrete random sequence developed by sampling the sample functions of Figure 6.1-2.

In this text we are concerned almost entirely with discrete and continuous random processes.

Deterministic and Nondeterministic Processes

In addition to the classes described above, a random process can be described by the form of its sample functions. If future values of any sample function cannot be predicted exactly from observed past values, the process is called *nondeterministic*. The process of Figure 6.1-1 is one example.

A process is called *deterministic* if future values of any sample function can be predicted from past values. An example is the random process defined by

$$X(t) = A \cos(\omega_0 t + \Theta) \tag{6.1-3}$$

Here A, Θ, or ω_0 (or all) may be random variables. Any one sample function corresponds to (6.1-3) with particular values of these random variables. Therefore, knowledge of the sample function prior to any time instant automatically allows prediction of the sample function's future values because its form is known.

6.2 STATIONARITY AND INDEPENDENCE

As previously stated, a random process becomes a random variable when time is fixed at some particular value. The random variable will possess statistical properties, such as a mean value, moments, variance, etc., that are related to its density function. If *two* random variables are obtained from the process for two time instants, they will have statistical properties (means, variances, joint moments, etc.) related to their joint density function. More generally, N random variables will possess statistical properties related to their N-dimensional joint density function.

Broadly speaking, a random process is said to be *stationary* if all its statistical properties do not change with time. Other processes are called *nonstationary*. These statements are not intended as definitions of stationarity but are meant to convey only a general meaning. More concrete definitions follow. Indeed, there are several "levels" of stationarity, all of which depend on the density functions of the random variables of the process.

Distribution and Density Functions

To define stationarity, we must first define distribution and density functions as they apply to a random process $X(t)$. For a particular time t_1, the distribution function associated with the random variable $X_1 = X(t_1)$ will be denoted $F_X(x_1; t_1)$. It is defined as†

$$F_X(x_1; t_1) = P\{X(t_1) \le x_1\} \tag{6.2-1}$$

† $F_X(x_1; t_1)$ is known as the *first-order distribution function* of the process $X(t)$.

for any real number x_1. This is the same definition used all along for the distribution function of one random variable. Only the notation has been altered to reflect the fact that it is possibly now a function of time choice t_1.

For two random variables $X_1 = X(t_1)$ and $X_2 = X(t_2)$, the *second-order joint distribution function* is the two-dimensional extension of (6.2-1):

$$F_X(x_1, x_2; t_1, t_2) = P\{X(t_1) \le x_1, X(t_2) \le x_2\} \tag{6.2-2}$$

In a similar manner, for N random variables $X_i = X(t_i)$, $i = 1, 2, \ldots, N$, the *Nth-order joint distribution function* is

$$F_X(x_1, \ldots, x_N; t_1, \ldots, t_N) = P\{X(t_1) \le x_1, \ldots, X(t_N) \le x_N\} \tag{6.2-3}$$

Joint density functions of interest are found from appropriate derivatives of the above three relationships:†

$$f_X(x_1; t_1) = dF_X(x_1; t_1)/dx_1 \tag{6.2-4}$$

$$f_X(x_1, x_2; t_1, t_2) = \partial^2 F_X(x_1, x_2; t_1, t_2)/(\partial x_1\, \partial x_2) \tag{6.2-5}$$

$$f_X(x_1, \ldots, x_N; t_1, \ldots, t_N) = \partial^N F_X(x_1, \ldots, x_N; t_1, \ldots, t_N)/(\partial x_1 \cdots \partial x_N) \tag{6.2-6}$$

Statistical Independence

Two processes $X(t)$ and $Y(t)$ are *statistically independent* if the random variable group $X(t_1), X(t_2), \ldots, X(t_N)$ is independent of the group $Y(t'_1), Y(t'_2), \ldots, Y(t'_M)$ for any choice of times $t_1, t_2, \ldots, t_N, t'_1, t'_2, \ldots, t'_M$. Independence requires that the joint density be factorable by groups:

$$\begin{aligned} f_{X,Y}(x_1, \ldots, x_N, y_1, \ldots, y_M; t_1, \ldots, t_N, t'_1, \ldots, t'_M) \\ = f_X(x_1, \ldots, x_N; t_1, \ldots, t_N) f_Y(y_1, \ldots, y_M; t'_1, \ldots, t'_M) \end{aligned} \tag{6.2-7}$$

First-Order Stationary Processes

A random process is called *stationary to order one* if its first-order density function does not change with a shift in time origin. In other words

$$f_X(x_1; t_1) = f_X(x_1; t_1 + \Delta) \tag{6.2-8}$$

must be true for any t_1 and any real number Δ if $X(t)$ is to be a first-order stationary process.

Consequences of (6.2-8) are that $f_X(x_1; t_1)$ is independent of t_1 and the process mean value $E[X(t)]$ is a constant:

$$E[X(t)] = \bar{X} = \text{constant} \tag{6.2-9}$$

† Analogous to distribution functions, these are called *first-*, *second-*, and *Nth-order density functions*, respectively.

To prove (6.2-9), we find mean values of the random variables $X_1 = X(t_1)$ and $X_2 = X(t_2)$. For X_1:

$$E[X_1] = E[X(t_1)] = \int_{-\infty}^{\infty} x_1 f_X(x_1; t_1)\, dx_1 \tag{6.2-10}$$

For X_2:

$$E[X_2] = E[X(t_2)] = \int_{-\infty}^{\infty} x_1 f_X(x_1; t_2)\, dx_1 \dagger \tag{6.2-11}$$

Now by letting $t_2 = t_1 + \Delta$ in (6.2-11), substituting (6.2-8), and using (6.2-10), we get

$$E[X(t_1 + \Delta)] = E[X(t_1)] \tag{6.2-12}$$

which must be a constant because t_1 and Δ are arbitrary.

Second-Order and Wide-Sense Stationarity

A process is called *stationary to order two* if its second-order density function satisfies

$$f_X(x_1, x_2; t_1, t_2) = f_X(x_1, x_2; t_1 + \Delta, t_2 + \Delta) \tag{6.2-13}$$

for all t_1, t_2, and Δ. After some thought, the reader will conclude that (6.2-13) is a function of time differences $t_2 - t_1$ and not absolute time (let arbitrary $\Delta = -t_1$). A second-order stationary process is also first-order stationary because the second-order density function determines the lower, first-order, density.

Now the correlation $E[X_1X_2] = E[X(t_1)X(t_2)]$ of a random process will, in general, be a function of t_1 and t_2. Let us denote this function by $R_{XX}(t_1, t_2)$ and call it the *autocorrelation function* of the random process $X(t)$:

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)]. \tag{6.2-14}$$

A consequence of (6.2-13), however, is that the autocorrelation function of a second-order stationary process is a function only of time differences and not absolute time; that is, if

$$\tau = t_2 - t_1 \tag{6.2-15}$$

then (6.2-14) becomes

$$R_{XX}(t_1, t_1 + \tau) = E[X(t_1)X(t_1 + \tau)] = R_{XX}(\tau) \tag{6.2-16}$$

Proof of (6.2-16) uses (6.2-13); it is left as a reader exercise (see Problem 6-6).

Many practical problems require that we deal with the autocorrelation function and mean value of a random process. Problem solutions are greatly

† Note that the variable x_2 of integration has been replaced by the alternative variable x_1 for convenience.

simplified if these quantities are not dependent on absolute time. Of course, second-order stationarity is sufficient to guarantee these characteristics. However, it is often more restrictive than necessary, and a more relaxed form of stationarity is desirable. The most useful form is the *wide-sense stationary process*, defined as that for which two conditions are true:

$$E[X(t)] = \bar{X} = \text{constant} \tag{6.2-17a}$$

$$E[X(t)X(t+\tau)] = R_{XX}(\tau) \tag{6.2-17b}$$

A process stationary to order 2 is clearly wide-sense stationary. However, the converse is not necessarily true.

