



---

# **PROBABILITY AND RANDOM PROCESSES**

---

---

## **WILEY SURVIVAL GUIDES IN ENGINEERING AND SCIENCE**

**Emmanuel Desurvire, Editor**

**Wiley Survival Guide in Global Telecommunications: Signaling Principles,  
Network Protocols, and Wireless Systems** *Emmanuel Desurvire*

**Wiley Survival Guide in Global Telecommunications: Broadband Access,  
Optical Components and Networks, and Cryptography**

*Emmanuel Desurvire*

**Fiber to the Home: The New Empowerment** *Paul E. Green, Jr.*

**Probability and Random Processes** *Venkatarama Krishnan*

# PROBABILITY AND RANDOM PROCESSES

---

**Venkatarama Krishnan**

Professor Emeritus of Electrical Engineering  
University of Massachusetts Lowell



**WILEY-  
INTERSCIENCE**

A John Wiley & Sons, Inc., Publication

---

Copyright © 2006 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at [www.copyright.com](http://www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at [www.wiley.com](http://www.wiley.com).

***Library of Congress Cataloging-in-Publication Data:***

Krishnan, Venkatarama,

Probability and random processes / by Venkatarama Krishnan.

p. cm. — (Wiley survival guides in engineering and science)

Includes bibliographical references and index.

ISBN-13: 978-0-471-70354-9 (acid-free paper)

ISBN-10: 0-471-70354-0 (acid-free paper)

1. Probabilities. 2. Stochastic processes. 3. Engineering—Statistical methods.

4. Science—Statistical methods. I. Title. II. Series

QA273.K74 2006

519.2—dc22

2005057726

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

श्रीराम राम रामेति रमे रामे मनोरमे ।  
सहस्रनाम तत्तुल्यं राम नाम वरानने ॥

Vishnu sahasranamam

Shree rāma rāma rāmēti  
ramē rāmē manoramē  
Sahasra nāmatathtulyam  
rāmānāma varānanē

This book is respectfully dedicated to the

memory of my mother

*Dharmambal Venkataraman*

and the memory of my father

*B. Venkataraman*

மன்னனும் மாசறக் கற்றிருநும் சீர் தூக்கின்  
மன்னனிர்க்கற்றிருள் சிறப்புகூடமோன்  
மன்னர்க்குத் தன்தோசமல்லார்ச் சிறப்பில்லை  
கற்றுவனுக்குச் சிகங்றுவிடமெல்லாம் சிறப்பு

Avvaiyar (Tamil Poet)

# Contents

---

**Preface, xi**

## **CHAPTER 1**

### **Sets, Fields, and Events, 1**

- 1.1 Set Definitions, 1**
- 1.2 Set Operations, 3**
- 1.3 Set Algebras, Fields, and Events, 8**

## **CHAPTER 2**

### **Probability Space and Axioms, 10**

- 2.1 Probability Space, 10**
- 2.2 Conditional Probability, 14**
- 2.3 Independence, 18**
- 2.4 Total Probability and Bayes' Theorem, 20**

## **CHAPTER 3**

### **Basic Combinatorics, 25**

- 3.1 Basic Counting Principles, 25**
- 3.2 Permutations, 26**
- 3.3 Combinations, 28**

## **CHAPTER 4**

### **Discrete Distributions, 37**

- 4.1 Bernoulli Trials, 37**
- 4.2 Binomial Distribution, 38**
- 4.3 Multinomial Distribution, 41**
- 4.4 Geometric Distribution, 42**
- 4.5 Negative Binomial Distribution, 44**

**4.6 Hypergeometric Distribution, 46**

**4.7 Poisson Distribution, 48**

**4.8 Logarithmic Distribution, 55**

**4.9 Summary of Discrete Distributions, 62**

## **CHAPTER 5**

### **Random Variables, 64**

- 5.1 Definition of Random Variables, 64**
- 5.2 Determination of Distribution and Density Functions, 66**
- 5.3 Properties of Distribution and Density Functions, 73**
- 5.4 Distribution Functions from Density Functions, 75**

## **CHAPTER 6**

### **Continuous Random Variables and Basic Distributions, 79**

- 6.1 Introduction, 79**
- 6.2 Uniform Distribution, 79**
- 6.3 Exponential Distribution, 80**
- 6.4 Normal or Gaussian Distribution, 84**

## **CHAPTER 7**

### **Other Continuous Distributions, 95**

- 7.1 Introduction, 95**
- 7.2 Triangular Distribution, 95**
- 7.3 Laplace Distribution, 96**
- 7.4 Erlang Distribution, 97**

<b>7.5</b>	Gamma Distribution, 99
<b>7.6</b>	Weibull Distribution, 101
<b>7.7</b>	Chi-Square Distribution, 102
<b>7.8</b>	Chi and Other Allied Distributions, 104
<b>7.9</b>	Student- <i>t</i> Density, 110
<b>7.10</b>	Snedecor <i>F</i> Distribution, 111
<b>7.11</b>	Lognormal Distribution, 112
<b>7.12</b>	Beta Distribution, 114
<b>7.13</b>	Cauchy Distribution, 115
<b>7.14</b>	Pareto Distribution, 117
<b>7.15</b>	Gibbs Distribution, 118
<b>7.16</b>	Mixed Distributions, 118
<b>7.17</b>	Summary of Distributions of Continuous Random Variables, 119

**CHAPTER 8**  
**Conditional Densities and Distributions, 122**

<b>8.1</b>	Conditional Distribution and Density for $P(A) \neq 0$ , 122
<b>8.2</b>	Conditional Distribution and Density for $P(A) = 0$ , 126
<b>8.3</b>	Total Probability and Bayes' Theorem for Densities, 131

**CHAPTER 9**  
**Joint Densities and Distributions, 135**

<b>9.1</b>	Joint Discrete Distribution Functions, 135
<b>9.2</b>	Joint Continuous Distribution Functions, 136
<b>9.3</b>	Bivariate Gaussian Distributions, 144

**CHAPTER 10**  
**Moments and Conditional Moments, 146**

<b>10.1</b>	Expectations, 146
<b>10.2</b>	Variance, 149
<b>10.3</b>	Means and Variances of Some Distributions, 150

<b>10.4</b>	Higher-Order Moments, 153
<b>10.5</b>	Bivariate Gaussian, 154

**CHAPTER 11**  
**Characteristic Functions and Generating Functions, 155**

<b>11.1</b>	Characteristic Functions, 155
<b>11.2</b>	Examples of Characteristic Functions, 157
<b>11.3</b>	Generating Functions, 161
<b>11.4</b>	Examples of Generating Functions, 162
<b>11.5</b>	Moment Generating Functions, 164
<b>11.6</b>	Cumulant Generating Functions, 167
<b>11.7</b>	Table of Means and Variances, 170

**CHAPTER 12**  
**Functions of a Single Random Variable, 173**

<b>12.1</b>	Random Variable $g(X)$ , 173
<b>12.2</b>	Distribution of $Y = g(X)$ , 174
<b>12.3</b>	Direct Determination of Density $f_Y(y)$ from $f_X(x)$ , 194
<b>12.4</b>	Inverse Problem: Finding $g(x)$ Given $f_X(x)$ and $f_Y(y)$ , 200
<b>12.5</b>	Moments of a Function of a Random Variable, 202

**CHAPTER 13**  
**Functions of Multiple Random Variables, 206**

<b>13.1</b>	Function of Two Random Variables, $Z = g(X,Y)$ , 206
<b>13.2</b>	Two Functions of Two Random Variables, $Z = g(X,Y)$ , $W = h(X,Y)$ , 222
<b>13.3</b>	Direct Determination of Joint Density $f_{ZW}(z,w)$ from $f_{XY}(x,y)$ , 227
<b>13.4</b>	Solving $Z = g(X,Y)$ Using an Auxiliary Random Variable, 233
<b>13.5</b>	Multiple Functions of Random Variables, 238

**CHAPTER 14****Inequalities, Convergences, and Limit Theorems, 241**

- 14.1** Degenerate Random Variables, 241
- 14.2** Chebyshev and Allied Inequalities, 242
- 14.3** Markov Inequality, 246
- 14.4** Chernoff Bound, 248
- 14.5** Cauchy–Schwartz Inequality, 251
- 14.6** Jensen’s Inequality, 254
- 14.7** Convergence Concepts, 256
- 14.8** Limit Theorems, 259

**CHAPTER 15****Computer Methods for Generating Random Variates, 264**

- 15.1** Uniform-Distribution Random Variates, 264
- 15.2** Histograms, 266
- 15.3** Inverse Transformation Techniques, 269
- 15.4** Convolution Techniques, 279
- 15.5** Acceptance–Rejection Techniques, 280

**CHAPTER 16****Elements of Matrix Algebra, 284**

- 16.1** Basic Theory of Matrices, 284
- 16.2** Eigenvalues and Eigenvectors of Matrices, 293
- 16.3** Vectors and Matrix Differentiations, 301
- 16.4** Block Matrices, 308

**CHAPTER 17****Random Vectors and Mean-Square Estimation, 311**

- 17.1** Distributions and Densities, 311
- 17.2** Moments of Random Vectors, 319
- 17.3** Vector Gaussian Random Variables, 323

**17.4** Diagonalization of Covariance Matrices, 330

- 17.5** Simultaneous Diagonalization of Covariance Matrices, 334
- 17.6** Linear Estimation of Vector Variables, 337

**CHAPTER 18****Estimation Theory, 340**

- 18.1** Criteria of Estimators, 340
- 18.2** Estimation of Random Variables, 342
- 18.3** Estimation of Parameters (Point Estimation), 350
- 18.4** Interval Estimation (Confidence Intervals), 364
- 18.5** Hypothesis Testing (Binary), 373
- 18.6** Bayesian Estimation, 384

**CHAPTER 19****Random Processes, 406**

- 19.1** Basic Definitions, 406
- 19.2** Stationary Random Processes, 420
- 19.3** Ergodic Processes, 439
- 19.4** Estimation of Parameters of Random Processes, 445
- 19.5** Power Spectral Density, 472

**CHAPTER 20****Classification of Random Processes, 490**

- 20.1** Specifications of Random Processes, 490
- 20.2** Poisson Process, 492
- 20.3** Binomial Process, 505
- 20.4** Independent Increment Process, 507
- 20.5** Random-Walk Process, 512
- 20.6** Gaussian Process, 521
- 20.7** Wiener Process (Brownian Motion), 523

- 20.8** Markov Process, 527
- 20.9** Markov Chain, 536
- 20.10** Martingale Process, 551
- 20.11** Periodic Random Process, 557
- 20.12** Aperiodic Random Process  
(Karhunen–Loeve Expansion), 563

## CHAPTER 21

### Random Processes and Linear Systems, 574

- 21.1** Review of Linear Systems, 574
- 21.2** Random Processes through Linear Systems, 578
- 21.3** Linear Filters, 592
- 21.4** Bandpass Stationary Random Processes, 606

## CHAPTER 22

### Weiner and Kalman Filters, 625

- 22.1** Review of Orthogonality Principle, 625
- 22.2** Wiener Filtering, 627
- 22.3** Discrete Kalman Filter, 647
- 22.4** Continuous Kalman Filter, 660

## CHAPTER 23

### Probabilistic Methods in Transmission Tomography, 666

- 23.1** Introduction, 666
- 23.2** Stochastic Model, 667
- 23.3** Stochastic Estimation Algorithm, 671
- 23.4** Prior Distribution  $P\{\mathbf{M}\}$ , 674
- 23.5** Computer Simulation, 676
- 23.6** Results and Conclusions, 678
- 23.7** Discussion of Results, 681
- 23.8** References for Chapter 23, 681

## APPENDICES

- A** A Fourier Transform Tables, 683
- B** Cumulative Gaussian Tables, 687
- C** Inverse Cumulative Gaussian Tables, 692
- D** Inverse Chi-Square Tables, 694
- E** Inverse Student- $t$  Tables, 701
- F** Cumulative Poisson Distribution, 704
- G** Cumulative Binomial Distribution, 708

### References, 714

### Index, 716

# Preface

---

Many good textbooks exist on probability and random processes written at the undergraduate level to the research level. However, there is no one handy and ready book that explains most of the essential topics, such as random variables and most of their frequently used discrete and continuous probability distribution functions; moments, transformation, and convergences of random variables; characteristic and generating functions; estimation theory and the associated orthogonality principle; vector random variables; random processes and their autocovariance and cross-covariance functions; stationarity concepts; and random processes through linear systems and the associated Wiener and Kalman filters. Engineering practitioners and students alike have to delve through several books to get the required formulas or tables either to complete a project or to finish a homework assignment. This book may alleviate this difficulty to some extent and provide access to a compendium of most distribution functions used by communication engineers, queuing theory specialists, signal processing engineers, biomedical engineers, and physicists. Probability tables with accuracy up to nine decimal places are given in the appendixes to enhance the utility of this book. A particular feature is the presentation of commonly occurring Fourier transforms where both the time and frequency functions are drawn to scale.

Most of the theory has been explained with figures drawn to scale. To understand the theory better, more than 300 examples are given with every step explained clearly. Following the adage that a figure is worth more than a thousand words, most of the examples are also illustrated with figures drawn to scale, resulting in more than 400 diagrams. This book will be of particular value to graduate and undergraduate students in electrical, computer, and civil engineering as well as students in physics and applied mathematics for solving homework assignments and projects. It will certainly be useful to communication and signal processing engineers, and computer scientists in an industrial setting. It will also serve as a good reference for research workers in biostatistics and financial market analysis.

The salient features of this book are

- Functional and statistical independence of random variables are explained.
- Ready reference to commonly occurring density and distribution functions and their means and variances is provided.
- A section on Benford's logarithmic law, which is used in detecting tax fraud, is included.
- More than 300 examples, many of them solved in different ways to illustrate the theory and various applications of probability, are presented.
- Most examples have been substantiated with graphs drawn to scale.

- More than 400 figures have been drawn to scale to improve the clarity of understanding the theory and examples.
- Bounds on the tails of Gaussian probability have been carefully developed.
- A chapter has been devoted to computer generation of random variates.
- Another chapter has been devoted to matrix algebra so that some estimation problems such as the Kalman filter can be cast in matrix framework.
- Estimation problems have been given considerable exposure.
- Random processes defined and classified.
- A section on martingale processes with examples has been included.
- Markov chains with interesting examples have been discussed.
- Wiener and Kalman filtering have been discussed with a number of examples.
- Important properties are summarized in tables.
- The final chapter is on applications of probability to tomographic imaging.
- The appendixes consist of probability tables with nine-place decimal accuracy for the most common probability distributions.
- An extremely useful feature are the Fourier transform tables with both time and frequency functions graphed carefully to scale.

After the introduction of functional independence in Section 1.2, the differences between functional and statistical independences are discussed in Section 2.3 and clarified in Example 2.3.2. The foundation for permutations and combinations are laid out in Chapter 3, followed by discrete distributions in Chapter 4 with an end-of-chapter summary of discrete distributions. Random variables are defined in Chapter 5, and most of the frequently occurring continuous distributions are explained in Chapters 6 and 7. Section 6.4 discusses the new bounds on Gaussian tails. A comprehensive table is given at the end of Chapter 7 for all the continuous distributions and densities. After discussion of conditional densities, joint densities, and moments in Chapters 8–10, characteristic and other allied functions are explained in Chapter 11 with a presentation of an end-of-chapter table of means and variances of all discrete and continuous distributions. Functions of random variables are discussed in Chapters 12 and 13 with numerous examples. Chapter 14 discusses the various bounds on probabilities along with some commonly occurring inequalities. Computer generation of random variates is presented in Chapter 15. Elements of matrix algebra along with vector and matrix differentiations are considered in Chapter 16. Vector random variables and diagonalization of covariance matrices and simultaneous diagonalization of two covariance matrices are taken up in Chapter 17. Estimation and allied hypothesis testing are discussed in Chapter 18. After random processes are explained in Chapter 19, they are carefully classified in Chapter 20, with Section 20.10 presenting martingale processes that find wide use in financial engineering. Chapter 21 discusses the effects of passing random processes through linear systems. A number of examples illustrate the basic ideas of Wiener and Kalman filters in Chapter 22. The final chapter, Chapter 23, presents a practical application of probability to tomographic imaging. The appendixes include probability tables up to nine decimal places for Gaussian, chi-square, Student- $t$ , Poisson, and binomial distributions, and Fourier transform tables with graphs for time and frequency functions carefully drawn to scale.

This book is suitable for students and practicing engineers who need a quick reference to any probability distribution or table. It is also useful for self-study where a number of

carefully solved examples illustrate the applications of probability. Almost every topic has solved examples. It can be used as an adjunct to any textbook on probability and may also be prescribed as a textbook with homework problems drawn from other sources.

During the more than four decades that I have been continuously teaching probability and random processes, many authors who have written excellent textbooks have influenced me to a considerable extent. Many of the ideas and concepts that are expanded in this book are the direct result of this influence. I owe a scientific debt of gratitude to all the authors who have contributed to this ubiquitous and exciting field of probability. Some of these authors are listed in the reference, and the list is by no means exhaustive.

Most of the graphs and all the probability tables in this book were created with Mathcad software. All algebraic calculations were verified with Mathcad software. Mathcad and Mathsoft are registered trademarks of Mathsoft Engineering and Education, Inc., <http://www.mathcad.com>.

While facing extreme personal difficulties in writing this book, the unequivocal support of my entire family have been a source of inspiration in finishing it.

I acknowledge with great pleasure the support given to me by the Wiley staff. George Telecki welcomed the idea of this book; Rachel Witmer did more than her share of keeping me in good humor with her cheerful disposition in spite of the ever-expanding deadlines; and Kellsee Chu who did an excellent job in managing the production of this book. Finally, the enthusiastic support given to me by the Series Editor Emmanuel Desurvire was very refreshing. This is my second book with Wiley, and the skills of their copyeditors and staff who transform highly mathematical manuscript into a finished book continue to amaze me.

I have revised this book several times and corrected errors. Nevertheless, I cannot claim that it is error-free since correcting errors in any book is a convergent process. I sincerely hope that readers will bring to my attention any errors of commission or omission that they may come across.

VENKATARAMA KRISHNAN  
Chelmsford, Massachusetts  
March 2006

# Sets, Fields, and Events

## 1.1 SET DEFINITIONS

The concept of sets play an important role in probability. We will define a set in the following paragraph.

### **Definition of Set**

A *set* is a collection of objects called *elements*. The elements of a set can also be sets. Sets are usually represented by uppercase letters  $A$ , and elements are usually represented by lowercase letters  $a$ . Thus

$$A = \{a_1, a_2, \dots, a_n\} \quad (1.1.1)$$

will mean that the set  $A$  contains the elements  $a_1, a_2, \dots, a_n$ . Conversely, we can write that  $a_k$  is an element of  $A$  as

$$a_k \in A \quad (1.1.2)$$

and  $a_k$  is not an element of  $A$  as

$$a_k \notin A \quad (1.1.3)$$

A finite set contains a finite number of elements, for example,  $S = \{2, 4, 6\}$ . Infinite sets will have either countably infinite elements such as  $A = \{x : x \text{ is all positive integers}\}$  or uncountably infinite elements such as  $B = \{x : x \text{ is real number } \leq 20\}$ .

**Example 1.1.1** The set  $A$  of all positive integers less than 7 is written as

$$A = \{x : x \text{ is a positive integer } < 7\}: \text{finite set.}$$

**Example 1.1.2** The set  $N$  of all positive integers is written as

$$N = \{x : x \text{ is all positive integers}\}: \text{countably infinite set.}$$

**Example 1.1.3** The set  $R$  of all real numbers is written as

$$R = \{x : x \text{ is real}\}: \text{uncountably infinite set.}$$

**Example 1.1.4** The set  $R^2$  of real numbers  $x, y$  is written as

$$R^2 = \{(x,y) : x \text{ is real, } y \text{ is real}\}$$

**Example 1.1.5** The set  $C$  of all real numbers  $x, y$  such that  $x + y \leq 10$  is written as

$$C = \{(x,y) : x + y \leq 10\}: \text{uncountably infinite set.}$$

### Venn Diagram

Sets can be represented graphically by means of a Venn diagram. In this case we assume tacitly that  $S$  is a universal set under consideration. In Example 1.1.5, the universal set  $S = \{x : x \text{ is all positive integers}\}$ . We shall represent the set  $A$  in Example 1.1.1 by means of a Venn diagram of Fig. 1.1.1.

### Empty Set

An empty is a set that contains no element. It plays an important role in set theory and is denoted by  $\emptyset$ . The set  $A = \{0\}$  is not an empty set since it contains the element 0.

### Cardinality

The number of elements in the set  $A$  is called the *cardinality* of set  $A$ , and is denoted by  $|A|$ . If it is an infinite set, then the cardinality is  $\infty$ .

**Example 1.1.6** The cardinality of the set  $A = \{2,4,6\}$  is 3, or  $|A| = 3$ . The cardinality of set  $R = \{x : x \text{ is real}\}$  is  $\infty$ .

**Example 1.1.7** The cardinality of the set  $A = \{x : x \text{ is positive integer} < 7\}$  is  $|A| = 6$ .

**Example 1.1.8** The cardinality of the set  $B = \{x : x \text{ is a real number} < 10\}$  is infinity since there are infinite real numbers  $< 10$ .

### Subset

A set  $B$  is a subset of  $A$  if every element in  $B$  is an element of  $A$  and is written as  $B \subseteq A$ .  $B$  is a proper subset of  $A$  if *every* element of  $A$  is not in  $B$  and is written as  $B \subset A$ .

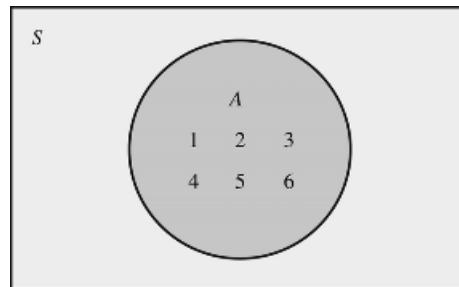


FIGURE 1.1.1

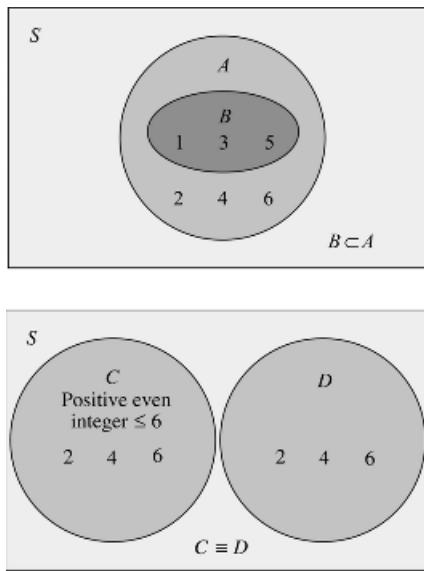


FIGURE 1.1.2

### Equality of Sets

Two sets  $A$  and  $B$  are equal if  $B \subseteq A$  and  $A \subseteq B$ , that is, if every element of  $A$  is contained in  $B$  and every element of  $B$  is contained in  $A$ . In other words, sets  $A$  and  $B$  contain exactly the same elements. Note that this is different from having the same cardinality, that is, containing the *same number* of elements.

**Example 1.1.9** The set  $B = \{1,3,5\}$  is a proper subset of  $A = \{1,2,3,4,5,6\}$ , whereas the set  $C = \{x : x \text{ is a positive even integer } \leq 6\}$  and the set  $D = \{2,4,6\}$  are the same since they contain the same elements. The cardinalities of  $B$ ,  $C$ , and  $D$  are 3 and  $C = D$ .

We shall now represent the sets  $A$  and  $B$  and the sets  $C$  and  $D$  in Example 1.1.9 by means of the Venn diagram of Fig. 1.1.2 on a suitably defined universal set  $S$ .

### Power Set

The power set of any set  $A$  is the set of all possible subsets of  $A$  and is denoted by  $PS(A)$ . Every power set of any set  $A$  must contain the set  $A$  itself and the empty set  $\emptyset$ . If  $n$  is the cardinality of the set  $A$ , then the cardinality of the power set  $|PS(A)| = 2^n$ .

**Example 1.1.10** If the set  $A = \{1,2,3\}$  then  $PS(A) = \{\emptyset, (1,2,3), (1,2), (2,3), (3,1), (1), (2), (3)\}$ . The cardinality  $|PS(A)| = 8 = 2^3$ .

## 1.2 SET OPERATIONS

### Union

Let  $A$  and  $B$  be sets belonging to the universal set  $S$ . The *union* of sets  $A$  and  $B$  is another set  $C$  whose elements are those that are in either  $A$  or  $B$ , and is denoted by  $A \cup B$ . Where there is no confusion, it will also be represented as  $A + B$ :

$$A \cup B = A + B = \{x : x \in A \text{ or } x \in B\} \quad (1.2.1)$$

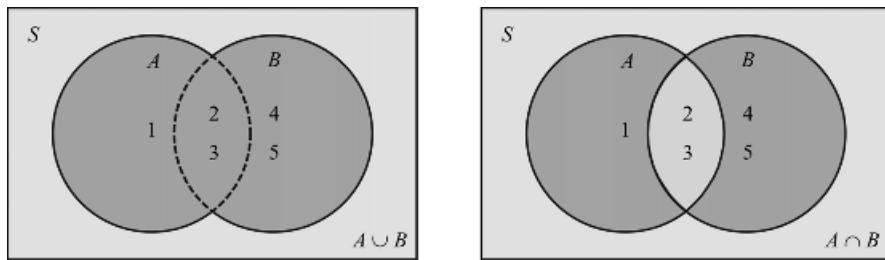


FIGURE 1.2.1

**Example 1.2.1** The union of sets  $A = \{1,2,3\}$  and  $B = \{2,3,4,5\}$  is the set  $C = A \cup B = \{1,2,3,4,5\}$ .

### Intersection

The *intersection* of the sets  $A$  and  $B$  is another set  $C$  whose elements are the same as those in both  $A$  and  $B$  and is denoted by  $A \cap B$ . Where there is no confusion, it will also be represented by  $AB$ .

$$A \cap B = AB = \{x : x \in A \text{ and } x \in B\} \quad (1.2.2)$$

**Example 1.2.2** The intersection of the sets  $A$  and  $B$  in Example 1.2.1 is the set  $C = \{2,3\}$ . Examples 1.2.1 and 1.2.2 are shown in the Venn diagram of Fig. 1.2.1.

### Mutually Exclusive Sets

Two sets  $A$  and  $B$  are called *mutually exclusive* if their intersection is empty. Mutually exclusive sets are also called *disjoint*.

$$A \cap B = \emptyset \quad (1.2.3)$$

One way to determine whether two sets  $A$  and  $B$  are mutually exclusive is to check whether set  $B$  can occur when set  $A$  has already occurred and vice versa. If it cannot, then  $A$  and  $B$  are mutually exclusive. For example, if a single coin is tossed, the two sets, {heads} and {tails}, are mutually exclusive since {tails} cannot occur when {heads} has already occurred and vice versa.

### Independence

We will consider two types of independence. The first is known as *functional independence* [42].\* Two sets  $A$  and  $B$  can be called functionally independent if the occurrence of  $B$  does not in any way influence the occurrence of  $A$  and vice versa. The second one is *statistical independence*, which is a different concept that will be defined later. As an example, the tossing of a coin is functionally independent of the tossing of a die because they do not depend on each other. However, the tossing of a coin and a die are not mutually exclusive since any one can be tossed irrespective of the other. By the same token, pressure and temperature are not functionally independent because the physics of the problem, namely, Boyle's law, connects these quantities. They are certainly not mutually exclusive.

\*Numbers in brackets refer to bibliographic entries in the References section at the end of the book.

### Cardinality of Unions and Intersections

We can now ascertain the cardinality of the union of two sets  $A$  and  $B$  that are not mutually exclusive. The cardinality of the union  $C = A \cup B$  can be determined as follows. If we add the cardinality  $|A|$  to the cardinality  $|B|$ , we have added the cardinality of the intersection  $|A \cap B|$  twice. Hence we have to subtract once the cardinality  $|A \cap B|$  as shown in Fig. 1.2.2. Or, in other words

$$|A \cup B| = |A| + |B| - |A \cap B| \quad (1.2.4a)$$

In Fig. 1.2.2 the cardinality  $|A| = 9$  and the cardinality  $|B| = 11$  and the cardinality  $|A \cup B|$  is  $11 + 9 - 4 = 16$ .

As a corollary, if sets  $A$  and  $B$  are mutually exclusive, then the cardinality of the union is the sum of the cardinalities; or

$$|A \cup B| = |A| + |B| \quad (1.2.4b)$$

The generalization of this result to an arbitrary union of  $n$  sets is called the *inclusion-exclusion* principle, given by

$$\begin{aligned} \left| \bigcup_{i=1}^n A_i \right| &= \sum_{i=1}^n |A_i| - \sum_{\substack{i,j=1 \\ i \neq j}}^n |A_i \cap A_j| + \sum_{\substack{i,j,k=1 \\ i \neq j \neq k}}^n |A_i \cap A_j \cap A_k| \\ &\quad - (\pm 1)^n \sum_{\substack{i,j,k=1 \\ i \neq j \neq k \dots \neq n}}^n |A_i \cap A_j \cap A_k \dots \cap A_n| \end{aligned} \quad (1.2.5a)$$

If the sets  $\{A_i\}$  are mutually exclusive, that is,  $A_i \cap A_j = \emptyset$  for  $i \neq j$ , then we have

$$\left| \bigcup_{i=1}^n A_i \right| = \sum_{i=1}^n |A_i| \quad (1.2.5b)$$

This equation is illustrated in the Venn diagram for  $n = 3$  in Fig. 1.2.3, where if  $|A \cup B \cup C|$  equals,  $|A| + |B| + |C|$  then we have added twice the cardinalities of  $|A \cap B|$ ,  $|B \cap C|$  and  $|C \cap A|$ . However, if we subtract once  $|A \cap B|$ ,  $|B \cap C|$ , and  $|C \cap A|$  and write

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |C \cap A|$$

then we have subtracted  $|A \cap B \cap C|$  thrice instead of twice. Hence, adding  $|A \cap B \cap C|$  we get the final result:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |C \cap A| - |A \cap B \cap C| \quad (1.2.6)$$

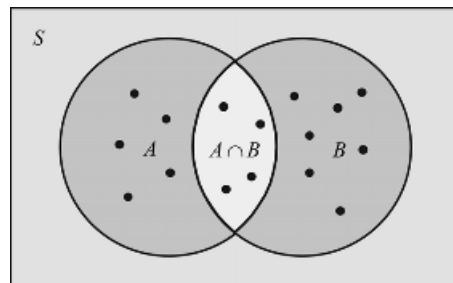


FIGURE 1.2.2

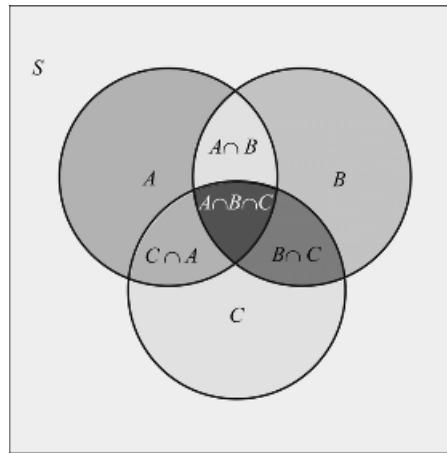


FIGURE 1.2.3

**Example 1.2.3** In a junior class, the number of students in electrical engineering (EE) is 100, in math (MA) 50, and in computer science (CS) 150. Among these 20 are taking both EE and MA, 25 are taking EE and CS, and 10 are taking MA and CS. Five of them are taking EE, CS and MA. Find the total number of students in the junior class.

From the problem we can identify the sets  $A = \text{EE students}$ ,  $B = \text{MA students}$ , and  $C = \text{CS students}$ , and the corresponding intersections are  $A \cap B$ ,  $B \cap C$ ,  $C \cap A$ , and  $A \cap B \cap C$ . Here,  $|A| = 100$ ,  $|B| = 50$ , and  $|C| = 150$ . We are also given  $|A \cap B| = 20$ ,  $|B \cap C| = 10$ , and  $|C \cap A| = 25$  and finally  $|A \cap B \cap C| = 5$ . Using Eq. (1.2.6) the total number of students are  $100 + 50 + 150 - 20 - 10 - 25 + 5 = 250$ .

### Complement

If  $A$  is a subset of the universal set  $S$ , then the complement of  $A$  denoted by  $\bar{A}$  is the elements in  $S$  not contained in  $A$ ; or

$$\bar{A} = S - A = \{x : x \notin A \subset S \text{ and } x \in S\} \quad (1.2.7)$$

**Example 1.2.4** If the universal set  $S = \{1,2,3,4,5,6\}$  and  $A = \{2,4,5\}$ , then the complement of  $A$  is given by  $\bar{A} = \{1,3,6\}$ .

### Difference

The complement of a subset  $A \subset S$  as given by Eq. (1.2.7) is the difference between the universal set  $S$  and the subset  $A$ . The difference between any two sets  $A$  and  $B$  denoted by  $A - B$  is the set  $C$  containing those elements that are in  $A$  but not in  $B$  and,  $B - A$  is the set  $D$  containing those elements in  $B$  but not in  $A$ :

$$\begin{aligned} C &= A - B = \{x : x \in A \text{ and } x \notin B\} \\ D &= B - A = \{x : x \in B \text{ and } x \notin A\} \end{aligned} \quad (1.2.8)$$

$A - B$  is also called the *relative complement* of  $B$  with respect to  $A$  and  $B - A$  is called the relative complement of  $A$  with respect to  $B$ . It can be verified that  $A - B \neq B - A$ .

**Example 1.2.5** If the set  $A$  is given as  $A = \{2,4,5,6\}$  and  $B$  is given as  $B = \{5,6,7,8\}$ , then the difference  $A - B = \{2,4\}$  and the difference  $B - A = \{7,8\}$ . Clearly,  $A - B \neq B - A$ .