Example 6.2-1 We show that the random process

$$X(t) = A\cos(\omega_0 t + \Theta)$$

is wide-sense stationary if it is assumed that A and ω_0 are constants and Θ is a uniformly distributed random variable on the interval $(0, 2\pi)$. The mean value is

$$E[X(t)] = \int_0^{2\pi} A\cos(\omega_0 t + \theta)\frac{1}{2\pi}\,d\theta = 0$$

The autocorrelation function, from (6.2-14) with $t_1 = t$ and $t_2 = t + \tau$, becomes

$$\begin{aligned} R_{XX}(t, t+\tau) &= E[A\cos(\omega_0 t + \Theta)A\cos(\omega_0 t + \omega_0\tau + \Theta)] \\ &= \frac{A^2}{2}E[\cos(\omega_0\tau) + \cos(2\omega_0 t + \omega_0\tau + 2\Theta)] \\ &= \frac{A^2}{2}\cos(\omega_0\tau) + \frac{A^2}{2}E[\cos(2\omega_0 t + \omega_0\tau + 2\Theta)] \end{aligned}$$

The second term easily evaluates to 0. Thus, the autocorrelation function depends only on τ and the mean value is a constant, so $X(t)$ is wide-sense stationary.

When we are concerned with two random processes $X(t)$ and $Y(t)$, we say they are *jointly* wide-sense stationary if each satisfies (6.2-17) and their *cross-correlation function*, defined in general by

$$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)] \tag{6.2-18}$$

is a function only of time difference $\tau = t_2 - t_1$ and not absolute time; that is, if

$$R_{XY}(t, t+\tau) = E[X(t)Y(t+\tau)] = R_{XY}(\tau) \tag{6.2-19}$$

N-Order and Strict-Sense Stationarity

By extending the above reasoning to N random variables $X_i = X(t_i)$, $i = 1, 2, \ldots, N$, we say a random process is *stationary to order N* if its Nth-order density function is invariant to a time origin shift; that is, if

$$f_X(x_1, \ldots, x_N; t_1, \ldots, t_N) = f_X(x_1, \ldots, x_N; t_1 + \Delta, \ldots, t_N + \Delta) \tag{6.2-20}$$

for all $t_1, \ldots, t_N$ and Δ. Stationarity of order N implies stationarity to all orders $k \le N$. A process stationary to *all* orders $N = 1, 2, \ldots$, is called *strict-sense stationary*.

Time Averages and Ergodicity

The time average of a quantity is defined as

$$A[\cdot] = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}[\cdot]\,dt \tag{6.2-21}$$

Here A is used to denote time average in a manner analogous to E for the statistical average. Time average is taken over all time because, as applied to random processes, sample functions of processes are presumed to exist for all time.

Specific averages of interest are the mean value $\bar{x} = A[x(t)]$ of a sample function (a lower case letter is used to imply a sample function), and the *time autocorrelation function*, denoted $\mathcal{R}_{xx}(\tau) = A[x(t)x(t+\tau)]$. These functions are defined by

$$\bar{x} = A[x(t)] = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)\,dt \tag{6.2-22}$$

$$\begin{aligned} \mathcal{R}_{xx}(\tau) &= A[x(t)x(t+\tau)] \\ &= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)x(t+\tau)\,dt \end{aligned} \tag{6.2-23}$$

For any *one* sample function of the process $X(t)$, these last two integrals simply produce two numbers (for a fixed value of τ). However, when all sample functions are considered, we see that $\bar{x}$ and $\mathcal{R}_{xx}(\tau)$ are actually *random variables*. By taking the expected value on both sides of (6.2-22) and (6.2-23), and assuming the expectation can be brought inside the integrals, we obtain†

$$E[\bar{x}] = \bar{X} \tag{6.2-24}$$

$$E[\mathcal{R}_{xx}(\tau)] = R_{XX}(\tau) \tag{6.2-25}$$

Now suppose by some theorem the random variables $\bar{x}$ and $\mathcal{R}_{xx}(\tau)$ could be made to have zero variances; that is, $\bar{x}$ and $\mathcal{R}_{xx}(\tau)$ actually become constants.

† We assume also that $X(t)$ is a stationary process so that the mean and the autocorrelation function are not time-dependent.

Then we could write

$$\bar{x} = \bar{X} \tag{6.2-26}$$

$$\mathcal{R}_{xx}(\tau) = R_{XX}(\tau) \tag{6.2-27}$$

In other words, the time averages $\bar{x}$ and $\mathcal{R}_{xx}(\tau)$ equal the statistical averages $\bar{X}$ and $R_{XX}(\tau)$ respectively. The *ergodic theorem* allows the validity of (6.2-26) and (6.2-27). Stated in loose terms, it more generally allows all time averages to equal the corresponding statistical averages. Processes that satisfy the ergodic theorem are called *ergodic processes.*

Ergodicity is a very restrictive form of stationarity and it may be difficult to prove that it constitutes a reasonable assumption in any physical situation. Nevertheless, we shall often assume a process is ergodic to simplify problems. In the real world, we are usually forced to work with only one sample function of a process and therefore must, like it or not, derive mean value, correlation functions, etc. from the time waveform. By assuming ergodicity, we may infer the similar statistical characteristics of the process. The reader may feel that our theory is on shaky ground based on these comments. However, it must be remembered that all our theory only serves to model real-world conditions. Therefore, what difference do our assumptions really make provided the assumed model does truly reflect real conditions?

Two random processes are called *jointly ergodic* if they are individually ergodic and also have a *time cross-correlation function* that equals the statistical cross-correlation function:†

$$\mathcal{R}_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(t)y(t+\tau)\,dt = R_{XY}(\tau) \tag{6.2-28}$$

6.3 CORRELATION FUNCTIONS

The autocorrelation and cross-correlation functions were introduced in the previous section. These functions are examined further in this section, along with their properties. In addition, other correlation-type functions are introduced that are important to the study of random processes.

Autocorrelation Function and Its Properties

Recall that the autocorrelation function of a random process $X(t)$ is the correlation $E[X_1X_2]$ of two random variables $X_1 = X(t_1)$ and $X_2 = X(t_2)$ defined by the process at times t_1 and t_2. Mathematically,

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)] \tag{6.3-1}$$

† As in ordinary stationarity, there are various *orders* of ergodic stationarity. For more detail on ergodic processes, the reader is referred to Papoulis (1965), pp. 323–332.

For time assignments, $t_1 = t$ and $t_2 = t_1 + \tau$, with τ a real number, (6.3-1) assumes the convenient form

$$R_{XX}(t, t+\tau) = E[X(t)X(t+\tau)] \tag{6.3-2}$$

If $X(t)$ is at least wide-sense stationary, it was noted in Section 6.2 that $R_{XX}(t, t+\tau)$ must be a function only of time difference $\tau = t_2 - t_1$. Thus, for wide-sense stationary processes

$$R_{XX}(\tau) = E[X(t)X(t+\tau)] \tag{6.3-3}$$

For such processes the autocorrelation function exhibits the following properties:

$$(1)\quad |R_{XX}(\tau)| \le R_{XX}(0) \tag{6.3-4}$$

$$(2)\quad R_{XX}(-\tau) = R_{XX}(\tau) \tag{6.3-5}$$

$$(3)\quad R_{XX}(0) = E[X^2(t)] \tag{6.3-6}$$

The first property shows that $R_{XX}(\tau)$ is bounded by its value at the origin, while the third property states that this bound is equal to the mean-squared value called the *power* in the process. The second property indicates that an autocorrelation function has even symmetry.