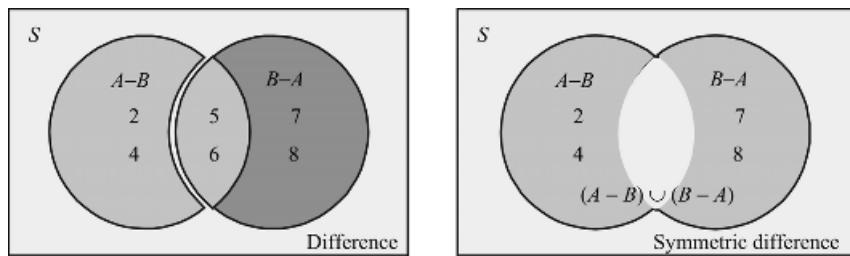


FIGURE 1.2.4

### Symmetric Difference

The symmetric difference between sets  $A$  and  $B$  is written as  $A \Delta B$  and defined by

$$A \Delta B = (A - B) \cup (B - A) = \{x : x \in A \text{ and } x \notin B\} \text{ or } \{x : x \in B \text{ and } x \notin A\} \quad (1.2.9)$$

**Example 1.2.6** The symmetric difference between the set  $A$  given by  $A = \{2,4,5,6\}$  and  $B$  given by  $B = \{5,6,7,8\}$  in Example 1.2.5 is  $A \Delta B = \{2,4\} \cup \{7,8\} = \{2,4,7,8\}$ .

The difference and symmetric difference for Examples 1.2.5 and 1.2.6 are shown in Fig. 1.2.4.

### Cartesian Product

This is a useful concept in set theory. If  $A$  and  $B$  are two sets, the Cartesian product  $A \times B$  is the set of ordered pairs  $(x,y)$  such that  $x \in A$  and  $y \in B$  and is defined by

$$A \times B = \{(x,y) : x \in A, y \in B\} \quad (1.2.10)$$

When ordering is considered the cartesian product  $A \times B \neq B \times A$ . The cardinality of a Cartesian product is the product of the individual cardinalities, or  $|A \times B| = |A| \cdot |B|$ .

**Example 1.2.7** If  $A = \{2,3,4\}$  and  $B = \{5,6\}$ , then  $C = A \times B = \{(2,5), (2,6), (3,5), (3,6), (4,5), (4,6)\}$  with cardinality  $3 \times 2 = 6$ . Similarly,  $D = B \times A = \{(5,2), (5,3), (5,4), (6,2), (6,3), (6,4)\}$  with the same cardinality of 6. However,  $C$  and  $D$  do not contain the same ordered pairs.

### Partition

A *partition* is a collection of disjoint sets  $\{A_i, i = 1, \dots, n\}$  of a set  $S$  such that  $\bigcup A_i$  over all  $i$  equals  $S$  and  $A_i \cap A_j$ , with  $i \neq j$  empty. Partitions play a useful role in conditional probability and Bayes' theorem. Figure 1.2.5 shows the concept of a partition.

$S$	$A_1$	$A_2$	$A_3$
$A_4$			
$A_7$			
$A_{10}$			
$A_{13}$			$A_n$

FIGURE 1.2.5

$S$ $A_1$ 1      5 3	$A_2$ 7      9
$A_3$ 2      4 6	$A_4$ 8      10

FIGURE 1.2.6

TABLE 1.2.1 Table of set properties

Property	Equation
Identity	$A \cup \emptyset = A$ $A \cap S = A$
Domination	$A \cap \emptyset = \emptyset$ $A \cup S = S$
Idempotent	$A \cap A = A$ $A \cup A = A$
Complementation	$\overline{\overline{A}} = A$
Commutative	$A \cup B = B \cup A$ $A \cap B = B \cap A$
Associative	$A \cup (B \cup C) = (A \cup B) \cup C$ $A \cap (B \cap C) = (A \cap B) \cap C$
Distributive	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
Noncommutative	$A - B \neq B - A$
De Morgan's law	$\overline{A \cup B} = \overline{A} \cap \overline{B}$ $\overline{A \cap B} = \overline{A} \cup \overline{B}$

**Example 1.2.8** If set  $S = \{1,2,3,4,5,6,7,8,9,10\}$ , then the collections of sets  $\{A_1, A_2, A_3, A_4\}$  where  $A_1 = \{1,3,5\}$ ,  $A_2 = \{7,9\}$ ,  $A_3 = \{2,4,6\}$ ,  $A_4 = \{8,10\}$  is a partition of  $S$  as shown in Fig. 1.2.6.

A tabulation of set properties is shown on Table 1.2.1. Among the identities, De Morgan's laws are quite useful in simplifying set algebra.

## 1.3 SET ALGEBRAS, FIELDS, AND EVENTS

### Boolean Field

We shall now consider a universal set  $S$  and a collection  $\mathcal{F}$  of subsets of  $S$  consisting of the sets  $\{A_i: i = 1,2,\dots,n,n+1,\dots\}$ . The collection  $\mathcal{F}$  is called a *Boolean field* if the

following two conditions are satisfied:

1. If  $A_i \in \mathcal{F}$ , then  $\overline{A}_i \in \mathcal{F}$ .
2. If  $A_1 \in \mathcal{F}$  and  $A_2 \in \mathcal{F}$ , then  $A_1 \cup A_2 \in \mathcal{F}$ .
3. If  $\{A_i, i = 1, \dots, n\} \in \mathcal{F}$ , then  $\bigcup_{i=1}^n A_i \in \mathcal{F}$ .

Let us see the consequences of these conditions being satisfied. If  $A_1 \in \mathcal{F}$  and  $\overline{A}_1 \in \mathcal{F}$ , then  $A_1 \cup \overline{A}_1 = S \in \mathcal{F}$ . If  $S \in \mathcal{F}$ , then  $\overline{S} = \emptyset \in \mathcal{F}$ . If  $A_1 \cup A_2 \in \mathcal{F}$ , then by condition 1  $\overline{A_1 \cup A_2} \in \mathcal{F}$ , and by De Morgan's law,  $\overline{A_1} \cup \overline{A_2} \in \mathcal{F}$ . Hence  $A_1 \cap A_2 \in \mathcal{F}$ . Also  $A_1 - A_2 \in \mathcal{F}$  and  $(A_1 - A_2) \cup (A_2 - A_1) \in \mathcal{F}$ . Thus the conditions listed above are sufficient for any field  $\mathcal{F}$  to be *closed* under all set operations.

### Sigma Field

The preceding definition of a field covers only finite set operations. Condition 3 for finite additivity may not hold for infinite additivity. For example, the sum of  $1 + 2 + 2^2 + 2^3 + 2^4 = (2^5 - 1)/(2 - 1) = 31$ , but the infinite sum  $1 + 2 + 2^2 + 2^3 + 2^4 + \dots$  diverges. Hence there is a need to extend finite additivity concept to infinite set operations. Thus, we have a  $\sigma$  field if the following extra condition is imposed for infinite additivity:

4. If  $A_1, A_2, \dots, A_n, \dots \in \mathcal{F}$  then  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Many  $\sigma$  fields may contain the subsets  $\{A_i\}$  of  $S$ , but the smallest  $\sigma$  field containing the subsets  $\{A_i\}$  is called the *Borel  $\sigma$  field*. The smallest  $\sigma$  field for  $S$  by itself is  $\mathcal{F} = \{S, \emptyset\}$ .

**Example 1.3.1** We shall assume that  $S = \{1,2,3,4,5,6\}$ . Then the collection of the following subsets of  $S$ ,  $\mathcal{F} = \{S, \emptyset, (1,2,3), (4,5,6)\}$  is a field since it satisfies all the set operations such as  $(1,2,3) \cup (4,5,6) = S$ ,  $(\overline{1,2,3}) = (4,5,6)$ . However, the collection of the following subsets of  $S$ ,  $\mathcal{F}_1 = \{S, \emptyset, (1,2,3), (4,5,6), (2)\}$  will not constitute a field because  $(2) \cup (4,5,6) = (2,4,5,6)$  is not in the field. But we can adjoin the missing sets and make  $\mathcal{F}_1$  into a field. This is known as *completion*. In the example above, if we adjoin the collection of sets  $\mathcal{F}_2 = \{(2,4,5,6), (1,3)\}$  to  $\mathcal{F}_1$ , then the resulting collection  $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2 = \{S, \emptyset, (1,2,3), (4,5,6), \{2\}, (2,4,5,6), (1,3)\}$  is a field.

### Event

Given the universal set  $S$  and the associated  $\sigma$  field  $\mathcal{F}$ , all subsets of  $S$  belonging to the field  $\mathcal{F}$  are called *events*. All events are subsets of  $S$  and can be assigned a probability, but not all subsets of  $S$  are events unless they belong to an associated  $\sigma$  field.

# Probability Space and Axioms

## 2.1 PROBABILITY SPACE

A *probability space* is defined as the triplet  $\{S, \mathcal{F}, P\}$ , where  $S$  is the sample space,  $\mathcal{F}$  is the field of subsets of  $S$ , and  $P$  is a probability measure. We shall now define these elements.

### **Sample Space $S$**

Finest-grain, mutually exclusive, and collectively exhaustive listing of all possible outcomes of a mathematical experiment is called the *sample space*. The outcomes can be either finite, countably infinite, or uncountably infinite. If  $S$  is countable, we call it a *discrete sample space*; if it is uncountable, we call it *continuous sample space*.

#### **Examples of Discrete Sample Space**

**Example 2.1.1** When a coin is tossed, the outcomes are heads  $\{H\}$  or tails  $\{T\}$ . Thus the sample space  $S = \{H, T\}$ .

**Example 2.1.2** When a die is tossed, the outcomes are 1,2,3,4,5,6. Thus the sample space is  $S = \{1, 2, 3, 4, 5, 6\}$ .

**Example 2.1.3** The set of all positive integers is countable, and the sample space is  $S = \{x : x \text{ is a positive integer}\}$

#### **Examples of Continuous Sample Space**

**Example 2.1.4** The set of all real numbers between 0 and 2 is the sample space  $S = \{x : 0 \leq x \leq 2\}$ .

**Example 2.1.5** The set of all real numbers  $x$  and  $y$  such that  $x$  is between 0 and 1 and  $y$  is between 1 and 2 is given by  $S = \{x, y : 0 \leq x \leq 1, 1 \leq y \leq 2\}$ .

The continuous sample space given in Examples 2.1.4 and 2.1.5 is shown in Fig. 2.1.1.

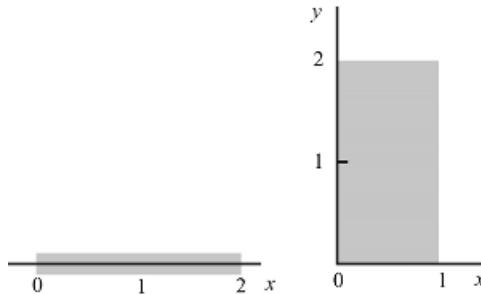


FIGURE 2.1.1

### Field $\mathcal{F}$

A *field* has already been defined in Section 1.3. We shall assume that a proper Borel  $\sigma$  field can be defined for all discrete and continuous sample spaces.

### Probability Measure $P$

A *probability measure* is defined on the field  $\mathcal{F}$  of events of the sample space  $S$ . It is a mapping of the sample space to the real line between 0 and 1 such that it satisfies the following three Kolmogorov axioms for all events belonging to the field  $\mathcal{F}$ :

*Axiom 1:* If  $A$  belongs to  $\mathcal{F}$ , then  $P(A) \geq 0$ : nonnegativity. (2.1.1)

*Axiom 2:* The probability of the sample space equals 1.

$$P(S) = 1: \text{normalization.} \quad (2.1.2)$$

*Axiom 3a:* If  $A$  and  $B$  are disjoint sets belonging to  $\mathcal{F}$ , that is,  $A \cap B = \emptyset$ , then

$$P\{A \cup B\} = P(A) + P(B): \text{Boolean additivity.} \quad (2.1.3)$$

*Axiom 3b:* If  $\{A_1, A_2, \dots\}$  is a sequence of sets belonging to the field  $\mathcal{F}$  such that  $A_i \cap A_j = \emptyset$  for all  $i \neq j$ , then

$$P\{A_1 \cup A_2 \cup \dots \cup A_i \cup \dots\} = \sum_{i=1}^{\infty} P\{A_i\}: \text{sigma additivity.} \quad (2.1.4)$$

From these axioms we can write the following corollaries:

*Corollary 1:*

$$P\{\emptyset\} \quad (2.1.5)$$

Note that the probability of the empty event is zero, but if the probability of an event  $A$  is zero, we cannot conclude that  $A = \emptyset$ . All sets with zero probability are defined as *null sets*. An empty set is also a null set, but null sets need not be empty.

*Corollary 2.* Since  $A \cup \bar{A} = S$  and  $A \cap \bar{A} = \emptyset$ , we can write

$$P(\bar{A}) = 1 - P(A) \quad (2.1.6)$$

*Corollary 3.* For any  $A$  and  $B$  that are subsets of  $S$ , not mutually exclusive, we can write

$$P\{A \cup B\} = P\{A \cup \bar{A} \cap B\} = P\{A\} + P\{\bar{A} \cap B\}$$

$$P\{B\} = P\{(A \cap B) \cup (\bar{A} \cap B)\} = P\{A \cap B\} + P\{\bar{A} \cap B\}$$

Eliminating  $P\{\bar{A} \cup B\}$  between the two equations, we obtain

$$P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\} \quad (2.1.7)$$

Clearly, if  $P\{A \cap B\} = 0$  implying  $A$  and  $B$  are mutually exclusive, then Eq. (2.1.7) reduces to Kolmogorov's axiom 3a. If the probabilities  $P\{A\}$  and  $P\{B\}$  are equal, then sets  $A$  and  $B$  are equal in probability. This does not mean that  $A$  is equal to  $B$ .

The probability of the union of three sets can be calculating from the Venn diagram for the three sets shown in Fig. 1.2.3 and redrawn in Fig. 2.1.2.

The results of the analysis for the cardinalities given in Eq. (1.2.6), can be extended to finding the  $P(A \cup B \cup C)$ . We subtract  $\{P(A \cap B) + P(B \cap C) + P(C \cap A)\}$  from  $\{P(A) + P(B) + P(C)\}$  and add  $P(A \cap B \cap C)$ , resulting in

$$\begin{aligned} P\{A \cup B \cup C\} &= P\{A\} + P\{B\} + P\{C\} \\ &\quad - P\{A \cap B\} - P\{B \cap C\} - P\{C \cap A\} \\ &\quad + P\{A \cap B \cap C\} \end{aligned} \quad (2.1.8)$$

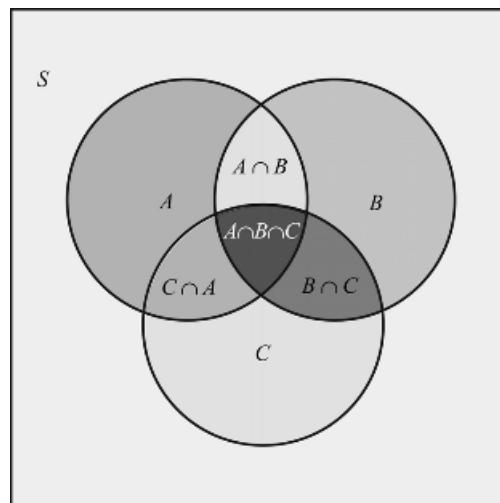


FIGURE 2.1.2

Using the inclusion–exclusion principle of Eq. (1.2.5a), we can also generalize Eq. (2.1.8) to the probabilities of the union of  $n$  sets  $A_1, A_2, \dots, A_n$  as follows:

$$\begin{aligned} P\{\bigcup_{i=1}^n A_i\} &= \sum_{i=1}^n P\{A_i\} - \sum_{\substack{i,j=1 \\ i \neq j}}^n P\{A_i \cap A_j\} + \sum_{\substack{i,j,k=1 \\ i \neq j \neq k}}^n P\{A_i \cap A_j \cap A_k\} \\ &\quad - (\pm 1)^n \sum_{\substack{i,j,k=1 \\ i \neq j \neq k \dots \neq n}}^n P\{A_i \cap A_j \cap A_k \dots \cap A_n\} \end{aligned} \quad (2.1.9)$$

**Example 2.1.6** A fair coin is tossed 3 times. Since each toss is a functionally independent trial and forms a Cartesian product, the cardinality of the three tosses is  $2 \times 2 \times 2 = 8$  and the sample space consists of the following 8 events:

$$S = \{[HHH], [HHT], [HTT], [HTH], [THH], [TTH], [THT], [TTT]\}$$

In this sample space

$$P\{2 \text{ heads}\} = P\{[HHH], [HHT], [HTH], [THH]\} = \frac{4}{8}$$

$$P\{\text{both heads and tails}\} = \{[HHT], [HTT], [HTH], [THH], [TTH], [THT]\} = \frac{6}{8}$$

**Example 2.1.7** A telephone call occurs at random between 9:00 a.m. and 11:00 a.m. Hence the probability that the telephone call occurs in the interval  $0 < t \leq t_0$  is given by  $t_0/120$ . Thus the probability that the call occurs between 9:30 a.m. and 10:45 a.m. is given by:

$$P\{30 < t \leq 105\} = \frac{105 - 30}{120} = \frac{75}{120} = \frac{5}{8}$$

**Example 2.1.8** Mortality tables give the probability density that a person who is born today assuming that the lifespan is 100 years as

$$f(t) = 3 \times 10^{-9} t^2 (100 - t^2) u(t)$$

and the distribution function that the person will die at the age  $t_0$  years as

$$F(t_0) = P\{t \leq t_0\} = \int_0^{t_0} 3 \times 10^{-9} t^2 (100 - t^2) dt$$

These probabilities are shown in Fig. 2.1.3. The probability that this person will die between the ages 70 to 80 is given by  $P\{70 < t \leq 80\} = 0.1052$ , obtained from the following integral:

$$P\{70 < t \leq 80\} = F(80) - F(70) = \int_{70}^{80} 3 \times 10^{-9} t^2 (100 - t^2) dt = 0.1052$$

The density and distribution functions are shown in Fig. 2.1.3.

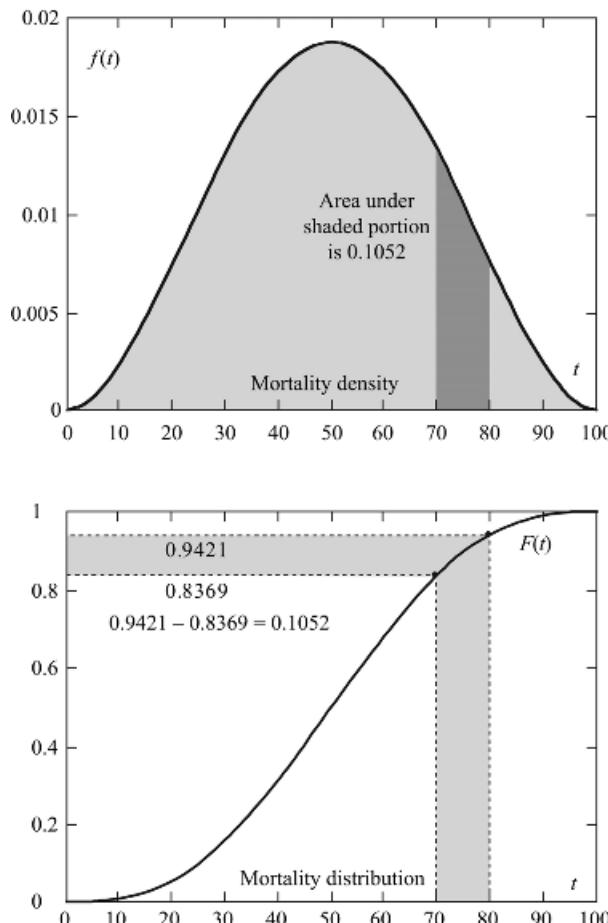


FIGURE 2.1.3

## 2.2 CONDITIONAL PROBABILITY

The conditional probability of an event  $B$  conditioned on the occurrence of another event  $B$  is defined by

$$\left. \begin{aligned} P\{B | A\} &= \frac{P\{B \cap A\}}{P\{A\}} \\ P\{B \cap A\} &= P\{B | A\}P\{A\} \end{aligned} \right\} \text{ if } P\{A\} \neq 0 \quad (2.2.1)$$

If  $P\{A\} = 0$ , then the conditional probability  $P\{B | A\}$  is not defined. Since  $P\{B | A\}$  is a probability measure, it also satisfies Kolmogorov axioms.

Conditional probability  $P\{B | A\}$  can also be interpreted as the probability of the event  $B$  in the reduced sample space  $S_1 = A$ . Sometimes it is simpler to use the reduced sample space to solve conditional probability problems. We shall use both these interpretations in obtaining conditional probabilities as shown in the following examples.

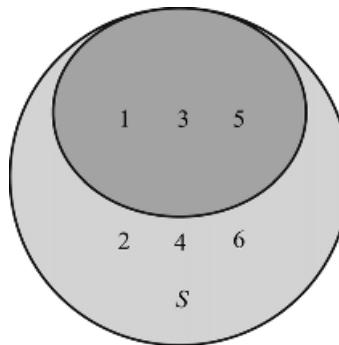


FIGURE 2.2.1

**Example 2.2.1** In the tossing of a fair die, we shall calculate the probability that 3 has occurred conditioned on the toss being odd. We shall define the following events:

$$A = \{\text{odd number}\}: B = 3$$

$P\{A\} = \frac{1}{2}$  and  $P\{B\} = \frac{1}{6}$  and the event  $A \cap B = \{3\}$  and hence

$$P\{3 \mid \text{odd}\} = \frac{P\{3\}}{P\{\text{odd}\}} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Note that  $P\{2 \mid \text{odd}\} = 0$  since 2 is an impossible event given that an odd number has occurred. Also note that  $P\{3 \mid 3\} = 1$ . The conditional probability of 3 occurring, given that 3 has occurred is obviously equal to 1.

We can also find the solution using the concept of reduced sample space. Since we are given that an odd number has occurred, we can reduce the sample space from  $S = \{1,2,3,4,5,6\}$  to  $S_1 = \{1,3,5\}$  as shown in Fig. 2.2.1.

The probability of 3 in this reduced sample space is  $\frac{1}{3}$ , agreeing with the previous result.

**Example 2.2.2** Consider a family of exactly two children. We will find the probabilities (1) that both are girls, (2) both are girls given that one of them is a girl, (3) both are girls given that the elder child is a girl, and (4) both are girls given that both are girls.

The sample space  $S$  has 4 points  $\{gg, gb, bg, bb\}$ , where  $g$  is girl and  $b$  is boy. The event  $B = \{\text{both children are girls}\}$ . The conditioning event  $A$  will depend on the following four cases:

1. The event  $A = S$ . Hence  $P\{B\} = \frac{1}{4}$ .
2. The event  $A = \{\text{one child is a girl}\}$ . The reduced sample space is  $S_1 = \{gg, gb, bg\}$  and the conditional probability  $P\{B \mid A\} = \frac{1}{3}$ .
3. The event  $A = \{\text{elder child is a girl}\}$ . In this case, the reduced sample space is  $S_2 = \{gg, gb\}$  and  $P\{B \mid A\} = \frac{1}{2}$ .
4. The event  $A = \{\text{both are girls}\}$ . The reduced sample space has only one point, namely,  $S_3 = \{gg\}$  and  $P\{B \mid A\} = 1$ .

As our knowledge of the experiment increases by the conditioning event, the uncertainty decreases and the probability increases from 0.25, through 0.333, 0.5, and finally to 1, meaning that there is no uncertainty.

**Example 2.2.3** The game of craps as played in Las Vegas has the following rules. A player rolls two dice. He wins on the first roll if he throws a 7 or a 11. He loses if the first throw is a 2, 3, or 12. If the first throw is a 4, 5, 6, 8, 9, or 10, it is called a *point* and the game continues. He goes on rolling until he throws the point for a win or a 7 for a loss. We have to find the probability of the player winning.

We will solve this problem both from the definition of conditional probability and the reduced sample space. Figure 2.2.2 shows the number of ways the sums of the pips on the two dice can equal 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and their probabilities.

*Solution Using Definition of Conditional Probability.* The probability of winning in the first throw is

$$P\{7\} + P\{11\} = \frac{6}{36} + \frac{2}{36} = \frac{8}{36} = 0.22222$$

The probability of losing in the first throw is

$$P\{2\} + P\{3\} + P\{12\} = \frac{1}{36} + \frac{2}{36} + \frac{1}{36} = \frac{4}{36} = 0.11111$$

To calculate the probability of winning in the second throw, we shall assume that  $i$  is the point with probability  $p$ . The probability of not winning in any given throw after the first is given by  $r = P\{\text{not } i \text{ and not 7}\} = 1 - p - \frac{1}{6}$ . We compute the conditional probability

$$P\{\text{win} \mid i \text{ in first throw}\} = p + rp + r^2p + \dots = \frac{p}{1-r} = \frac{p}{p+\frac{1}{6}}$$

as an infinite geometric series:

$$P\{\text{win in second or subsequent throws}\} = \frac{p^2}{p+\frac{1}{6}}$$

The probability  $p$  of making the point  $i = 4, 5, 6, 8, 9, 10$  is obtained from Fig. 2.2.2.

	Total	$P\{\text{total}\}$
Die A	2	$1/36$
Die B	3	$2/36$
1	4	$3/36$
2	5	$4/36$
3	6	$5/36$
4	7	$6/36$
5	8	$5/36$
6	9	$4/36$
	10	$3/36$
	11	$2/36$
	12	$1/36$

FIGURE 2.2.2

Thus, for  $i = 4,5,6$ , we obtain

$$P\{\text{win after point } i = 4\} = \left(\frac{3}{36}\right)^2 \Big/ \left(\frac{3}{36} + \frac{1}{6}\right) = \frac{1}{36}$$

$$P\{\text{win after point } i = 5\} = \left(\frac{4}{36}\right)^2 \Big/ \left(\frac{4}{36} + \frac{1}{6}\right) = \frac{2}{45}$$

$$P\{\text{win after point } i = 6\} = \left(\frac{5}{36}\right)^2 \Big/ \left(\frac{5}{36} + \frac{1}{6}\right) = \frac{25}{396}$$

with similar probabilities for 8,9,10. Thus the probability of winning in craps is

$$\begin{aligned} P\{\text{win}\} &= P\{\text{win in roll 1}\} + P\{\text{win in roll 2,3,}\dots\} \\ &= \frac{8}{36} + 2 \left[ \frac{1}{36} + \frac{2}{45} + \frac{25}{396} \right] = \frac{244}{495} = 0.4929 \end{aligned}$$

*Solution Using Reduced Sample Space.* We can now use the reduced sample space to arrive at the same result in a simpler manner. After the point  $i$  has been rolled in the first throw, the reduced sample space consists of  $(i - 1 + 6)$  points for  $i = 4,5,6$  and  $(13 - i + 6)$  points for  $i = 8,9,10$  in which the game will terminate. We can see from the figure that there are 6 ways of rolling 7 for a loss,  $(i - 1, i = 4,5,6)$  ways of rolling the point for a win, and  $(13 - i, i = 8,9,10)$  ways of rolling the point for a win. Thus

$$\begin{aligned} P\{\text{win in roll 2,3,}\dots \mid i \text{ in roll 1}\} &= P\{W \mid i\} \\ &= \begin{cases} \frac{i-1}{i-1+6}, & i = 4,5,6 \\ \frac{13-i}{13-i+6}, & i = 8,9,10 \end{cases} \end{aligned}$$

Thus,  $P\{W \mid 4\} = \frac{3}{9}$ ,  $P\{W \mid 5\} = \frac{4}{10}$ ,  $P\{W \mid 6\} = \frac{5}{11}$ ,  $P\{W \mid 8\} = \frac{5}{11}$ ,  $P\{W \mid 9\} = \frac{4}{10}$ ,  $P\{W \mid 10\} = \frac{3}{9}$ . Hence probability of a win after the first roll is  $P\{W \mid i\} \cdot P\{i\}$ . Hence

$$P\{\text{win after first roll}\} = 2 \left[ \frac{3}{9} \cdot \frac{3}{36} + \frac{4}{10} \cdot \frac{4}{36} + \frac{5}{11} \cdot \frac{5}{36} \right] = \frac{134}{495}$$

$$P\{\text{win}\} = 2 \left[ \frac{3}{9} \cdot \frac{3}{36} + \frac{4}{10} \cdot \frac{4}{36} + \frac{5}{11} \cdot \frac{5}{36} \right] = \frac{8}{36} + \frac{134}{495} + \frac{244}{495} = 0.4929$$

a result obtained with comparative ease.

**Example 2.2.4** We shall continue Example 2.1.8 and find the probability that a person will die between the ages 70 and 80 given that this individual has lived upto the age 70. We will define the events  $A$  and  $B$  as follows:

$$A = \{\text{person has lived upto 70}\}$$

$$B = \{\text{person dies between ages 70 and 80}\}$$

The required conditional probability equals the number of persons who die between the ages 70 and 80 divided by the number of persons who have lived upto the age 70:

$$\begin{aligned} P\{B|A\} &= \frac{P\{70 < t \leq 80 \text{ and } t \geq 70\}}{P\{t \geq 70\}} = \frac{P\{70 < t \leq 80\}}{P\{t \geq 70\}} \\ &= \frac{\int_{70}^{80} 3 \times 10^{-9} t^2 (100 - t^2) dt}{\int_{70}^{100} 3 \times 10^{-9} t^2 (100 - t^2) dt} = \frac{0.105}{0.163} = 0.644 \end{aligned}$$

This type of analysis is useful in setting up actuarial tables for life insurance purposes.

## 2.3 INDEPENDENCE

We have already seen the concept of *functional independence* of sets in Chapter 1. We will now define *statistical independence* of events. If  $A$  and  $B$  are two events in a field of events  $\mathcal{F}$ , then  $A$  and  $B$  are statistically independent if and only if

$$P\{A \cap B\} = P\{A\}P\{B\} \quad (2.3.1)$$

Statistical independence neither implies nor is implied by functional independence. Unless otherwise indicated, we will call statistically independent events “independent” without any qualifiers.

Three events  $A_1, A_2, A_3$  are independent if they satisfy, in addition to

$$P\{A_1 \cap A_2 \cap A_3\} = P\{A_1\}P\{A_2\}P\{A_3\} \quad (2.3.2)$$

the pairwise independence

$$P\{A_i \cap A_j\} = P\{A_i\}P\{A_j\}, \quad i, j = 1, 2, 3, \quad i \neq j \quad (2.3.3)$$

Note that if  $A$  and  $B$  are mutually exclusive, then, from Kolmogorov’s axiom 3a, we obtain

$$P\{A \cup B\} = P\{A\} + P\{B\} \quad (2.3.4)$$

The concepts of independence and mutual exclusivity are fundamental to probability theory. For any two sets  $A$  and  $B$ , we recall from Eq. (2.1.7) that

$$P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\} \quad (2.1.7)$$

If  $A$  and  $B$  are mutually exclusive, then  $P\{A \cap B\} = 0$ , but if they are independent, then  $P\{A \cap B\} = P\{A\} \cdot P\{B\}$ . If this has to be zero for mutual exclusivity, then one of the probabilities must be zero. Therefore, we conclude that mutual exclusivity and independence are incompatible unless we are dealing with null events.

We shall now give examples to clarify the concepts of functional independence and statistical independence.

**Example 2.3.1** The probability of an ace on the throw of a die is  $\frac{1}{6}$ . The probability of a head on the throw of a coin is  $\frac{1}{2}$ . Since the throw of a die is not dependent on the throw of a coin, we have functional independence and the combined sample space is the Cartesian product with 12 elements as shown in Table 2.3.1.

TABLE 2.3.1

1, H	2, H	3, H	4, H	5, H	6, H
1, T	2, T	3, T	4, T	5, T	6, T

We assume as usual that all the 12 elements are equiprobable events and the probability of ace and head is  $\frac{1}{12}$  since there is only one occurrence of {1, H} among the 12 points as seen in the first row and first column in Table 2.3.1. Using Eq. (2.3.1), we have  $P\{1, H\} = P\{1\} \cdot P\{H\} = \frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$ , illustrating that in this example, functional independence coincides with statistical independence.

**Example 2.3.2** Two dice, one red and the other blue, are tossed. These tosses are functionally independent, and we have the Cartesian product of  $6 \times 6 = 36$  elementary events in the combined sample space, where each event is equiprobable. We seek the probability that an event  $B$  defined by the sum of the numbers showing on the dice equals 9. There are four points {(6,3), (5,4), (4,5), (3,6)}, and hence  $P\{B\} = \frac{4}{36} = \frac{1}{9}$ . We now condition the event  $B$  with an event  $A$  defined as the red die shows odd numbers. The probability of the event  $A$  is  $P\{A\} = \frac{18}{36} = \frac{1}{2}$ . We want to determine whether the events  $A$  and  $B$  are statistically independent. These events are shown in Fig. 2.3.1, where the first number is for the red die and the second number is for the blue die.

From Fig. 2.3.1  $P\{A \cap B\} = P\{(3,6), (5,4)\} = \frac{2}{36} = \frac{1}{18}$  and  $= P\{A\} \times P\{B\} = \frac{1}{2} \times \frac{1}{9} = \frac{1}{18}$  showing statistical independence. We compute  $P\{B | A\}$  from the point of view of reduced sample space. The conditioning event reduces the sample space from 36 points to 18 equiprobable points and the event  $\{B | A\} = \{(3,6), (5,4)\}$ . Hence  $P\{B | A\} = \frac{2}{18} = \frac{1}{9} = P\{B\}$ , or, the conditioning event has no influence on  $B$ . Here, even though the events  $A$  and  $B$  are functionally dependent, they are statistically independent.