Other properties of stationary processes may also be stated [see Cooper and McGillem (1971), p. 113, and Melsa and Sage (1973), pp. 207–208]:

(4) If $E[X(t)] = \bar{X} \ne 0$ and $X(t)$ has no periodic components then

$$\lim_{|\tau|\to\infty} R_{XX}(\tau) = \bar{X}^2 \tag{6.3-7}$$

(5) If $X(t)$ has a periodic component, then $R_{XX}(\tau)$ will have a periodic component with the same period. (6.3-8)

(6) If $X(t)$ is ergodic, zero-mean, and has no periodic component, then

$$\lim_{|\tau|\to\infty} R_{XX}(\tau) = 0 \tag{6.3-9}$$

(7) $R_{XX}(\tau)$ cannot have an arbitrary shape. (6.3-10)

Properties 4 through 6 are more or less self-explanatory. Property 7 simply says that any arbitrary function cannot be an autocorrelation function. This fact will be more apparent when the *power density spectrum* is introduced in Chapter 7. It will be shown there that $R_{XX}(\tau)$ is related to the power density spectrum through the Fourier transform and the form of the spectrum is not arbitrary.

Example 6.3-1 Given the autocorrelation function for a stationary process is

$$R_{XX}(\tau) = 25 + \frac{4}{1 + 6\tau^2}$$

we shall find the mean value and variance of the process $X(t)$. From property

4, the mean value is $E[X(t)] = \bar{X} = \sqrt{25} = \pm 5$. The variance is given by (3.2-6), so

$$\sigma_X^2 = E[X^2(t)] - (E[X(t)])^2$$

But $E[X^2(t)] = R_{XX}(0) = 25 + 4 = 29$ from property 3, so

$$\sigma_X^2 = 29 - 25 = 4$$

Cross-Correlation Function and Its Properties

The cross-correlation function of two random processes $X(t)$ and $Y(t)$ was defined in (6.2-18). Setting $t_1 = t$ and $\tau = t_2 - t_1$, we may write (6.2-18) as

$$R_{XY}(t, t + \tau) = E[X(t)Y(t + \tau)] \tag{6.3-11}$$

If $X(t)$ and $Y(t)$ are at least jointly wide-sense stationary, $R_{XY}(t, t + \tau)$ is independent of absolute time and we can write

$$R_{XY}(\tau) = E[X(t)Y(t + \tau)] \tag{6.3-12}$$

If

$$R_{XY}(t, t + \tau) = 0 \tag{6.3-13}$$

then $X(t)$ and $Y(t)$ are called *orthogonal processes*. If the two processes are statistically independent, the cross-correlation function becomes

$$R_{XY}(t, t + \tau) = E[X(t)]E[Y(t + \tau)] \tag{6.3-14}$$

If, in addition to being independent, $X(t)$ and $Y(t)$ are at least wide-sense stationary, (6.3-14) becomes

$$R_{XY}(\tau) = \bar{X}\bar{Y} \tag{6.3-15}$$

which is a constant.

We may list some properties of the cross-correlation function applicable to processes that are at least wide-sense stationary:

$$(1) \quad R_{XY}(-\tau) = R_{YX}(\tau) \tag{6.3-16}$$

$$(2) \quad |R_{XY}(\tau)| \le \sqrt{R_{XX}(0)R_{YY}(0)} \tag{6.3-17}$$

$$(3) \quad |R_{XY}(\tau)| \le \tfrac{1}{2}[R_{XX}(0) + R_{YY}(0)] \tag{6.3-18}$$

Property 1 follows from the definition (6.3-12). It describes the symmetry of $R_{XY}(\tau)$. Property 2 can be proven by expanding the inequality

$$E[\{Y(t + \tau) + \alpha X(t)\}^2] \ge 0 \tag{6.3-19}$$

where α is a real number (see Problem 6-27). Properties 2 and 3 both constitute bounds on the magnitude of $R_{XY}(\tau)$. Equation (6.3-17) represents a tighter bound than that of (6.3-18), because the geometric mean of two positive numbers cannot exceed their arithmetic mean; that is

$$\sqrt{R_{XX}(0)R_{YY}(0)} \le \tfrac{1}{2}[R_{XX}(0) + R_{YY}(0)] \tag{6.3-20}$$

Example 6.3-2 Let two random processes $X(t)$ and $Y(t)$ be defined by

$$X(t) = A\cos(\omega_0 t) + B\sin(\omega_0 t)$$

$$Y(t) = B\cos(\omega_0 t) - A\sin(\omega_0 t)$$

where A and B are random variables and ω_0 is a constant. It can be shown (Problem 6-12) that $X(t)$ is wide-sense stationary if A and B are uncorrelated, zero-mean random variables with the same variance (they may have different density functions, however). With these same constraints on A and B, $Y(t)$ is also wide-sense stationary. We shall now find the cross-correlation function $R_{XY}(t, t + \tau)$ and show that $X(t)$ and $Y(t)$ are *jointly* wide-sense stationary. By use of (6.3-11) we have

$$\begin{aligned}
R_{XY}(t, t + \tau) &= E[X(t)Y(t + \tau)] \\
&= E[AB\cos(\omega_0 t)\cos(\omega_0 t + \omega_0 \tau) \\
&\quad + B^2 \sin(\omega_0 t)\cos(\omega_0 t + \omega_0 \tau) \\
&\quad - A^2 \cos(\omega_0 t)\sin(\omega_0 t + \omega_0 \tau) \\
&\quad - AB\sin(\omega_0 t)\sin(\omega_0 t + \omega_0 \tau)] \\
&= E[AB]\cos(2\omega_0 t + \omega_0 \tau) \\
&\quad + E[B^2]\sin(\omega_0 t)\cos(\omega_0 t + \omega_0 \tau) \\
&\quad - E[A^2]\cos(\omega_0 t)\sin(\omega_0 t + \omega_0 \tau)
\end{aligned}$$

Since A and B are assumed to be zero-mean, uncorrelated random variables, $E[AB] = 0$. Also, since A and B are assumed to have equal variances, $E[A^2] = E[B^2] = \sigma^2$ and we obtain

$$R_{XY}(t, t + \tau) = -\sigma^2 \sin(\omega_0 \tau)$$

Thus, $X(t)$ and $Y(t)$ are jointly wide-sense stationary because $R_{XY}(t, t + \tau)$ depends only on τ.

Note from the above result that cross-correlation functions are not necessarily even functions of τ with the maximum at $\tau = 0$, as is the case with autocorrelation functions.

Covariance Functions

The concept of the covariance of two random variables, as defined by (5.1-13), can be extended to random processes. The *autocovariance function* is defined by

$$C_{XX}(t, t + \tau) = E[\{X(t) - E[X(t)]\}\{X(t + \tau) - E[X(t + \tau)]\}] \tag{6.3-21}$$

which can also be put in the form

$$C_{XX}(t, t+\tau) = R_{XX}(t, t+\tau) - E[X(t)]E[X(t+\tau)] \tag{6.3-22}$$

The *cross-covariance function* for two processes $X(t)$ and $Y(t)$ is defined by

$$C_{XY}(t, t+\tau) = E[\{X(t) - E[X(t)]\}\{Y(t+\tau) - E[Y(t+\tau)]\}] \tag{6.3-23}$$

or, alternatively,

$$C_{XY}(t, t+\tau) = R_{XY}(t, t+\tau) - E[X(t)]E[Y(t+\tau)] \tag{6.3-24}$$

For processes that are at least jointly wide-sense stationary, (6.3-22) and (6.3-24) reduce to

$$C_{XX}(\tau) = R_{XX}(\tau) - \bar{X}^2 \tag{6.3-25}$$

and

$$C_{XY}(\tau) = R_{XY}(\tau) - \bar{X}\bar{Y} \tag{6.3-26}$$

The *variance* of a random process is given in general by (6.3-21) with $\tau = 0$. For a wide-sense stationary process, variance does not depend on time and is given by (6.3-25) with $\tau = 0$:

$$\sigma_X^2 = E[\{X(t) - E[X(t)]\}^2] = R_{XX}(0) - \bar{X}^2 \tag{6.3-27}$$

For two random processes, if

$$C_{XY}(t, t+\tau) = 0 \tag{6.3-28}$$

they are called *uncorrelated*. From (6.3-24) this means that

$$R_{XY}(t, t+\tau) = E[X(t)]E[Y(t+\tau)] \tag{6.3-29}$$

Since this result is the same as (6.3-14), which applies to independent processes, we conclude that independent processes are uncorrelated. The converse case is not necessarily true, although it is true for *jointly gaussian processes*, which we consider in Section 6.5.