However, if another set  $C$  is defined by the sum being equal to 8 as shown in Fig. 2.3.2, then  $P\{C\} = \frac{5}{36}$ .

Here the events  $C$  and  $A$  are not statistically independent because  $P\{C\} \cdot P\{A\} = \frac{5}{36} \times \frac{1}{2} = \frac{5}{72} \neq P\{C \cap A\} = P\{(3,5), (5,3)\} = \frac{2}{18} = \frac{4}{72}$ .

In this example, we have the case where the events  $A$  and  $C$  are neither statistically independent nor functionally independent.

1, 1	1, 2	1, 3	1, 4	1, 5	1, 6
2, 1	2, 2	2, 3	2, 4	2, 5	2, 6
3, 1	3, 2	3, 3	3, 4	3, 5	3, 6
4, 1	4, 2	4, 3	4, 4	4, 5	4, 6
5, 1	5, 2	5, 3	5, 4	5, 5	5, 6
6, 1	6, 2	6, 3	6, 4	6, 5	6, 6

FIGURE 2.3.1

1, 1	1, 2	1, 3	1, 4	1, 5	1, 6
2, 1	2, 2	2, 3	2, 4	2, 5	2, 6
3, 1	3, 2	3, 3	3, 4	3, 5	3, 6
4, 1	4, 2	4, 3	4, 4	4, 5	4, 6
5, 1	5, 2	5, 3	5, 4	5, 5	5, 6
6, 1	6, 2	6, 3	6, 4	6, 5	6, 6

FIGURE 2.3.2

## 2.4 TOTAL PROBABILITY AND BAYES' THEOREM

Let  $\{A_i, i = 1, \dots, n\}$  be a partition of the sample space and let  $B$  an arbitrary event in the sample space  $S$  as shown in Fig. 2.4.1.

We will determine the *total probability*  $P\{B\}$ , given the conditional probabilities  $P\{B | A_1\}, P\{B | A_2\}, \dots, P\{B | A_n\}$ . The event  $B$  can be given in terms of the partition as

$$\begin{aligned} B &= \{B \cap S\} = B \cap \{A_1 \cup A_2 \cup \dots \cup A_n\} \\ &= \{B \cap A_1\} \cup \{B \cap A_2\} \cup \dots \cup \{B \cap A_n\} \end{aligned} \quad (2.4.1)$$

The events  $\{B \cap A_1\}, \{B \cap A_2\}, \dots, \{B \cap A_n\}$  in Eq. (2.4.1) are mutually exclusive since  $A_1, A_2, \dots, A_n$  are mutually exclusive and we have from Kolmogorov's axiom 3a:

$$P\{B\} = P\{B \cap A_1\} + P\{B \cap A_2\} + \dots + P\{B \cap A_n\} \quad (2.4.2)$$

Using the definition of conditional probability from Eq. (2.2.1), we can rewrite Eq. (2.4.2) as

$$\begin{aligned} P\{B\} &= P\{B | A_1\}P\{A_1\} + P\{B | A_2\}P\{A_2\} + \dots + P\{B | A_n\}P\{A_n\} \\ &= \sum_{i=1}^n P\{B | A_i\}P\{A_i\} \end{aligned} \quad (2.4.3)$$

Equation (2.4.3) is the expression for the total probability  $P\{B\}$  in terms of the conditional probabilities  $P\{B | A_i, i = 1, \dots, n\}$ . We shall give an example to illustrate this concept.

$S$	$A_1$	$A_2$	$A_3$
	$A_4$		
	$A_7$	$B$	
	$A_{10}$		
	$A_{13}$		$A_n$

FIGURE 2.4.1

**Example 2.4.1** Four computer companies  $A, B, C, D$  supply microprocessors to a company. From previous experience it is known that the probability of the processor being bad if it comes from  $A$  is 4%,  $B$  is 2%,  $C$  is 5%, and  $D$  is 1%. The probabilities of picking supplier  $A$  is 20%,  $B$  is 30%,  $C$  is 10%, and  $D$  are 40%. We want to find the probability that a processor chosen at random is bad.

Let us define  $X$  as the event that the processor is bad. We need to find  $P\{X\}$  from the conditional probabilities  $P\{X | A\} = 0.04$ ,  $P\{X | B\} = 0.02$ ,  $P\{X | C\} = 0.05$ , and  $P\{X | D\} = 0.01$ . We are also given that  $P\{A\} = 0.2$ ,  $P\{B\} = 0.3$ ,  $P\{C\} = 0.1$ , and  $P\{D\} = 0.4$ . Substituting these quantities in Eq. (2.4.3), we obtain

$$\begin{aligned} P\{X\} &= P\{X | A\}P\{A\} + P\{X | B\}P\{B\} + P\{X | C\}P\{C\} + P\{X | D\}P\{D\} \\ &= 0.04 \times 0.2 + 0.02 \times 0.3 + 0.05 \times 0.1 + 0.01 \times 0.4 = 0.023 \end{aligned}$$

Thus the probability that the processor is bad is 2.3%.

### Bayes' Theorem

The next logical question to ask is, “Given that the event  $B$  has occurred, what are the probabilities that  $A_1, A_2, \dots, A_n$  are involved in it?” In other words, given the conditional probabilities  $P\{B | A_1\}, P\{B | A_2\}, \dots, P\{B | A_n\}$ , we have to find the reverse conditional probabilities  $P\{A_1 | B\}, P\{A_2 | B\}, \dots, P\{A_n | B\}$ . Using the total probability  $P\{B\}$  from Eq. (2.4.3), we can write

$$P\{A_i | B\} = \frac{P\{A_i B\}}{P\{B\}} = \frac{P\{B | A_i\}P\{A_i\}}{P\{B\}} = \frac{P\{B | A_i\}P\{A_i\}}{\sum_{j=1}^n P\{B | A_j\}P\{A_j\}} \quad (2.4.4)$$

This is the celebrated Bayes' theorem. This theorem connects the a priori probability  $P\{A_i\}$  to the a posteriori probability  $P\{A_i | B\}$ . This useful theorem has wide applicability in communications and medical imaging, as we will see in Chapter 23.

**Example 2.4.2** We will now continue Example 2.4.1 and find the probability that the processor came from companies  $A, B, C$ , or  $D$  given that the processor is bad. Using the result from Example 2.4.1 that  $P\{X\} = 0.023$ , we can write from Eq. (2.4.4)

$$P\{X | A\} = \frac{P\{X | A\}P\{A\}}{P\{X\}} = \frac{0.04 \times 0.2}{0.023} = 0.348$$

$$P\{X | B\} = \frac{P\{X | B\}P\{B\}}{P\{X\}} = \frac{0.02 \times 0.3}{0.023} = 0.261$$

$$P\{X | C\} = \frac{P\{X | C\}P\{C\}}{P\{X\}} = \frac{0.05 \times 0.1}{0.023} = 0.217$$

$$P\{X | D\} = \frac{P\{X | D\}P\{D\}}{P\{X\}} = \frac{0.01 \times 0.4}{0.023} = 0.174$$

Note that the sum of the resulting probabilities, namely,  $0.348 + 0.261 + 0.217 + 0.174 = 1$ , as it should be because the bad processor has to come only from any one of the four companies  $A, B, C$ , or  $D$  resulting in the probability being 1.

**Example 2.4.3** A communication receiver sends numbers 1,2,3. As a result of inherent noise in the channel, the conditional probabilities of receiving 1,2,3 under the condition that 1,2,3 were sent are shown in Fig. 2.4.2.

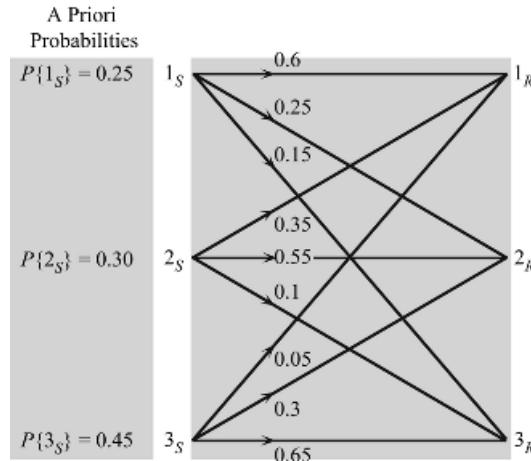


FIGURE 2.4.2

*Transition Probabilities:*

$P\{1_R   1_S\} = 0.60$	$P\{2_R   1_S\} = 0.25$	$P\{3_R   1_S\} = 0.15$
$P\{1_R   2_S\} = 0.35$	$P\{2_R   2_S\} = 0.55$	$P\{3_R   2_S\} = 0.10$
$P\{1_R   3_S\} = 0.05$	$P\{2_R   3_S\} = 0.30$	$P\{3_R   3_S\} = 0.65$

The total probability that 1 was received is obtained from Eq. (2.4.3):

$$\begin{aligned} P\{1_R\} &= P\{1_R | 1_S\}P\{1_S\} + P\{1_R | 2_S\}P\{2_S\} + P\{1_R | 3_S\}P\{3_S\} \\ &= 0.6 \times 0.25 + 0.35 \times 0.3 + 0.05 \times 0.45 = 0.2775 \end{aligned}$$

From Bayes' theorem, Eq. (2.4.4), the a posteriori probabilities of 1, 2, and 3 being sent conditioned on 1 received are found as follows:

$$\begin{aligned} P\{1_S | 1_R\} &= \frac{P\{1_R | 1_S\}P\{1_S\}}{P\{1_R\}} = \frac{0.6 \times 0.25}{0.2775} = 0.5405: \quad P\{1_S\} = 0.25 \\ P\{2_S | 1_R\} &= \frac{P\{1_R | 2_S\}P\{2_S\}}{P\{1_R\}} = \frac{0.35 \times 0.3}{0.2775} = 0.3784: \quad P\{2_S\} = 0.3 \\ P\{3_S | 1_R\} &= \frac{P\{1_R | 3_S\}P\{3_S\}}{P\{1_R\}} = \frac{0.05 \times 0.45}{0.2775} = 0.0811: \quad P\{3_S\} = 0.45 \end{aligned}$$

The sum of these probabilities is equal to 1, as it should be. Note that the a priori probability that 3 was sent is 0.45 and after 1 has been received, the a posteriori probability is 0.0811, a very small probability. Similar posterior probabilities of 1, 2, and 3 being sent conditioned on 2 and 3 received can be found.

**Example 2.4.4 (Monty Hall Problem)** This classic problem is called the “Monty Hall problem” because of the game show host Monty Hall who designed this game. There are three doors A, B, and C, and behind one of them is a car and behind the other two are goats. The contestant is asked to select any door, and the host Monty Hall opens one of the other two doors and reveals a goat. He then offers the choice to the contestant of switching to the other unopened door or keep the original door. The question now is whether the

probability of winning the car is improved if she switches or it is immaterial whether she switches or not.

The answer is counterintuitive in the sense that one may be misled into thinking that it is immaterial whether one switches or not. We will analyze this problem in two different ways.

*Mathematical Analysis.* Let us analyze this problem from the point of view of Bayes' theorem. We shall first assume that the host has prior knowledge of the door behind which the car is located. Let us also assume that the host opens door  $B$  given that the contestant's choice is door  $A$ . The a priori probabilities of a car behind the doors  $A$ ,  $B$ , and  $C$  are

$$P\{A\} = \frac{1}{3}; \quad P\{B\} = \frac{1}{3}; \quad P\{C\} = \frac{1}{3}$$

We can make the following observations. If the contestant's choice is door  $A$  and the car is behind door  $A$ , then the host will open the door  $B$  with probability of  $\frac{1}{2}$  (since she has a choice between  $B$  and  $C$ ). On the other hand, if the car is behind door  $B$ , then there is zero probability that she will open door  $B$  because of her prior knowledge. If the car is behind door  $C$ , then she will open door  $B$  with probability 1. Thus we can write the following conditional probabilities for the host opening door  $B$ :

$$P\{B | A\} = \frac{1}{2}; \quad P\{B | B\} = 0; \quad P\{B | C\} = 1$$

We can now calculate the total probability  $P(B)$  of the host opening door  $B$ .

$$\begin{aligned} P\{B\} &= P\{B | A\}P\{A\} + P\{B | B\}P\{B\} + P\{B | C\}P\{C\} \\ &= \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{1}{2} \end{aligned}$$

Using Bayes' theorem [Eq. (2.4.4)] we can now find the a posteriori probabilities of the car behind the doors  $A$  or  $C$  ( $B$  has already been opened by the host) conditioned on  $B$ :

$$P\{A | B\} = \frac{P\{BA\}}{P\{B\}} = \frac{P\{B | A\}P\{A\}}{P\{B\}} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$$

$$P\{C | B\} = \frac{P\{BC\}}{P\{B\}} = \frac{P\{B | C\}P\{C\}}{P\{B\}} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Similar analysis holds for other cases of opening the doors  $A$  or  $C$ . We can see from the result above that the contestant *must* switch if she wants to double her probability of winning the car.

TABLE 2.4.1

First Choice	Host Opens	Second Choice	Probability
$A$	$B$	$A$	$P\{B\} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
$A$	$C$	$A$	$P\{C\} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
$B$	$C$	$B$	$P\{C\} = 1 \cdot \frac{1}{3} = \frac{1}{3}$
$C$	$B$	$B$	$P\{B\} = 1 \cdot \frac{1}{3} = \frac{1}{3}$

**TABLE 2.4.2 For not switching**

First Choice	Host Opens	Second Choice	Win or Loss	Probability
A	B	A	Win	$P\{B\} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
A	C	A	Win	$P\{C\} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
B	C	B	Loss	$P\{C\} = 1 \cdot \frac{1}{3} = \frac{1}{3}$
C	B	B	Loss	$P\{B\} = 1 \cdot \frac{1}{3} = \frac{1}{3}$

**TABLE 2.4.3 For switching**

First Choice	Host Opens	Second Choice	Win or Loss	Probability
A	B	C	Loss	$P\{B\} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
A	C	B	Loss	$P\{C\} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$
B	C	A	Win	$P\{C\} = 1 \cdot \frac{1}{3} = \frac{1}{3}$
C	B	A	Win	$P\{B\} = 1 \cdot \frac{1}{3} = \frac{1}{3}$

*Heuristic Analysis.* In this case we shall enumerate all the possibilities and determine which possibility gives a better chance. Assume that the car is behind door A. The probabilities of  $A, B, C$  are once again

$$P\{A\} = \frac{1}{3}; \quad P\{B\} = \frac{1}{3}; \quad P\{C\} = \frac{1}{3}$$

Conditioned on the contestant choosing door A, the host has two equiprobable choices of opening either door B or door C. However, conditioned on the contestant choosing B or C, the host has only one choice of either opening C or B. Thus

$$P\{B|A\} = \frac{1}{2}; \quad P\{C|A\} = \frac{1}{2}; \quad P\{C|B\} = 1; \quad P\{B|C\} = 1$$

From these conditional probabilities we can evaluate the unconditional probabilities as shown in Table 2.4.1.

We can now enumerate the various win or loss possibilities for not switching and for switching in Tables 2.4.2 and 2.4.3. From the Table 2.4.2 we can see that the probability of winning the car by not switching is  $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$ .

From Table 2.4.3 we can see that the probability of winning the car by switching is  $\frac{1}{3} + \frac{1}{3} = \frac{2}{3}$ . Thus the probability of winning the car doubles with switching.

# Basic Combinatorics

## 3.1 BASIC COUNTING PRINCIPLES

Combinatorics is the study of arrangement and manipulation of mathematical elements in sets and essentially involves counting. We will discuss two basic counting principles.

### ***The Addition Rule***

A task  $A$  consists of  $m$  subtasks  $A_1, A_2, \dots, A_m$  that can be done in  $n(A_1), n(A_2), \dots, n(A_m)$  ways, respectively. If no two of them can be done at the same time, then the total number of ways  $n(A)$  of doing the task  $A$  is

$$n(A) = n(A_1) + n(A_2) + \dots + n(A_m) \quad (3.1.1)$$

This rule is the same as in set theory [Eq. (1.2.5a)] where the cardinality of the union of mutually exclusive sets is the sum of the cardinalities of the individual sets. Thus, the addition rule is a restatement of one of the properties of mutually exclusive sets.

**Example 3.1.1** An engineering student undertaking a capstone project  $T$  has a choice of 4 professors  $A, B, C, D$  who can give him a list of projects from which he can choose one.  $A$  has 5 projects,  $B$  has 7 projects,  $C$  has 6 projects, and  $D$  has 2 projects. Since the student can not take any other project once he has chosen a project, the number of choices he has is

$$n(T) = n(A) + n(B) + n(C) + n(D) = 5 + 7 + 6 + 2 = 20$$

### ***The Multiplication Rule***

A task  $A$  consists of  $m$  sequential subtasks  $A_1, A_2, \dots, A_m$  that can be performed only after the previous one is completed. The numbers of ways  $n(A_1), n(A_2), \dots, n(A_m)$  of performing the subtasks are functionally independent; that is,  $n(A_2)$  ways are needed for the second step regardless of  $n(A_1)$  of the first step. The number of ways  $n(A)$  of performing the task  $A$  is

$$n(A) = n(A_1) \times n(A_2) \times \dots \times n(A_m) \quad (3.1.2)$$

This is very similar to the Cartesian product in set theory, where the cardinality of the product is the product of the cardinalities of the individual sets. These sets are also functionally independent. They may or may not be statistically independent.