6.4 MEASUREMENT OF CORRELATION FUNCTIONS

In the real world, we can never measure the true correlation functions of two random processes $X(t)$ and $Y(t)$ because we never have *all* sample functions of the ensemble at our disposal. Indeed, we may typically have available for measurements only a portion of one sample function from each process. Thus, our only recourse is to determine time averages based on finite time portions of single sample functions, taken large enough to approximate true results for ergodic processes. Because we are able to work only with time functions, we are forced, like it or not, to presume that given processes are ergodic. This fact should not prove too disconcerting, however, if we remember that assumptions only reflect the details of our mathematical model of a real-world situation. Provided that the model gives consistent agreement with the real situation, it is of little importance whether ergodicity is assumed or not.

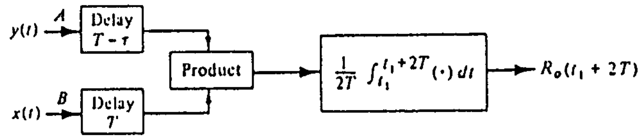


Figure 6.4-1 A time cross-correlation function measurement system. Autocorrelation function measurement is possible by connecting points A and B and applying either $x(t)$ or $y(t)$.

Figure 6.4-1 illustrates the block diagram of a possible system for measuring the approximate time cross-correlation function of two jointly ergodic random processes $X(t)$ and $Y(t)$. Sample functions $x(t)$ and $y(t)$ are delayed by amounts T and $T - \tau$, respectively, and the product of the delayed waveforms is formed. This product is then integrated to form the output which equals the integral at time $t_1 + 2T$, where t_1 is arbitrary and $2T$ is the integration period. The integrator can be of the integrate-and-dump variety described by Peebles (1976, p. 361).

If we assume $x(t)$ and $y(t)$ exist at least during the interval $-T < t$ and t_1 is an arbitrary time except $0 \le t_1$, then the output is easily found to be

$$R_o(t_1 + 2T) = \frac{1}{2T}\int_{t_1 - T}^{t_1 + T} x(t)y(t+\tau)\,dt \tag{6.4-1}$$

Now if we choose $t_1 = 0$† and assume T is large, then we have

$$R_o(2T) = \frac{1}{2T}\int_{-T}^{T} x(t)y(t+\tau)\,dt \approx \mathcal{R}_{xy}(\tau) = R_{XY}(\tau) \tag{6.4-2}$$

Thus, for jointly ergodic processes, the system of Figure 6.4-1 can approximately measure their cross-correlation function (τ is varied to obtain the complete function).

Clearly, by connecting points A and B and applying either $x(t)$ or $y(t)$ to the system, we can also measure the autocorrelation functions $R_{XX}(\tau)$ and $R_{YY}(\tau)$.

Example 6.4-1 We connect points A and B together in Figure 6.4-1 and use the system to measure the autocorrelation function of the process $X(t)$ of Example 6.2-1. From (6.4-2)

$$R_o(2T) = \frac{1}{2T}\int_{-T}^{T} A^2 \cos(\omega_0 t + \theta)\cos(\omega_0 t + \theta + \omega_0\tau)\,dt$$

$$= \frac{A^2}{4T}\int_{-T}^{T} [\cos(\omega_0\tau) + \cos(2\omega_0 t + 2\theta + \omega_0\tau)]\,dt$$

In writing this result θ represents a specific value of the random variable Θ;

† Since the processes are assumed jointly ergodic and therefore jointly stationary, the integral (6.4-1) will tend to be independent of t_1 if T is large enough.

the value that corresponds to the specific ensemble member being used in (6.4-2). On straightforward reduction of the above integral we obtain

$$R_o(2T) = R_{XX}(\tau) + \varepsilon(T)$$

where

$$R_{XX}(\tau) = (A^2/2)\cos(\omega_0 \tau)$$

is the true autocorrelation function of $X(t)$, and

$$\varepsilon(T) = (A^2/2)\cos(\omega_0 \tau + 2\theta)\,\frac{\sin(2\omega_0 T)}{2\omega_0 T}$$

is an error term. If we require the error term's magnitude to be at least 20 times smaller than the largest value of the true autocorrelation function then $|\varepsilon(T)| < 0.05R_{XX}(0)$ is necessary. Thus, we must have $1/2\omega_0 T \leq 0.05$ or

$$T \geq 10/\omega_0$$

In other words, if $T \geq 10/\omega_0$ the error in using Figure 6.4-1 to measure the autocorrelation function of the process $X(t) = A\cos(\omega_0 t + \Theta)$ will be 5% or less of the largest value of the true autocorrelation function.

6.5 GAUSSIAN RANDOM PROCESSES

A number of random processes are important enough to have been given names. We shall discuss only the most important of these, the *gaussian random process*.

Consider a continuous random process such as illustrated in Figure 6.1-1 and define N random variables $X_1 = X(t_1), \ldots, X_i = X(t_i), \ldots, X_N = X(t_N)$ corresponding to N time instants $t_1, \ldots, t_i, \ldots, t_N$. If, for any $N = 1, 2, \ldots$ and any times $t_1, \ldots, t_N$, these random variables are jointly gaussian, that is, they have a joint density as given by (5.3-12), the process is called gaussian. Equation (5.3-12) can be written in the form

$$f_X(x_1, \ldots, x_N; t_1, \ldots, t_N) = \frac{\exp\{-(1/2)[x - \bar{X}]^t[C_X]^{-1}[x - \bar{X}]\}}{\sqrt{(2\pi)^N |[C_X]|}} \tag{6.5-1}$$

where matrices $[x - \bar{X}]$ and $[C_X]$ are defined in (5.3-13) and (5.3-14) and (5.3-15), respectively. The mean values $\bar{X}_i$ of $X(t_i)$ are

$$\bar{X}_i = E[X_i] = E[X(t_i)] \tag{6.5-2}$$

The elements of the covariance matrix $[C_X]$ are

$$\begin{aligned} C_{ik} = C_{X_iX_k} &= E[(X_i - \bar{X}_i)(X_k - \bar{X}_k)] \\ &= E[\{X(t_i) - E[X(t_i)]\}\{X(t_k) - E[X(t_k)]\}] \\ &= C_{XX}(t_i, t_k) \end{aligned} \tag{6.5-3}$$

which is the autocovariance of $X(t_i)$ and $X(t_k)$ from (6.3-21).

From (6.5-2) and (6.5-3), when used in (6.5-1), we see that the mean and autocovariance functions are all that are needed to completely specify a gaussian random process. By expanding (6.5-3) to get

$$C_{XX}(t_i, t_k) = R_{XX}(t_i, t_k) - E[X(t_i)]E[X(t_k)] \tag{6.5-4}$$

we see that an alternative specification using only the mean and autocorrelation function $R_{XX}(t_i, t_k)$ is possible.

If the gaussian process is not stationary the mean and autocovariance functions will, in general, depend on absolute time. However, for the important case where the process is wide-sense stationary, the mean will be constant,

$$\bar{X}_i = E[X(t_i)] = \bar{X} \qquad \text{(constant)} \tag{6.5-5}$$

while the autocovariance and autocorrelation functions will depend only on time differences and not absolute time,

$$C_{XX}(t_i, t_k) = C_{XX}(t_k - t_i) \tag{6.5-6}$$

$$R_{XX}(t_i, t_k) = R_{XX}(t_k - t_i) \tag{6.5-7}$$

It follows from the preceding discussions that a wide-sense stationary gaussian process is also strictly stationary.