**Example 3.1.2** A professor assigns a capstone project  $T$  to senior engineering students consisting of four parts  $A, B, C, D$  that must be done in sequence independent of the previous ones. Part  $A$  can be done in 2 ways, part  $B$  in 3 ways, part  $C$  in 5 ways, and part  $D$  in 4 ways. Since  $A, B, C, D$  are functionally independent, the project can be done in

$$n(T) = n(A) \times n(B) \times n(C) \times n(D) = 2 \times 3 \times 5 \times 4 = 120 \text{ ways}$$

## 3.2 PERMUTATIONS

Many problems in combinatorics can be modeled as placing  $r$  balls into  $n$  cells. The two ways of considering permutations are (1) sampling with replacement and with ordering and (2) sampling without replacement and with ordering.

### ***Sampling with Replacement and with Ordering***

We consider placing  $r$  balls into  $n$  cells with replacement and with ordering. The first ball can be placed in  $n$  cells in  $n$  ways. The second ball can be placed also in  $n$  ways. Since the placing of the second ball is functionally independent of that of the first ball, there are now  $n^2$  ordered 2-permutations. Thus, for placing  $r$  balls into  $n$  cells, there will be  $n^r$  ways of placing  $r$  balls into  $n$  cells with replacement and ordering:

$$\text{Number of distinct ordered } r\text{-tuples} = n^r \quad (3.2.1)$$

The probability of each one of these  $r$ -tuples assuming that they are equally likely is  $n^{-r}$ . Clearly we can place 100 balls into one cell or 1 ball in 100 cells. Hence in Eq. (3.2.1) it does not matter whether  $r > n$ ,  $r = n$ , or  $r < n$ .

**Example 3.2.1** The number of distinct ordered ways we can draw 2 balls from an urn containing 4 balls numbered from 1 to 4 is obtained as follows. The cells correspond to  $n = 4$ , the total number of balls. The balls correspond to  $r = 2$ . The number of ways 2 balls can be placed into 4 cells is, from Eq. (3.2.1),  $4^2 = 16$  ways. These 16 ordered ways are shown in Table 3.2.1 for clarity.

### ***Sampling without Replacement and with Ordering***

We now consider placing  $r$  balls into  $n$  cells without replacement and with ordering. This situation is different from the previous case. The first ball can be placed in  $n$  cells in

**TABLE 3.2.1**

1,1	1,2	1,3	1,4	2,1	2,2	2,3	2,4
3,1	3,2	3,3	3,4	4,1	4,2	4,3	4,4

$n$  different ways. Since one of the cells is occupied, the second ball can be placed only in  $(n - 1)$  ways. The  $r$ th ball can be placed in  $(n - r + 1)$  ways. Since placements of the balls are again functionally independent, we have

$$\text{Number of distinct } r\text{-tuples} = n(n - 1) \cdots (n - r + 1) \quad (3.2.2)$$

These will be called *r-permutations* and will be defined by  $(n)_r$  or  $nP_r$ . Using the factorial notation, we can write

$$(n)_r = \frac{n!}{(n - r)!} \quad (3.2.3)$$

We cannot place 5 balls into 2 cells without replacement because at least some cells will have more than 1 ball, violating the nonreplacement condition. Hence we define  $(n)_r = 0$  whenever  $r > n$ . If  $r = n$ , then  $(n)_r = (n)_n = n!$ .

**Example 3.2.2** We continue Example 3.2.1 and now draw 2 balls from 4 balls without replacement and with ordering. The number of ways this can be done is obtained by applying Eq. (3.2.3) with  $n = 4$  and  $r = 2$  which is equal to  $4 \cdot 3 = 12$  ways instead of the 16 ways that we obtained in Table 3.2.1. The 12 ways are shown in Table 3.2.2.

In Table 3.2.2, the arrangement with ordering, (1,2), (1,3), (1,4), (2,3), (2,4), (3,4) represents an ordering different from (2,1), (3,1), (4,1), (3,2), (4,2), (4,3).

We can now summarize the preceding results as follows. The number of different ways in which  $r$  balls can be placed into  $n$  cells with ordering and replacement is  $n^r$ , and with ordering and without replacement is  $(n)_r$ .

**Example 3.2.3** In the Olympic 100-m dash, there are 9 runners. In how many ways can the gold, silver, and bronze can be given to these runners?

Here the 9 runners are the cells  $n$  and the 3 medals are the balls  $r$ . Hence, from Eq. (3.2.3), the number of ways that the medals can be distributed is  $(9)_3 = 9 \times 8 \times 7 = 504$  ways.

**Example 3.2.4 (Birthdays Problem)** We want to determine the probability that in a class size of  $r$  two or more birth dates match. We shall assume that the year consists of 365 days and the year of birth is inconsequential.

Here the 365 days in a year correspond to  $n$  cells and the class size, to  $r$ . If the class size is over 365, then two or more birth dates will match with probability 1. The number of ways  $r$  students can be placed into 365 birth dates with replacement is  $365^r$ . The number of ways they can be placed without replacement is  $(365)_r$ , meaning that all the birth dates are different. Thus, the probability of no two birth dates match is

$$P\{\text{no birthdays match}\} = \frac{(365)_r}{365^r}$$

TABLE 3.2.2

1,2	1,3	1,4	2,1	2,3	2,4
3,1	3,2	3,4	4,1	4,2	4,3

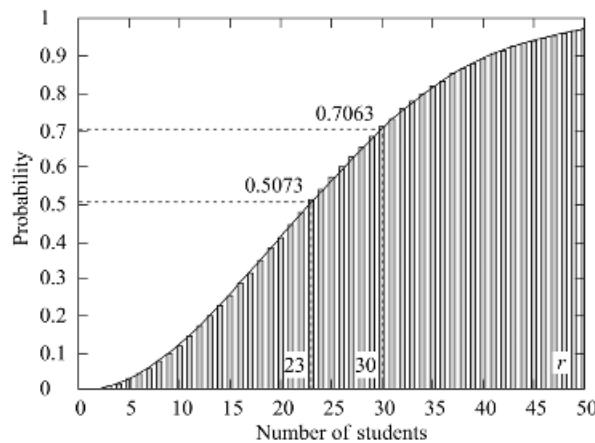


FIGURE 3.2.1

TABLE 3.2.3

#	Probability	#	Probability	#	Probability
20	0.411438	30	0.706316	44	0.932885
21	0.443668	31	0.730455	45	0.940976
22	0.475695	32	0.753348	46	0.948253
23	0.507297	33	0.774972	46	0.954774
24	0.538344	34	0.795317	48	0.960598
25	0.5687	35	0.814383	49	0.96578
26	0.598241	36	0.832182	50	0.970374

The probability  $P(r)$  that two or more birth dates match is the complement of the preceding equation and is given by

$$P\{r\} = 1 - \frac{(365)_r}{365^r}$$

$P(r)$  is graphed for  $r = 0-50$  in Fig. 3.2.1 and tabulated for  $r = 20-26$ ,  $30-36$ , and  $44-50$  in Table 3.2.3.

Note that for two or more birthdays to match, there is about 50% chance for 23 students, 71% chance for 30 students, and 95% chance for 46 students. In other words, we need only 46 students for 2 or more birthdays to match with 95% probability, but to get to 100% probability, we need to go to 366 students!

### 3.3 COMBINATIONS

Similar to the case of permutations we have two ways of considering combinations:

1. Sampling without replacement and without ordering
2. Sampling with replacement and without ordering

### Sampling without Replacement and without Ordering

We now discuss the case where we want to obtain a sample of  $r$  balls from  $n$  balls without replacement and without ordering. We have already seen that the number of ways to get  $r$  balls from  $n$  balls with ordering is  $(n)_r$ . If we disregard ordering, then the number of ways is called the  $r$ -combination from the  $n$  balls. We can define this quantity as

$$\{r\text{-combination from } n \text{ balls}\} = \binom{n}{r} \quad (3.3.1)$$

By ordering the elements in each  $r$ -combination, we can obtain from  $\binom{n}{r}$  the  $r$ -permutations  $(n)_r$ . This can be done in  $r!$  ways. Hence we can determine  $\binom{n}{r}$  from either of the following formulas:

$$\begin{aligned} \binom{n}{r} r! &= (n)_r \\ \binom{n}{r} &= \frac{(n)_r}{r!} = \frac{n!}{r!(n-r)!} \end{aligned} \quad (3.3.2)$$

The quantity  $\binom{n}{r}$  is also called the *binomial coefficient* since it occurs as the coefficient of the binomial expansion, given by

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \cdots + \binom{n}{r}a^{n-r}b^r + \cdots + \binom{n}{n}b^n$$

### Properties of Binomial Coefficients

$$1. \quad \binom{n}{r} = \binom{n}{n-r} \quad (3.3.3)$$

Expanding the righthand side of Eq. (3.3.3), we obtain

$$\binom{n}{n-r} = \frac{n!}{(n-r)![n-(n-r)]!} = \frac{n!}{(n-r)!r!} = \binom{n}{r}$$

### 2. Pascal's Identity:

$$\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r} \quad (3.3.4)$$

Expanding the righthand side of Eq. (3.3.4), we obtain

$$\begin{aligned} \binom{n}{r-1} + \binom{n}{r} &= \frac{n!}{(r-1)!(n-r+1)!} + \frac{n!}{r!(n-r)!} \\ &= \frac{n!}{r!(n-r)!} \left[ \frac{r}{n-r+1} + 1 \right] \\ &= \frac{n!}{r!(n-r)!} \left[ \frac{n+1}{n-r+1} \right] = \frac{(n+1)!}{r!(n-r+1)!} = \binom{n+1}{r} \end{aligned}$$

Pascal's identity shows that if two adjacent binomial coefficients are added, the binomial coefficient in the next row in between these two coefficients results, as shown in Fig. 3.3.1. It is called the *Pascal triangle*.

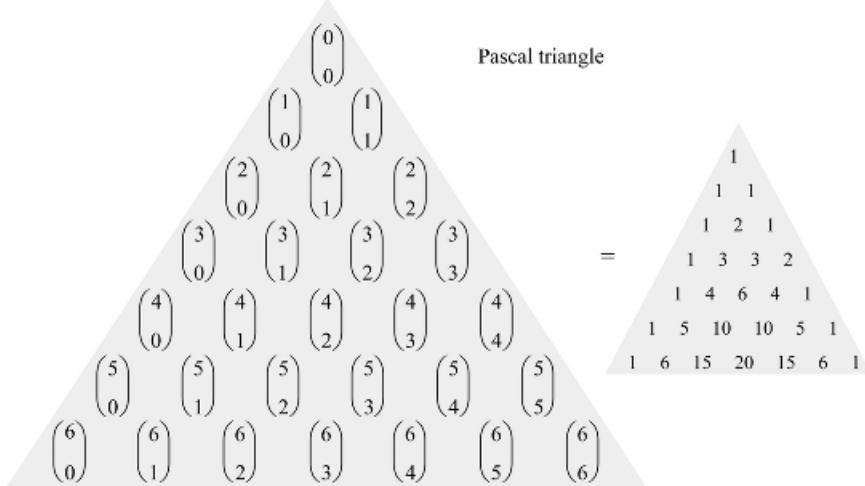


FIGURE 3.3.1

**3. Sum of Binomial Coefficients:**

$$\sum_{r=0}^n \binom{n}{r} = 2^n \quad (3.3.5)$$

Substituting in the binomial expansion

$$(1+x)^n = 1 + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n$$

$x = 1$ , we obtain Eq. (3.3.5).

**4. Vandermonde's Identity:**

$$\binom{n+m}{r} = \sum_{k=0}^r \binom{n}{k} \binom{m}{r-k} \quad (3.3.6)$$

Let us assume that there are two urns with  $n$  balls in the first and  $m$  balls in the second. The total number of ways of picking  $r$  balls from both of these urns is  $\binom{n+m}{r}$ . We can also choose these  $r$  balls by first picking  $k$  balls from the first urn and  $(r-k)$  balls from the second urn. Since the choice is functionally independent, the number of ways of accomplishing this is  $\binom{n}{k} \binom{m}{r-k}$ . The realizations for every  $0 \leq k \leq r$  are mutually exclusive, and hence these numbers add, resulting in Eq. (3.3.6). This is shown in Fig. 3.3.2.

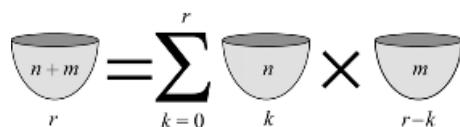


FIGURE 3.3.2

**Example 3.3.1** We continue Example 3.2.2 with sampling 2 balls from 4 balls without replacement and without ordering. Thus means that there is no distinction between (12) and (21), (23) and (32), (34), and (43). Thus we have only 6 ways and this is obtained by applying Eq. (3.3.1) with  $n = 4$  and  $r = 2$ , yielding  $(4 \times 3)/(2 \times 1) = 6$ . These are shown in Table 3.3.1.

TABLE 3.3.1

1,2	1,3	1,4
2,3	2,4	3,4

**Example 3.3.2**

1. A committee of 3 women and 4 men must be formed from a group of 6 women and 8 men. The number of ways 3 women can be chosen from 6 women is  $\binom{6}{3}$  from Eq. (3.2.2) and the number of ways 4 men can be chosen from 8 men is  $\binom{8}{4}$ . Since these choices are functionally independent, the total number of ways the committee can be formed is  $\binom{6}{3} \cdot \binom{8}{4} = 1400$ .
2. We modify the problem above by having 2 particular men be always on the committee. In this case we have a choice of only 2 men from the 6 men after taking out the 2 required men. Thus, there are  $\binom{6}{3} \cdot \binom{6}{2} = 300$  ways.
3. We again modify the problem by requiring that two particular women should not be on the committee. In this case, we calculate the number of ways that the two women will be together and subtract it from the total number of ways. There are  $\binom{5}{1} \cdot \binom{8}{4} = 350$  ways that the two women will be together, and hence there are  $\binom{6}{3} \cdot \binom{8}{4} - \binom{5}{1} \cdot \binom{8}{4} = 1400 - 350 = 1050$  ways in which two given women will not be together.

**Example 3.3.3 (Catalan Numbers)** There are  $n$  objects to be placed in computer stacks in the following way. These objects are first stored and deleted later with the result that when all the  $n$  objects are deleted, the stack is empty. Hence, the operation store ( $S$ ) has to precede the operation delete ( $D$ ), and at any point along the stack the number of stores should always be more than or equal to the number of deletes. Finally, the stack should begin with store and end with delete. We need to find the number  $C(n)$  of operations required for  $n$  objects. For example, if there are  $n = 3$  objects, the number of ways of ordering the store and delete operations are  $2 \times 3 = 6$  given by  $SSSSDD$ ,  $SDSDSD$ ,  $SSDDSD$ ,  $SDSSDD$ ,  $SSDSDD$ ,  $SDSDSSD$ . Thus, the operation starts with an  $S$  and ends in a  $D$ , and at no point along the chain can the number of  $D$  operations exceed the  $S$  operations. We will show that the number  $C(n)$  is given by

$$C_{(n)} = \frac{1}{n+1} \cdot \frac{(2n)!}{n!n!}$$

Since there are a total of  $2n$  operations, the total number of ways that we can obtain an  $n$ -combination is  $\binom{2n}{n}$ . Out of these combinations there will be many realizations in which the number of  $D$ s will exceed the number of  $S$ s. We will now find the number of ways in which the  $D$ s will exceed the  $S$ s. If we are at the point  $(n+1)$

TABLE 3.3.2

#	Stacks	#	Stacks	#	Stacks
1	1	8	1430	15	9694845
2	2	9	4862	16	35357670
3	5	10	16796	17	129644790
4	14	11	58786	18	477638700
5	42	12	208012	19	1767263190
6	132	13	742900	20	6564120420
7	429	14	2674440		

along the stack, the number of ways that  $D$  will exceed  $S$  will be  $\binom{2n}{n+1} - \binom{2n}{n-1}$ . Hence the required number of stacks is  $\binom{2n}{n} - \binom{2n}{n-1} = \frac{1}{n+1} \binom{2n}{n}$ , which is the desired result. The Catalan numbers are tabulated in Table 3.3.2 for  $n = 1$  to 20.

### ***Sampling with Replacement and without Ordering***

We shall now consider forming  $r$ -combinations with replacement and without ordering. In this case we pick  $r$  balls with replacement from balls in  $n$  distinct categories such as red, blue, green, and brown. We do not distinguish among balls of the same color, and the distinction is among different colors. For example, if we are picking 6 balls from the 4 color categories, then it can be arranged as shown in Figure 3.3.3.

There are now 2 red balls, 1 blue ball, no green balls, and 3 brown balls. The bar symbol | separates the four categories. The arrangement cannot start with a bar or end with a bar; hence we have  $4 - 1 = 3$  bars. We now have to choose 6 balls from 6 balls and 3 bars without replacement and without ordering. There are then

$$\binom{9}{6} = \binom{9}{3} = \frac{9 \times 8 \times 7}{3 \times 2 \times 1} = 84$$

ways of picking 6 balls from 4 different categories.