We illustrate some of the above remarks with an example.

Example 6.5-1 A gaussian random process is known to be wide-sense stationary with a mean of $\bar{X} = 4$ and autocorrelation function

$$R_{XX}(\tau) = 25e^{-3|\tau|}$$

We seek to specify the joint density function for three random variables $X(t_i)$, $i = 1, 2, 3$, defined at times $t_i = t_0 + [(i - 1)/2]$, with t_0 a constant.

Here $t_k - t_i = (k - i)/2$, i and $k = 1, 2, 3$, so

$$R_{XX}(t_k - t_i) = 25e^{-3|k-i|/2}$$

and

$$C_{XX}(t_k - t_i) = 25e^{-3|k-i|/2} - 16$$

from (6.5-4) through (6.5-7). Elements of the covariance matrix are found from (6.5-3). Thus,

$$[C_X] = \begin{bmatrix} (25 - 16) & (25e^{-3/2} - 16) & (25e^{-6/2} - 16) \\ (25e^{-3/2} - 16) & (25 - 16) & (25e^{-3/2} - 16) \\ (25e^{-6/2} - 16) & (25e^{-3/2} - 16) & (25 - 16) \end{bmatrix}$$

and $\bar{X}_i = 4$ completely determine (6.5-1) for this case where $N = 3$.

Two random processes $X(t)$ and $Y(t)$ are said to be *jointly gaussian* if the random variables $X(t_1), \ldots, X(t_N), Y(t'_1), \ldots, Y(t'_M)$ defined at times $t_1, \ldots, t_N$ for $X(t)$ and times $t'_1, \ldots, t'_M$ for $Y(t)$, are jointly gaussian for any $N, t_1, \ldots, t_N, M, t'_1, \ldots, t'_M$.

*6.6 COMPLEX RANDOM PROCESSES

If the complex random variable of Section 5.6 is generalized to include time, the result is a *complex random process* $Z(t)$ *given by*

$$Z(t) = X(t) + jY(t) \tag{6.6-1}$$

where $X(t)$ and $Y(t)$ are real processes. $Z(t)$ is called stationary if $X(t)$ and $Y(t)$ are jointly stationary. If $X(t)$ and $Y(t)$ are jointly wide-sense stationary, then $Z(t)$ is said to be wide-sense stationary.

Two complex processes $Z_i(t)$ and $Z_j(t)$ are jointly wide-sense stationary if each is wide-sense stationary and their cross-correlation function (defined below) is a function of time differences only and not absolute time.

We may extend the operations involving process mean value, autocorrelation function, and autocovariance function to include complex processes. The *mean value* of $Z(t)$ is

$$E[Z(t)] = E[X(t)] + jE[Y(t)] \tag{6.6-2}$$

Autocorrelation function is defined by

$$R_{ZZ}(t, t+\tau) = E[Z^*(t)Z(t+\tau)] \tag{6.6-3}$$

where the asterisk * denotes the complex conjugate. *Autocovariance function* is defined by

$$C_{ZZ}(t, t+\tau) = E[\{Z(t) - E[Z(t)]\}^*\{Z(t+\tau) - E[Z(t+\tau)]\}] \tag{6.6-4}$$

If $Z(t)$ is at least wide-sense stationary, the mean value becomes a constant

$$\bar{Z} = \bar{X} + j\bar{Y} \tag{6.6-5}$$

and the correlation functions are independent of absolute time:

$$R_{ZZ}(t, t+\tau) = R_{ZZ}(\tau) \tag{6.6-6}$$

$$C_{ZZ}(t, t+\tau) = C_{ZZ}(\tau) \tag{6.6-7}$$

For two complex processes $Z_i(t)$ and $Z_j(t)$, *cross-correlation* and *cross-covariance functions* are defined by

$$R_{Z_iZ_j}(t, t+\tau) = E[Z_i^*(t)Z_j(t+\tau)] \qquad i \neq j \tag{6.6-8}$$

and

$$C_{Z_iZ_j}(t, t+\tau) = E[\{Z_i(t) - E[Z_i(t)]\}^*\{Z_j(t+\tau) - E[Z_j(t+\tau)]\}] \qquad i \neq j \tag{6.6-9}$$

respectively. If the two processes are at least jointly wide-sense stationary, we obtain

$$R_{Z_iZ_j}(t, t+\tau) = R_{Z_iZ_j}(\tau) \qquad i \neq j \tag{6.6-10}$$

$$C_{Z_iZ_j}(t, t+\tau) = C_{Z_iZ_j}(\tau) \qquad i \neq j \tag{6.6-11}$$

$Z_i(t)$ and $Z_j(t)$ are said to be *uncorrelated processes* if $C_{Z_iZ_j}(t, t+\tau) = 0$, $i \neq j$. They are called *orthogonal processes* if $R_{Z_iZ_j}(t, t+\tau) = 0$, $i \neq j$.

Example 6.6-1 A complex random process $V(t)$ is comprised of a sum of N complex signals:

$$V(t) = \sum_{n=1}^{N} A_n e^{j\omega_0 t + j\Theta_n}$$

Here $\omega_0/2\pi$ is the (constant) frequency of each signal. A_n is a random variable representing the random amplitude of the nth signal. Similarly, Θ_n is a random variable representing a random phase angle. We assume all the variables A_n and Θ_n, for $n = 1, 2, \ldots, N$, are statistically independent and the Θ_n are uniformly distributed on $(0, 2\pi)$. We find the autocorrelation function of $V(t)$.

From (6.6-3):

$$\begin{aligned} R_{VV}(t, t+\tau) &= E[V^*(t)V(t+\tau)] \\ &= E\left[\sum_{n=1}^{N} A_n e^{-j\omega_0 t - j\Theta_n} \sum_{m=1}^{N} A_m e^{j\omega_0 t + j\omega_0\tau + j\Theta_m}\right] \\ &= \sum_{n=1}^{N}\sum_{m=1}^{N} e^{j\omega_0\tau} E[A_n A_m e^{j(\Theta_m - \Theta_n)}] = R_{VV}(\tau) \end{aligned}$$

From statistical independence:

$$R_{VV}(\tau) = e^{j\omega_0\tau} \sum_{n=1}^{N}\sum_{m=1}^{N} E[A_n A_m] E[\exp\{j(\Theta_m - \Theta_n)\}]$$

However,

$$\begin{aligned} E[\exp\{j(\Theta_m - \Theta_n)\}] &= E[\cos(\Theta_m - \Theta_n)] + jE[\sin(\Theta_m - \Theta_n)] \\ &= \int_0^{2\pi}\int_0^{2\pi} \frac{1}{(2\pi)^2}[\cos(\theta_m - \theta_n) + j\sin(\theta_m - \theta_n)]\, d\theta_n\, d\theta_m \\ &= \begin{cases} 0 & m \neq n \\ 1 & m = n \end{cases} \end{aligned}$$

so

$$R_{VV}(\tau) = e^{j\omega_0\tau} \sum_{n=1}^{N} \overline{A_n^2}$$

PROBLEMS

6-1 A random experiment consists of selecting a point on some city street that has two-way automobile traffic. Define and classify a random process for this experiment that is related to traffic flow.

6-2 A 10-meter section of a busy downtown sidewalk is actually the platform of a scale that produces a voltage proportional to the total weight of people on the scale at any time.

(a) Sketch a typical sample function for this process.
(b) What is the underlying random experiment for the process?
(c) Classify the process.

*6-3 An experiment consists of measuring the weight W of some person each 10 minutes. The person is randomly male or female (which is not known though) with equal probability. A two-level discrete random process $X(t)$ is generated where

$$X(t) = \pm 10$$

The level -10 is generated in the period following a measurement if the measured weight does not exceed W_0 (some constant). Level $+10$ is generated if weight exceeds W_0. Let the weight of men in kg be a random variable having the gaussian density

$$f_W(w\,|\,\text{male}) = \frac{1}{\sqrt{2\pi}11.3} \exp\left[-(w-77.1)^2/2(11.3)^2\right]$$

Similarly, for women

$$f_W(w\,|\,\text{female}) = \frac{1}{\sqrt{2\pi}6.8} \exp\left[-(w-54.4)^2/2(6.8)^2\right]$$

(a) Find W_0 so that $P\{W > W_0\,|\,\text{male}\}$ is equal to $P\{W \le W_0\,|\,\text{female}\}$.
(b) If the levels ± 10 are interpreted as "decisions" about whether the weight measurement of a person corresponds to a male or female, give a physical significance to their generation.
(c) Sketch a possible sample function.

6-4 The two-level *semirandom binary process* is defined by

$$X(t) = A \text{ or } -A \qquad (n-1)T < t < nT$$

where the levels A and $-A$ occur with equal probability, T is a positive constant, and $n = 0, \pm 1, \pm 2, \ldots$.

(a) Sketch a typical sample function.
(b) Classify the process.
(c) Is the process deterministic?

6-5 Sample functions in a discrete random process are constants; that is

$$X(t) = C = \text{constant}$$

where C is a discrete random variable having possible values $c_1 = 1$, $c_2 = 2$, and $c_3 = 3$ occurring with probabilities 0.6, 0.3, and 0.1 respectively.

(a) Is $X(t)$ deterministic?
(b) Find the first-order density function of $X(t)$ at any time t.

6-6 Utilize (6.2-13) to prove (6.2-16).

*6-7 A random process $X(t)$ has periodic sample functions as shown in Figure P6-7 where B, T, and $4t_0 \le T$ are constants but ε is a random variable uniformly distributed on the interval $(0, T)$.

(a) Find the first-order distribution function of $X(t)$.
(b) Find the first-order density function.
(c) Find $E[X(t)]$, $E[X^2(t)]$, and σ_X^2.

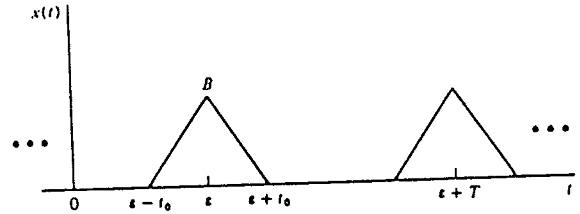


Figure P6-7

*6-8 Work Problem 6-7 for the waveform of Figure P6-8. Assume $2t_0 < T$.

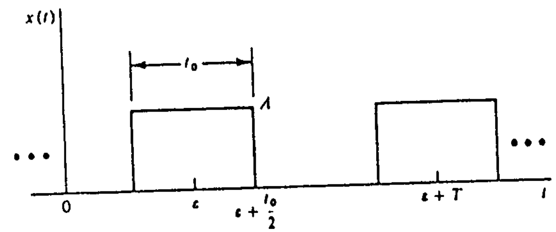


Figure P6-8

*6-9 Work Problem 6-7 for the waveform of Figure P6-9. Assume $4t_0 \le T$.

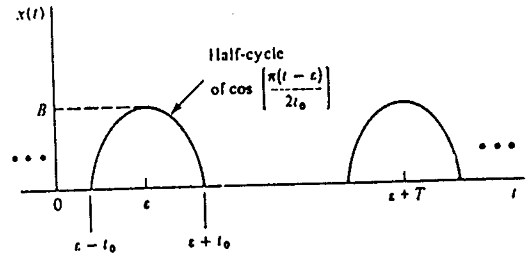


Figure P6-9

6-10 Given the random process

$$X(t) = A \sin(\omega_0 t + \Theta)$$

where A and ω_0 are constants and Θ is a random variable uniformly distributed on the interval $(-\pi, \pi)$. Define a new random process $Y(t) = X^2(t)$.

(*a*) Find the autocorrelation function of $Y(t)$.
(*b*) Find the cross-correlation function of $X(t)$ and $Y(t)$.
(*c*) Are $X(t)$ and $Y(t)$ wide-sense stationary?
(*d*) Are $X(t)$ and $Y(t)$ jointly wide-sense stationary?

6-11 A random process is defined by

$$Y(t) = X(t) \cos(\omega_0 t + \Theta)$$

where $X(t)$ is a wide-sense stationary random process that amplitude-modulates a carrier of constant angular frequency ω_0 with a random phase Θ independent of $X(t)$ and uniformly distributed on $(-\pi, \pi)$.

(*a*) Find $E[Y(t)]$.
(*b*) Find the autocorrelation function of $Y(t)$.
(*c*) Is $Y(t)$ wide-sense stationary?

6-12 Given the random process

$$X(t) = A \cos(\omega_0 t) + B \sin(\omega_0 t)$$

where ω_0 is a constant, and A and B are uncorrelated zero-mean random variables having different density functions but the same variances σ^2. Show that $X(t)$ is wide-sense stationary but not strictly stationary.

6-13 If $X(t)$ is a stationary random process having a mean value $E[X(t)] = 3$ and autocorrelation function $R_{XX}(\tau) = 9 + 2e^{-|\tau|}$, find:

(*a*) the mean value and
(*b*) the variance of the random variable

$$Y = \int_0^2 X(t)\,dt$$

(*Hint:* Assume expectation and integration operations are interchangeable.)

6-14 Define a random process by

$$X(t) = A \cos(\pi t)$$

where A is a gaussian random variable with zero mean and variance σ_A^2.

(*a*) Find the density functions of $X(0)$ and $X(1)$.
(*b*) Is $X(t)$ stationary in any sense?

6-15 For the random process of Problem 6-4, calculate:

(*a*) the mean value $E[X(t)]$ (*b*) $R_{XX}(t_1 = 0.5T,\ t_2 = 0.7T)$ (*c*) $R_{XX}(t_1 = 0.2T,\ t_2 = 1.2T)$.

6-16 A random process consists of three sample functions $X(t, s_1) = 2$, $X(t, s_2) = 2\cos(t)$, and $X(t, s_3) = 3\sin(t)$, each occurring with equal probability. Is the process stationary in any sense?

6-17 Statistically independent, zero-mean, random processes $X(t)$ and $Y(t)$ have autocorrelation functions

$$R_{XX}(\tau) = e^{-|\tau|}$$

and

$$R_{YY}(\tau) = \cos(2\pi\tau)$$

respectively.

(*a*) Find the autocorrelation function of the sum $W_1(t) = X(t) + Y(t)$.
(*b*) Find the autocorrelation function of the difference $W_2(t) = X(t) - Y(t)$.
(*c*) Find the cross-correlation function of $W_1(t)$ and $W_2(t)$.

6-18 Define a random process as $X(t) = p(t + \varepsilon)$, where $p(t)$ is any periodic waveform with period T and ε is a random variable uniformly distributed on the interval $(0, T)$. Show that

$$E[X(t)X(t+\tau)] = \frac{1}{T}\int_0^T p(\xi)p(\xi+\tau)\,d\xi = R_{XX}(\tau)$$

***6-19** Use the result of Problem 6-18 to find the autocorrelation function of random processes having periodic sample function waveforms $p(t)$ defined

(*a*) by Figure P6-7 with $\varepsilon = 0$ and $4t_0 \le T$, and
(*b*) by Figure P6-8 with $\varepsilon = 0$ and $2t_0 \le T$.

6-20 Define two random processes by $X(t) = p_1(t + \varepsilon)$ and $Y(t) = p_2(t + \varepsilon)$ when $p_1(t)$ and $p_2(t)$ are both periodic waveforms with period T and ε is a random variable uniformly distributed on the interval $(0, T)$. Find an expression for the cross-correlation function $E[X(t)Y(t+\tau)]$.