Generalizing this result, if there are  $n$  distinct categories of balls, and if we have to pick  $r$  balls *with replacement and without regard to order*, then

$$\text{Number of } r\text{-combinations} = \binom{n-1+r}{r} \quad (3.3.7)$$

This is the celebrated Bose–Einstein statistics, where the elementary particles in the same energy state are not distinguishable.

Red	Blue	Green	Brown
● ●	●		● ● ●

FIGURE 3.3.3

**Example 3.3.4** An urn contains red, blue, green, yellow, white, and black balls. In how many ways can we choose

1. 10 balls? This is a direct application of Eq. (3.3.7) with  $n = 6$ ,  $r = 10$  and the number of ways is

$$\binom{6-1+10}{10} = \binom{15}{10} = \binom{15}{5} = 3003$$

2. 20 balls with least 2 colors each? If we draw 2 balls of each color, then we have drawn 12 balls, leaving us to draw  $20-12=8$  balls to be drawn without restriction. Hence, applying Eq. (3.2.7) with  $n = 6$  and  $r = 8$ , we obtain the number of ways as

$$\binom{6-1+8}{8} = \binom{13}{8} = \binom{13}{5} = 1287$$

3. 20 balls with no more than 3 blue balls? We first calculate the number of unrestricted ways of drawing 20 balls as

$$\binom{6-1+20}{20} = \binom{25}{5} = 53,130$$

We next draw at least 4 blue balls out of 20 balls. Having drawn the 4 blue balls, the number of ways to draw the remaining 16 balls is

$$\binom{6-1+16}{26} = \binom{21}{16} = \binom{21}{5} = 20,349$$

Thus the number of ways of drawing 20 balls with no more than 3 blue balls is  $53,130 - 20,349 = 32,781$ .

4. 20 balls with at least 4 red balls, at least 2 green balls, at least 6 white balls, and no more than 3 brown balls? We first draw the lower bounded balls, namely,  $4+2+6=12$ , before we go to the upper-bounded brown balls. We then draw the remaining 8 balls in an unrestricted manner, and the number of ways is

$$\binom{6-1+8}{8} = \binom{13}{5} = 1287$$

We have to subtract from this number the number of ways 4 brown balls can be picked from the 6 colors. This number is

$$\binom{6-1+4}{4} = \binom{9}{4} = 126$$

Hence the number of ways in which all the restrictions will be satisfied is  $1287 - 126 = 1161$ .

**Example 3.3.5** We tackle a different kind of a problem. How many solutions are there to the equation

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 24$$

where  $\{x_k, k = 1, \dots, 6\}$  are positive integers such that

1.  $x_k > 2$  for all  $k$ ? We need  $x_k \geq 3$  for all  $k$ . This takes up  $6 \times 3 = 18$  of the required total of 24, leaving us to solve the equation

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 6$$

in an unrestricted manner. Here we identify, in Eq. (3.2.7),  $n = 6$  corresponding to the number of  $x_k$  values and  $r = 6$  corresponding to the righthand side of the equation. Hence the number of solutions is

$$\binom{6-1+6}{6} = \binom{11}{6} = \binom{11}{5} = 462$$

2.  $x_1 > 2, x_2 \geq 1, x_3 > 3, x_4 \geq 4, x_5 > 2$ , and  $x_6 \geq 2$ ? Using  $x_1 \geq 3, x_3 \geq 4, x_5 \geq 3$ , the number of restrictions add upto  $3 + 1 + 4 + 4 + 3 + 2 = 17$ . This leaves  $24 - 17 = 7$  unrestricted choices. Hence the number of solutions is

$$\binom{6-1+7}{7} = \binom{12}{7} = \binom{12}{5} = 792$$

3.  $x_2 < 5, x_3 > 10$ ? The number of balls that can be drawn with the restriction  $x_3 \geq 11$  corresponding to  $n = 6$  and  $r = (24 - 11) = 13$  is given by

$$\binom{6-1+13}{13} = \binom{18}{13} = \binom{18}{5} = 8568$$

The number of draws that violate the additional restriction  $x_2 < 5$  is obtained by first finding the number of draws with the additional restriction  $x_2 \geq 6$ . The number of ways corresponding to  $n = 6$  and  $r = (24 - 11 - 6) = 7$  is given by

$$\binom{6-1+7}{7} = \binom{12}{7} = \binom{12}{5} = 792$$

Hence the required result is  $8568 - 792 = 7776$  ways.

**Example 3.3.6** Here is an interesting example to illustrate the versatility of the formula in Eq. (3.2.7). Four dice, each with only 4 faces marked 1,2,3,4, are tossed. We have to calculate the number of ways in which the tosses will result in exactly 1 number, 2 numbers, 3 numbers, and 4 numbers showing on the faces of the dice without regard to order and with replacement. The total number of ways is tabulated in Table 3.3.3.

TABLE 3.3.3

1,1,1,1	2,2,2,2	3,3,3,3	4,4,4,4	4 single digits
1,1,1,2	2,2,2,3	3,3,3,4	4,4,4,1	
1,1,2,2	2,2,3,3	3,3,4,4	4,4,1,1	
1,2,2,2	2,3,3,3	3,4,4,4	4,1,1,1	
1,1,1,3	2,2,2,4	3,3,3,1	4,4,4,2	
1,1,3,3	2,2,4,4			
1,1,2,3	2,2,3,4	3,3,4,1	4,4,1,2	18 double digits
1,1,3,4	2,2,4,1	3,3,1,2	4,4,2,3	
1,1,4,2	2,2,1,3	3,3,2,4	4,4,3,1	
1,2,3,4				1 quadruple digit

We can see from Table 3.3.3 that there are 4 single digits, 18 double digits, 12 triple digits, and 1 quadruple digit totaling 35 arrangements without regard to order and with replacement. The problem can be stated as follows. Four dice each with 4 faces are tossed without regard to order and with replacement. Find the number of ways in which

1. This can be achieved. With  $n = 4$  and  $r = 4$  in Eq. (3.2.7), the number of ways is

$$\binom{4-1+4}{4} = \binom{7}{4} = \binom{7}{3} = 35$$

which agrees with the total number in the table.

2. Exactly one digit is present. With  $n = 4$  and  $r = 1$ , the number of ways is

$$\binom{4-1+1}{1} = \binom{4}{1} = 4$$

which agrees with the table.

3. Exactly 2 digits are present. The number of ways 2 digits can be drawn from 4 digits is  $\binom{4}{2}$ . The number of ways any particular 2 digits, for example {1,2}, will occupy a 4-digit number is a restricted case as in Example 3.2.4 with both 1 and 2 occupying two slots and the rest of the two remaining unrestricted. With  $n = 2$  and  $r = 4 - 2 = 2$  in Eq. (3.2.7), the number of ways for this particular sequence {1,2} is  $\binom{2-1+2}{2} = \binom{3}{2} = 3$ . The total number using all the combinations are  $\binom{4}{2} \cdot \binom{3}{2} = 6 \times 3 = 18$  ways. This agrees with the table.
4. Exactly 3 digits are present. The number of ways 3 digits can be drawn from 4 digits is  $\binom{4}{3} = 4$  ways. The number of ways 3 digits can be arranged in 4 slots such that all the three digits are present is again given by the restriction formula Eq. (3.2.7)  $\binom{3-1+(4-3)}{2} = \binom{3}{1} = 3$ . Hence the total number of ways is  $\binom{4}{3} \binom{3}{1} = 12$  ways, agreeing with the table.
5. Exactly 4 digits are present. The number of ways 4 digits will be present is obtained from the argument in item 3 above, giving  $\binom{4}{4} \binom{4-1+(4+4)}{0} = 1$

### Generalization

The result in Example 3.3.6 can be generalized to the tossing of  $n$   $n$ -sided dice. The total number of sample points without regard to order and with replacement is, from Eq. (3.2.7)

$$\binom{n-1+n}{n} = \binom{2n}{n} \quad (3.3.8)$$

Out of these, the number of ways exactly  $r$  digits will occur can be derived as follows: The number of ways  $r$  digits can be chosen from  $n$  is  $\binom{n}{r}$ , and if any particular sequence of  $r$  digits has to fill  $n$  slots, then the number of ways this can be done is obtained by substituting  $r = (n - r)$  in Eq. (3.2.7), yielding

$$\binom{r-1+(n-r)}{n-r} = \binom{n-1}{n-r} \quad (3.3.9)$$

		n-sided die thrown n times							
		1	2	3	4	5	6	7	8
k → number of digits	0	0	0	0	0	0	0	0	0
	1	1	2	3	4	5	6	7	8
	2	0	1	6	18	40	75	126	196
	3	0	0	1	12	60	200	525	1176
	4	0	0	0	1	20	150	700	2450
	5	0	0	0	0	1	30	315	1960
	6	0	0	0	0	0	1	42	588
	7	0	0	0	0	0	0	1	56
	8	0	0	0	0	0	0	0	1
T		1	3	10	35	126	462	1716	6435

FIGURE 3.3.4

Since all the  $\binom{n}{r}$  sequences are functionally independent, the number of ways  $N(n,r)$  is as follows, from Eq. (3.1.2):

$$N(n,r) = \binom{n}{r} = \binom{n-1}{n-r} \quad (3.3.10)$$

Using Eq. (3.3.10),  $N(n,r)$  tabulated for  $n = 1, \dots, 8$  and  $r = 1, \dots, 8$  in Fig. 3.3.4.

In addition, we have the following interesting identity. Combining Eqs. (3.3.8) and (3.3.9), we have

$$\binom{2n-1}{n} = \sum_{r=0}^n \binom{n}{r} \binom{n-1}{n-r} \quad (3.3.11)$$

Equation (3.3.11) can also be expressed by substituting  $m = n - 1$ ,  $r = n$ , and  $k = r$  in the Vandermonde identity of Eq. (3.3.6)

$$\binom{n+m}{r} = \sum_{k=0}^r \binom{n}{k} \binom{m}{r-k} = \binom{n+n-1}{n} = \sum_{k=0}^n \binom{n}{k} \binom{n-1}{n-k}. \quad (3.3.12)$$

# Discrete Distributions

## 4.1 BERNoulli TRIALS

We will elaborate on the Cartesian product already discussed in Chapter 1. Consider the tossing of an unfair coin with success given by  $\{\text{head} = s\}$  with probability  $p$  and failure given by  $\{\text{tail} = f\}$  with probability  $q$ , and  $p + q = 1$ . The sample space is  $S_1 = \{s, f\}$  with field  $\mathcal{F} = [S_1, \emptyset, \{s\}, \{f\}]$ . If the coin is tossed twice, then the sample space  $S_2 = \{ss, sf, fs, ff\}$  is an ordered set from the Cartesian product  $S_1 \times S_1$ . The cardinality of  $S_2$  is  $2 \times 2 = 4$ . The probability of two successes is  $p^2$  and two failures is  $q^2$ . If we toss the coin  $n$  times, then the resulting sample space is  $S_n = S_1 \times S_1 \times \cdots \times S_1 \times S_1$  and the cardinality of  $S_n$  is  $2^n$ . The probability of  $n$  successes is  $p^n$ .

### **Bernoulli Trials**

Bernoulli trials are repeated functionally independent trials with only two events  $s$  and  $f$  for each trial. The trials are also statistically independent with the two events  $s$  and  $f$  in each trial having probabilities  $p$  and  $(1 - p)$ , respectively. These are called *independent identically distributed* (i.i.d.) *trials*.

**Example 4.1.1** Tossing an unfair coin 4 times constitutes 4 Bernoulli trials with heads =  $s$  and tails =  $f$  with  $P\{s\} = p$  and  $P\{f\} = 1 - p$ . The sample space  $S$  is the Cartesian product with ordered 4-tuples. The cardinality of  $S$  is  $2^4 = 16$  points.

**Example 4.1.2** Tossing a die 5 times can be considered Bernoulli trials if we define getting an ace as success,  $s$  with  $P\{s\} = \frac{1}{6}$  and not getting an ace as failure,  $f$  with  $P\{f\} = \frac{5}{6}$ . The resulting sample space  $S$  is the Cartesian product with ordered quintuples with number of points equaling the cardinality  $2^5 = 32$ .

### **Bernoulli Distribution**

Bernoulli distribution is the simplest of a discrete distribution. A single event is Bernoulli distributed if the probability of success is  $p$  and probability of failure is  $q$  with  $p + q = 1$ .

**Example 4.1.3** A single die is thrown. The probability of getting an ace is  $p = \frac{1}{6}$  and the probability of not getting an ace is  $q = \frac{5}{6}$ . Hence this experiment is Bernoulli distributed.

## 4.2 BINOMIAL DISTRIBUTION

We are given an experiment  $S$  with an event  $A \subset S$  with  $P\{A\} = P\{\text{success}\} = p$  and  $P\{\bar{A}\} = P\{\text{failure}\} = q$  with  $p + q = 1$ . We repeat this experiment  $n$  times, and the resulting sample space  $S_n$  is the Cartesian product

$$S_n = \{S \times S \times \cdots \times S\}$$

$n$  times

containing  $2^n$  elementary events. We will now determine the probability  $b(k; n, p)$  that the event  $A$  occurs exactly  $k$  times in  $n$  Bernoulli trials. We shall address this problem in two parts. In the first part we will find the probability of  $k$  successes in a particular sequence in  $n$  trials. For example, one such sequence can be

$$\{s, s, f, s, f, f, \dots, s, f\}$$

exactly  $k$  successes

For the  $k$  independent successes the probability is  $p^k$  and for the  $(n - k)$  independent failures the probability is  $q^{n-k}$ . Since both successes and failures are independent, we have

$$\text{Probability}\{k \text{ successes in a given sequence}\} = p^k q^{n-k} \quad (4.2.1)$$

In the second part we address the situation of  $k$  successes in any sequence in  $n$  trials. The number of ways  $k$  successes can be obtained from  $n$  trials is by sampling without replacement and without regard to order. This number from Eq. (3.3.2) is  $\binom{n}{k}$ . These realizations are mutually exclusive since the occurrence of one realization precludes the occurrence of another. Hence the probabilities will add, giving rise to

$$\text{Probability}\{k \text{ successes in any sequence in } n \text{ trials}\} = b(k; n, p) = \binom{n}{k} p^k q^{n-k} \quad (4.2.2)$$

This is called the *binomial distribution*, and  $b(k; n, p)$  is called a *probability mass function*. We will now show that this is a probability distribution by summing over all  $k$ . Since  $\binom{n}{k} p^k q^{n-k}$  is the  $k$ th term in the binomial expansion of  $(p + q)^n$ , we can write

$$\sum_{k=0}^n b(k; n, p) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1 \quad (4.2.3)$$

The probability of getting utmost  $r$  successes is given by

$$P\{\text{number of successes} \leq r\} = \sum_{k=0}^r b(k; n, p) = \sum_{k=0}^r \binom{n}{k} p^k q^{n-k} \quad (4.2.4)$$

and the probability of getting at least  $r$  successes is given by

$$P\{\text{number of successes} \geq r\} = \sum_{k=r}^n b(k; n, p) = \sum_{k=r}^n \binom{n}{k} p^k q^{n-k} \quad (4.2.5)$$

Since  $n!$  grows very rapidly with  $n$ , computational difficulties in calculating  $b(k;n,p)$  can be avoided by using the recursive formula shown below with starting value  $b(0;n,p)$ .

$$\begin{aligned} b(k+1;n,p) &= \binom{n}{k+1} p^{k+1} q^{n-k-1} = \frac{n-k}{k+1} \cdot \frac{p}{q} \binom{n}{k} p^k q^{n-k} \\ &= \frac{n-k}{k+1} \cdot \frac{p}{q} b(k;n,p) \end{aligned} \quad (4.2.6)$$

**Example 4.2.1** A fair die is rolled 10 times, and we will calculate

1. The probability that an ace comes up exactly 4 times. Substituting  $p = \frac{1}{6}$  and  $q = \frac{5}{6}$ ,  $n = 10$ , and  $k = 4$  in Eq. (4.2.3), we obtain

$$P\{6 \text{ aces in 10 trials}\} = \frac{10!}{6!4!} \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^6 = 0.05427$$

2. The probability that utmost 4 aces show up. In probability 1 (above) we wanted exactly 4 aces coming up. Now we are looking into the possibility of 4 or fewer aces showing up. This probability is obtained as follows:

$$P\{\text{number of aces} \leq 4\} = \sum_{k=0}^4 \frac{10!}{k!(10-k)!} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{10-k} = 0.98454$$

3. The probability that the number of aces showing is between 2 and 6 inclusive. Here we have to sum  $b(k; 10, \frac{1}{6})$  between 2 and 6. Performing the indicated summation, we have

$$P\{2 \leq \text{number of aces} \leq 6\} = \sum_{k=2}^6 \frac{10!}{k!(10-k)!} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{10-k} = 0.22451$$

The graph of the probability mass function  $\binom{10}{i} \left(\frac{1}{6}\right)^i \left(\frac{5}{6}\right)^{10-i}$  is shown in Fig. 4.2.1, and that of the cumulative distribution function  $\sum_{i=0}^k \binom{10}{i} \left(\frac{1}{6}\right)^i \left(\frac{5}{6}\right)^{10-i}$  is shown in Fig. 4.2.2 for  $k = 0, \dots, 10$ .