6-21 Prove:

(*a*) (6.3-4) and (*b*) (6.3-5).

6-22 Give arguments to justify (6.3-9).

6-23 For the random process having the autocorrelation function shown in Figure P6-23, find:

(*a*) $E[X(t)]$ (*b*) $E[X^2(t)]$ and (*c*) σ_X^2.

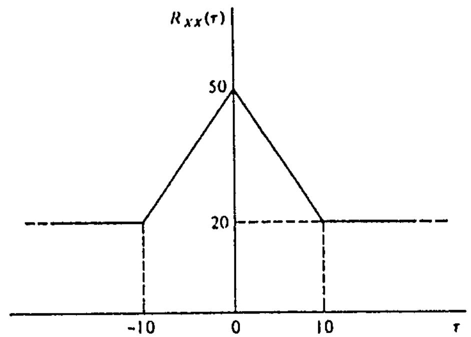


Figure P6-23

6-24 A random process $Y(t) = X(t) - X(t + \tau)$ is defined in terms of a process $X(t)$ that is at least wide-sense stationary.

(*a*) Show that the mean value of $Y(t)$ is 0 even if $X(t)$ has a nonzero mean value.

(*b*) Show that

$$\sigma_Y^2 = 2[R_{XX}(0) - R_{XX}(\tau)]$$

(*c*) If $Y(t) = X(t) + X(t + \tau)$, find $E[Y(t)]$ and σ_Y^2. How do these results compare to those of parts (*a*) and (*b*)?

6-25 For two zero-mean, jointly wide-sense stationary random processes $X(t)$ and $Y(t)$, it is known that $\sigma_X^2 = 5$ and $\sigma_Y^2 = 10$. Explain why each of the following functions cannot apply to the processes if they have no periodic components.

(*a*) $R_{XX}(\tau) = 6u(\tau)\exp(-3\tau)$ (*b*) $R_{XX}(\tau) = 5\sin(5\tau)$

(*c*) $R_{XY}(\tau) = 9(1 + 2\tau^2)^{-1}$ (*d*) $R_{YY}(\tau) = -\cos(6\tau)\exp(-|\tau|)$

(*e*) $R_{YY}(\tau) = 5\left[\dfrac{\sin(3\tau)}{3\tau}\right]^2$ (*f*) $R_{YY}(\tau) = 6 + 4\left[\dfrac{\sin(10\tau)}{10\tau}\right]$

6-26 Given two random processes $X(t)$ and $Y(t)$. Find expressions for the autocorrelation function of $W(t) = X(t) + Y(t)$ if:

(*a*) $X(t)$ and $Y(t)$ are correlated.

(*b*) They are uncorrelated.

(*c*) They are uncorrelated with zero means.

6-27 Use (6.3-19) to prove (6.3-17).

6-28 Let $X(t)$ be a stationary continuous random process that is differentiable. Denote its time-derivative by $\dot{X}(t)$.

(*a*) Show that $E[\dot{X}(t)] = 0$.

(*b*) Find $R_{X\dot{X}}(\tau)$ in terms of $R_{XX}(\tau)$.

(*c*) Find $R_{\dot{X}\dot{X}}(\tau)$ in terms of $R_{XX}(\tau)$. (*Hint:* Use the definition of the derivative

$$\dot{X}(t) = \lim_{\varepsilon \to 0} \frac{X(t + \varepsilon) - X(t)}{\varepsilon}$$

and assume the order of the limit and expectation operations can be interchanged.)

6-29 A gaussian random process has an autocorrelation function

$$R_{XX}(\tau) = 6\exp(-|\tau|/2)$$

Determine a covariance matrix for the random variables $X(t)$, $X(t + 1)$, $X(t + 2)$, and $X(t + 3)$.

6-30 Work Problem 6-29 if

$$R_{XX}(\tau) = 6\frac{\sin(\pi\tau)}{\pi\tau}$$

6-31 An ensemble member of a stationary random process $X(t)$ is sampled at N times t_i, $i = 1, 2, \ldots, N$. By treating the samples as random variables $X_i = X(t_i)$, an *estimate* or *measurement* $\hat{X}$ of the mean value $\bar{X} = E[X(t)]$ of the process is sometimes formed by averaging the samples:

$$\hat{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

(*a*) Show that $E[\hat{X}] = \bar{X}$.

(*b*) If the samples are separated far enough in time so that the random variables X_i can be considered statistically independent, show that the variance of the estimate of the process mean is

$$(\sigma_{\hat{X}})^2 = \sigma_X^2/N$$

6-32 For the random process and samples defined in Problem 6-31, let an estimate of the variance of the process be defined by

$$\widehat{\sigma_X^2} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{X})^2$$

Show that the mean value of this estimate is

$$E[\widehat{\sigma_X^2}] = \frac{N-1}{N}\sigma_X^2$$

6-33 Assume that $X(t)$ of Problem 6-31 is a zero-mean stationary gaussian process and let

$$\widehat{\sigma_X^2} = \frac{1}{N}\sum_{i=1}^{N} X_i^2$$

be an estimate of the variance σ_X^2 of $X(t)$ formed from the samples. Show that the variance of the estimate is

$$\text{variance of } \widehat{\sigma_X^2} = \frac{2\sigma_X^4}{N}$$

(*Hint:* Use the facts that $E[X^2] = \sigma_X^2$, $E[X^3] = 0$, and $E[X^4] = 3\sigma_X^4$ for a gaussian random variable having mean zero.)

6-34 How many samples must be taken in Problem 6-33 if the standard deviation of the estimate of the variance of $X(t)$ is to not exceed 5% of σ_X^2?

***6-35** A complex random process $Z(t) = X(t) + jY(t)$ is defined by jointly stationary real processes $X(t)$ and $Y(t)$. Show that

$$E[|Z(t)|^2] = R_{XX}(0) + R_{YY}(0)$$

***6-36** Let $X_1(t)$, $X_2(t)$, $Y_1(t)$ and $Y_2(t)$ be real random processes and define

$$Z_1(t) = X_1(t) + jY_1(t) \qquad Z_2(t) = X_2(t) - jY_2(t)$$

Find expressions for the cross-correlation function of $Z_1(t)$ and $Z_2(t)$ if:

(a) All the real processes are correlated.

(b) They are uncorrelated.

(c) They are uncorrelated with zero means.

*6-37 Let $Z(t)$ be a stationary complex random process with an autocorrelation function $R_{ZZ}(\tau)$. Define the random variable

$$W = \int_a^{a+T} Z(t)\, dt$$

where $T > 0$ and a are real numbers. Show that

$$E[|W|^2] = \int_{-T}^{T} (T - |\tau|) R_{ZZ}(\tau)\, d\tau$$

ADDITIONAL PROBLEMS

6-38 For a random process $X(t)$ it is known that $f_X(x_1, x_2, x_3; t_1, t_2, t_3) = f_X(x_1, x_2, x_3; t_1 + \Delta, t_2 + \Delta, t_3 + \Delta)$ for any t_1, t_2, t_3 and Δ. Indicate which of the following statements are unequivocably true: $X(t)$ is (a) stationary to order 1, (b) stationary to order 2, (c) stationary to order 3, (d) strictly stationary, (e) wide-sense stationary, (f) not stationary in any sense, and (g) ergodic.

6-39 A random process is defined by $X(t) = X_0 + Vt$ where X_0 and V are statistically independent random variables uniformly distributed on intervals $[X_{01}, X_{02}]$ and $[V_1, V_2]$, respectively. Find (a) the mean, (b) the autocorrelation, and (c) the autocovariance functions of $X(t)$. (d) Is $X(t)$ stationary in any sense? If so, state the type.

*6-40 (a) Find the first-order density of the random process of Problem 6-39. (b) Plot the density for $t = k(X_{02} - X_{01})/(V_2 - V_1)$ with $k = 0, \frac{1}{2}, 1,$ and 2. Assume $V_2 = 3V_1$ in all plots.

6-41 Assume a wide-sense stationary process $X(t)$ has a known mean $\bar{X}$ and a known autocorrelation function $R_{XX}(\tau)$. Now suppose the process is observed at time t_1 and we wish to *estimate*, that is, *predict*, what the process will be at time $t_1 + \tau$ with $\tau > 0$. We assume the estimate has the form

$$\hat{X}(t_1 + \tau) = \alpha X(t_1) + \beta$$

where α and β are constants.

(a) Find α and β so that the mean-squared prediction error

$$\overline{\varepsilon^2} = E[\{X(t_1 + \tau) - \hat{X}(t_1 + \tau)\}^2]$$

is minimum.

(b) Find the minimum mean-squared error in terms of $R_{XX}(\tau)$. Develop an alternative form in terms of the autocovariance function.

6-42 Find the time average and time autocorrelation function of the random process of Example 6.2-1. Compare these results with the statistical mean and autocorrelation found in the example.

6-43 Assume that an ergodic random process $X(t)$ has an autocorrelation function

$$R_{XX}(\tau) = 18 + \frac{2}{6 + \tau^2}[1 + 4\cos(12\tau)]$$

(a) Find $|\bar{X}|$.

(b) Does this process have a periodic component?

(c) What is the average power in $X(t)$?

6-44 Define a random process $X(t)$ as follows: (1) $X(t)$ assumes only one of two possible levels 1 or -1 at any time, (2) $X(t)$ switches back and forth between its two levels randomly with time, (3) the number of level transitions in any time interval τ is a Poisson random variable, that is, the probability of exactly k transitions, when the average rate of transitions is λ, is given by $[(\lambda\tau)^k/k!]\exp(-\lambda\tau)$, (4) transitions occurring in any time interval are statistically independent of transitions in any other interval, and (5) the levels at the start of any interval are equally probable. $X(t)$ is usually called the *random telegraph process*. It is an example of a discrete random process.

(a) Find the autocorrelation function of the process.

(b) Find probabilities $P\{X(t) = 1\}$ and $P\{X(t) = -1\}$ for any t.

(c) What is $E[X(t)]$?

(d) Discuss the stationarity of $X(t)$.

6-45 Work Problem 6-44 assuming the random telegraph signal has levels 0 and 1.

6-46 $\bar{X} = 6$ and $R_{XX}(t, t + \tau) = 36 + 25\exp(-|\tau|)$ for a random process $X(t)$. Indicate which of the following statements are true based on what is known with certainty. $X(t)$ (a) is first-order stationary, (b) has total average power of 61 W, (c) is ergodic, (d) is wide-sense stationary, (e) has a periodic component, and (f) has an ac power of 36 W.

6-47 A zero-mean random process $X(t)$ is ergodic, has average power of 24 W, and has no periodic components. Which of the following can be a valid autocorrelation function? If one cannot, state at least one reason why. (a) $16 + 18\cos(3\tau)$, (b) $24\mathrm{Sa}^2(2\tau)$, (c) $[1 + 3\tau^2]^{-1}\exp(-6\tau)$, and (d) $24\delta(t - \tau)$.

6-48 Use the result of Problem 6-18 to find the autocorrelation function of a random process with periodic sample function waveform $p(t)$ defined by

$$p(t) = A\cos^2(2\pi t/T)$$

where A and $T > 0$ are constants.

6-49 An engineer wants to measure the mean value of a noise signal that can be well-modeled as a sample function of a gaussian process. He uses the sampling estimator of Problem 6-31. After 100 samples he wishes his estimate to be within ± 0.1 V of the true mean with probability 0.9606. What is the largest variance the process can have such that his wishes will be true?

6-50 Let $X(t)$ be the sum of a deterministic signal $s(t)$ and a wide-sense stationary noise process $N(t)$. Find the mean value, and autocorrelation and autocovariance functions of $X(t)$. Discuss the stationarity of $X(t)$.

6-51 Random processes $X(t)$ and $Y(t)$ are defined by

$$X(t) = A \cos(\omega_0 t + \Theta)$$

$$Y(t) = B \cos(\omega_0 t + \Theta)$$

where A, B, and ω_0 are constants while Θ is a random variable uniform on $(0, 2\pi)$. By the procedures of Example 6.2-1 it is easy to find that $X(t)$ and $Y(t)$ are zero-mean, wide-sense stationary with autocorrelation functions

$$R_{XX}(\tau) = (A^2/2)\cos(\omega_0 \tau)$$

$$R_{YY}(\tau) = (B^2/2)\cos(\omega_0 \tau)$$

(a) Find the cross-correlation function $R_{XY}(t, t+\tau)$ and show that $X(t)$ and $Y(t)$ are jointly wide-sense stationary.

(b) Solve (6.4-2) and show that the response of the system of Figure 6.4-1 equals the true cross-correlation function plus an error term $\varepsilon(T)$ that decreases as T increases.

(c) Sketch $|\varepsilon(T)|$ versus T to show its behavior. How large must T be to make $|\varepsilon(T)|$ less than 1% of the largest value the correct cross-correlation function can have?

6-52 Consider random processes

$$X(t) = A \cos(\omega_0 t + \Theta)$$

$$Y(t) = B \cos(\omega_1 t + \Phi)$$

where A, B, ω_1, and ω_0 are constants, while Θ and Φ are statistically independent random variables uniform on $(0, 2\pi)$.

(a) Show that $X(t)$ and $Y(t)$ are jointly wide-sense stationary.

(b) If $\Theta = \Phi$ show that $X(t)$ and $Y(t)$ are not jointly wide-sense stationary unless $\omega_1 = \omega_0$.

6-53 A zero-mean gaussian random process has an autocorrelation function

$$R_{XX}(\tau) = \begin{cases} 13[1 - (|\tau|/6)] & |\tau| \le 6 \\ 0 & \text{elsewhere} \end{cases}$$

Find the covariance function necessary to specify the joint density of random variables defined at times $t_i = 2(i-1)$, $i = 1, 2, \ldots, 5$. Give the covariance matrix for the $X_i = X(t_i)$.

6-54 If the gaussian process of Problem 6-53 is shifted to have a constant mean $\bar{X} = -2$ but all else is unchanged, discuss how the autocorrelation function and covariance matrix change. What is the effect on the joint density of the five random variables?

*6-55 Extend Example 6.6-1 to allow the sum of complex-amplitude unequal-frequency phasors. Let Z_i, $i = 1, 2, \ldots, N$ be N complex zero-mean, uncorrelated random variables with variances $\sigma_{Z_i}^2$. Form a random process

$$Z(t) = \sum_{i=1}^{N} Z_i e^{j\omega_i t}$$

where ω_i are the frequencies of the phasors.

(a) Show that $E[Z(t)] = 0$.

(b) Derive the autocorrelation function and show that $Z(t)$ is wide-sense stationary.

*6-56 A complex random process is defined by

$$Z(t) = \exp(j\Omega t)$$

where Ω is a zero-mean random variable uniformly distributed on the interval from $\omega_0 - \Delta\omega$ to $\omega_0 + \Delta\omega$, where ω_0 and $\Delta\omega$ are positive constants. Find:

(a) the mean value, and (b) the autocorrelation function of $Z(t)$.

(c) Is $Z(t)$ wide-sense stationary?

*6-57 Work Problem 6-56 except assume the process

$$Z(t) = e^{j\Omega t} + e^{-j\Omega t} = 2\cos(\Omega t)$$

*6-58 Let $X(t)$ and $Y(t)$ be statistically independent wide-sense stationary real processes having the same autocorrelation function $R(\tau)$. Define the complex process

$$Z(t) = X(t)\cos(\omega_0 t) + jY(t)\sin(\omega_0 t)$$

where ω_0 is a positive constant. Find the autocorrelation function of $Z(t)$. Is $Z(t)$ wide-sense stationary?