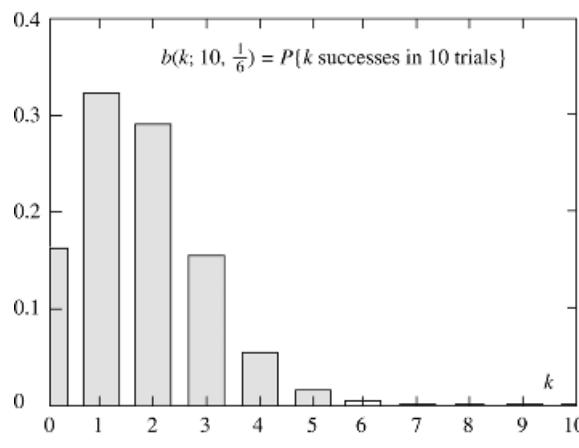


FIGURE 4.2.1

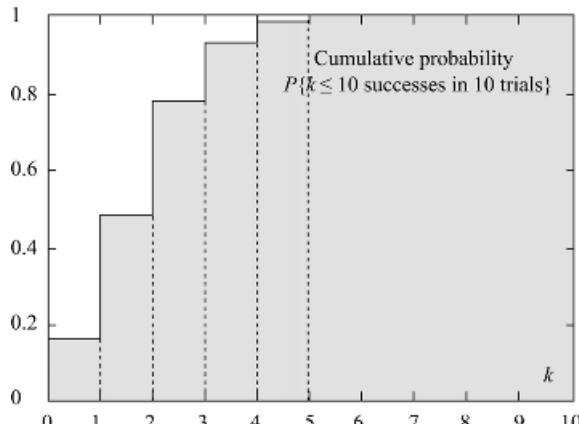


FIGURE 4.2.2

**Example 4.2.2** We shall use the recursive form [Eq. (4.2.6)] to compute the binomial coefficients in Example 4.2.1 with initial condition  $b(0;n,p) = 0.161506$ . These values are substituted in Eq. (4.2.4), and the first four steps are shown:

$$b\left(0; 10, \frac{1}{6}\right) = 0.161506$$

$$b\left(1; 10, \frac{1}{6}\right) = \frac{10}{1} \cdot \frac{1}{5} \cdot 0.161506 = 0.323011$$

$$b\left(2; 10, \frac{1}{6}\right) = \frac{9}{2} \cdot \frac{1}{5} \cdot 0.323011 = 0.29071$$

$$b\left(3; 10, \frac{1}{6}\right) = \frac{8}{3} \cdot \frac{1}{5} \cdot 0.29071 = 0.155045$$

The true values and the computed values are shown in Table 4.2.1, and they are the same.

**Example 4.2.3** Two experiments are performed. In the first experiment 6 fair dice are tossed and the person wins if she gets 1 or more aces. In the second experiment 12 dice are tossed and the person wins if she gets 2 or more aces. We want to find in which experiment the probability of winning is higher.

TABLE 4.2.1

$b(k;n,p)$					
$k$	Computed	Calculated	$k$	Computed	Calculated
0	0.161506	0.161506	6	0.002171	0.002171
1	0.323012	0.323011	7	0.000248	0.000248
2	0.290711	0.29071	8	0.000019	0.000019
3	0.155046	0.155045	9	0.000001	0.000001
4	0.054266	0.054266	10	0	0
5	0.013024	0.013024			

The probability of getting one or more aces out of 6 tosses is the same as the complement of the probability of getting no aces in 6 throws. This is obtained from Eq. (4.2.5) with  $n = 6$ ,  $p = \frac{1}{6}$  and  $q = \frac{5}{6}$ .

$$1 - P\{\text{no aces in 6 throws}\} = 1 - \binom{6}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^6 = 0.6651$$

The probability of getting 2 or more aces in 12 throws is the same as the complement of the sum of the probabilities of getting no aces and one ace in 12 throws. With  $n = 12$ , we obtain from Eq. (4.2.5)

$$\begin{aligned} 1 - P\{\text{no aces in 12 throws}\} - P\{\text{one ace in 12 throws}\} \\ = 1 - \binom{12}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^{12} - 1 - \binom{12}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^{11} \\ = 1 - 0.1122 - 0.2692 = 0.6186 \end{aligned}$$

Hence the probability of getting one or more aces in 6 throws is better than getting two or more aces in 12 throws.

**Example 4.2.4** An urn consists of 40 red balls and 60 green balls. What is the probability of getting exactly  $k$  red balls in a sample size of 10 if the sampling is done

1. With replacement? The meaning of “with replacement” is to pick a ball from the 100 in the urn note its color and replace it and pick again, and this process is continued for 10 times. Thus  $P\{\text{red ball}\} = p = \frac{2}{5}$  and  $P\{\text{green ball}\} = q = \frac{3}{5}$  and with  $n = 10$ , the probability of  $k$  red balls in 10 trials is given by

$$b\left(k; n, \frac{2}{5}\right) = \binom{10}{k} \left(\frac{2}{5}\right)^k \left(\frac{3}{5}\right)^{10-k}$$

2. Without replacement? Here we sample the balls and do not replace it in the urn. Thus, the number of ways of getting 10 balls from 100 balls is  $\binom{100}{10}$ . Out of these 10 balls, the number of ways of getting  $k$  red balls is  $\binom{40}{k}$  and the number of ways of getting  $(10 - k)$  blue balls is  $\binom{60}{10-k}$ . Since getting  $k$  red balls and  $(10 - k)$  blue balls are functionally independent, the total number of ways are  $\binom{40}{k} \cdot \binom{60}{10-k}$  using the product rule. Hence the probability  $P(k)$  of  $k$  red balls without replacement is given by

$$P(k) = \frac{\binom{40}{k} \binom{60}{10-k}}{\binom{100}{10}}$$

## 4.3 MULTINOMIAL DISTRIBUTION

The binomial distribution of the previous section can be generalized to  $n$  repeated independent trials where each trial has  $m$  outcomes  $\{E_i, i = 1, \dots, m\}$  with probabilities  $P\{E_i\} = \{p_i, i = 1, \dots, m\}$  and

$$p_1 + p_2 + \dots + p_m = 1 \quad (4.3.1)$$

Each of the outcomes  $\{E_i, i = 1, \dots, m\}$  occurs  $\{k_i, i = 1, \dots, m\}$  times with

$$k_1 + k_2 + \dots + k_m = n \quad (4.3.2)$$

The probability  $P\{k_1, k_2, \dots, k_m\}$  is given by the multinomial probability distribution,

$$\begin{aligned} P\{k_1, k_2, \dots, k_m\} &= \binom{n}{k_1, k_2, \dots, k_m} p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m} \\ &= \frac{n!}{k_1! k_2! \cdots k_m!} p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m} \end{aligned} \quad (4.3.3)$$

If  $m = 2$ , the multinomial distribution reduces to the binomial distribution.

**Example 4.3.1** A pair of dice is rolled 8 times. We will find the probability that 7 or 11 occurs thrice; 2, 3, or 12 twice; any pair twice; and any other combination once. For each rolling of a pair there are four events  $E_1, E_2, E_3, E_4$  and their probabilities are obtained from Table 2.2.1 and are shown below:

$$\begin{aligned} E_1 &= \{7 \text{ or } 11\} & P\{E_1\} &= p_1 = \frac{8}{36} \\ E_2 &= \{2, 3 \text{ or } 12\} & P\{E_2\} &= p_2 = \frac{4}{36} \\ E_3 &= \{\text{any pair}\} & P\{E_3\} &= p_3 = \frac{6}{36} \\ E_4 &= \{\text{other combination}\} & P\{E_4\} &= p_4 = \frac{18}{36} \end{aligned}$$

In Eq. (4.3.3) we substitute  $k_1 = 3, k_2 = 2, k_3 = 2, k_4 = 1$  with  $p_1, p_2, p_3, p_4$  as given above, yielding

$$P\{3, 2, 2, 1\} = \frac{8!}{3!2!2!1!} \left(\frac{8}{36}\right)^3 \left(\frac{4}{36}\right)^2 \left(\frac{6}{36}\right)^2 \left(\frac{18}{36}\right)^1 = 0.003161$$

## 4.4 GEOMETRIC DISTRIBUTION

Closely allied to the binomial distribution is the *geometric distribution*. A sequence of independent Bernoulli trials with  $p$  as the probability of success can be modeled in two different ways. The distribution of the number of successes  $k$  is binomial. The waiting time until the first success occurs is geometrically distributed. If the first success is to occur at the  $k$ th trial, then we must have  $(k - 1)$  failures preceding the first success. Since the trials are independent, this probability is given by

$$g(k, p) = (1 - p)^{k-1} p, k = 1, 2, \dots \quad (4.4.1)$$

This is the geometric distribution with parameter  $p$ . Further, the waiting times between each success and the next are independent and are geometrically distributed.

We can show that the sum to infinity of  $g(k, p)$  is indeed 1 as shown below:

$$\sum_{k=1}^{\infty} g(k, p) = p \sum_{k=1}^{\infty} (1 - p)^{k-1} = \frac{p}{1 - (1 - p)} = 1$$

**Example 4.4.1** In an urn there are  $r$  red balls and  $b$  blue balls. Balls are selected at random with replacement so that the first blue ball is obtained. We have to find the probability that

1. Exactly  $k$  draws are needed for the first blue ball. The probability of a blue ball is  $p = b/(b+r)$ . Substituting this value  $p$  in Eq. (4.4.1), the required probability is

$$\begin{aligned} g\left(k, \frac{b}{b+r}\right) &= \left(1 - \frac{b}{b+r}\right)^{k-1} \frac{b}{b+r} \\ &= \left(\frac{r}{b+r}\right)^{k-1} \frac{b}{b+r} = \frac{br^{k-1}}{(b+r)^k} \end{aligned}$$

2. At least  $m$  draws needed for the first blue ball. The first success occurs at the  $m$ th draw after  $(m-1)$  failures. Since the failure probability is  $q = [r/(b+r)]$ , we have from the definition of the geometric distribution this probability is  $[r/(b+r)]^{m-1}$ .

The same result can also be obtained from the cumulative summation as follows:

$$\begin{aligned} \sum_{k=m}^{\infty} g\left(k, \frac{b}{b+r}\right) &= \frac{b}{b+r} \sum_{k=m}^{\infty} \left(\frac{r}{b+r}\right)^{k-1} \\ &= \frac{b}{b+r} \left(\frac{r}{b+r}\right)^{m-1} \left[ 1 + \frac{r}{b+r} + \left(\frac{r}{b+r}\right)^2 + \dots \right] \\ &= \frac{b}{b+r} \left(\frac{r}{b+r}\right)^{m-1} \left[ \frac{1}{1 - \frac{r}{b+r}} \right] = \left(\frac{r}{b+r}\right)^{m-1} \end{aligned}$$

This geometric distribution is shown in Fig. 4.4.1 for  $r = 100$ ,  $b = 20$ , and  $k = 1, \dots, 10$  and the cumulative distribution in Fig. 4.4.2.

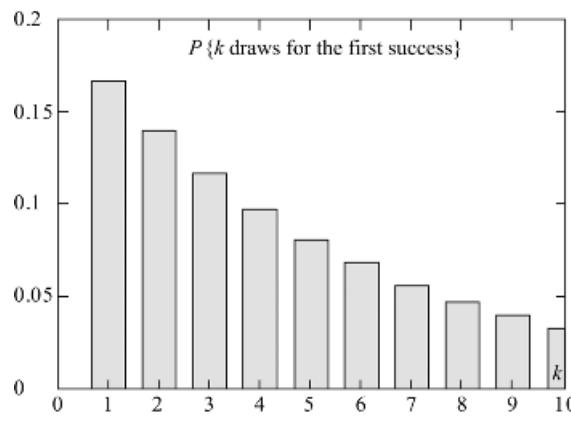


FIGURE 4.4.1

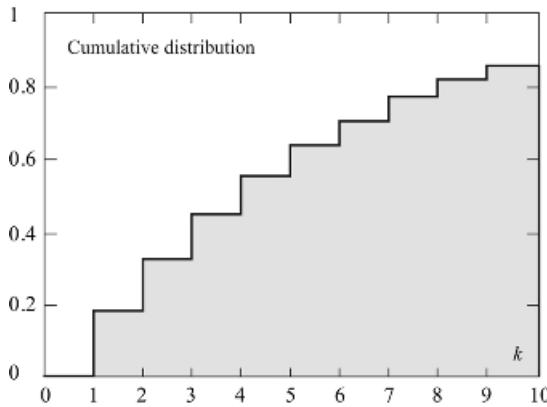


FIGURE 4.4.2

## 4.5 NEGATIVE BINOMIAL DISTRIBUTION

The binomial distribution finds the probability of exactly  $k$  successes in  $n$  independent trials. The geometric distribution found the number of independent trials needed for the first success. We can generalize these two results and find the number of independent trials required for  $k$  successes. This distribution is known as the *negative binomial* or the *Pascal distribution*. Let us now define the following events to find this distribution:

$E = n$  trials for exactly  $k$  successes

$A =$  exactly  $(k - 1)$  successes occur in  $(n - 1)$  trials

$B =$   $n$ th trial results in  $k$ th success

Since  $E = A \cap B$  and  $A$  and  $B$  are independent, we have

$$P\{E\} = P\{A\} \cdot P\{B\} \quad (4.5.1)$$

We will now compute  $P\{A\}$ . In a sequence of  $(n - 1)$  trials we have exactly  $(k - 1)$  successes and  $[n - 1 - (k - 1)] = (n - k)$  failures. Since these trials are independent, the probability of that particular sequence is  $p^{k-1}q^{n-1}$ . There are  $\binom{n-1}{k-1}$  mutually exclusive sequences, and hence

$$P\{A\} = b(k - 1; n - 1, p) = \binom{n-1}{k-1} p^{k-1} q^{n-k} \quad (4.5.2)$$

Since the probability  $P\{B\} = p$ , we can substitute this and Eq. (4.5.2) in Eq. (4.5.1) and write

$$nb(k; n, p) = P\{E\} = p \binom{n-1}{k-1} p^{k-1} q^{n-k} = \binom{n-1}{k-1} p^k q^{n-k}, n = k, k + 1, \dots \quad (4.5.3)$$

We have to show that the sum of  $nb(k; n, p)$  from  $n = k$  to  $\infty$  equals 1. Or

$$\sum_{n=k}^{\infty} nb(k; n, p) = \sum_{n=k}^{\infty} \binom{n-1}{k-1} p^k q^{n-k} = 1 \quad (4.5.4)$$

TABLE 4.5.1 Negative binomial

$n$	Probability	$n$	Probability	$n$	Probability
5	0.16807	9	0.095296	13	0.005458
6	0.252105	10	0.05146	14	0.002365
7	0.226894	11	0.02573	15	0.000993
8	0.158826	12	0.01213		

By substituting  $(n - k) = m$ , we can rewrite Eq. (4.5.4) as

$$\sum_{m=0}^{\infty} nb(k; k + m, p) = \sum_{m=0}^{\infty} \binom{k+m-1}{k-1} p^k q^m = \sum_{m=0}^{\infty} \binom{k+m-1}{m} p^k q^m = 1 \quad (4.5.5)$$

The Taylor series expansion of  $(1 - x)^{-k}$  yields

$$\begin{aligned} (1 - x)^{-k} &= 1 + kx + \frac{(-k)(-k-1)}{2!}(-x)^2 + \frac{(-k)(-k-1)(-k-2)}{3!}(-x)^3 + \dots \\ &= \sum_{m=0}^{\infty} \binom{k+m-1}{m} x^m \end{aligned} \quad (4.5.6)$$

Substituting  $x = q$  in Eq. (4.5.6) and  $(1 - q) = p$ , we obtain

$$p^{-k} = \sum_{m=0}^{\infty} \binom{k+m-1}{m} q^m, \quad \text{or} \quad 1 = \sum_{m=0}^{\infty} \binom{k+m-1}{m} p^k q^m, \text{ which is Eq. (4.5.5)}$$

In Eq. (4.5.3), if we substitute  $k = 1$ , then we get the geometric distribution.

**Example 4.5.1** A computer software company wants to recruit 5 new engineers. The probability that a person interviewed is given a job offer is 0.7. We want to find the distribution of the number of people interviewed for the 5 positions.

We use Eq. (4.5.3) with  $p = 0.7$  and  $k = 5$ . Thus

$$P\{E = n\} = \binom{n-1}{5-1} (0.7)^5 \cdot (0.3)^{n-5}, \quad n = 5, 6, \dots$$

$P\{E = n\}$  is shown for  $n = 5, 6, \dots, 15$  in Table 4.5.1, and the probability mass function is shown in Fig. 4.5.1.

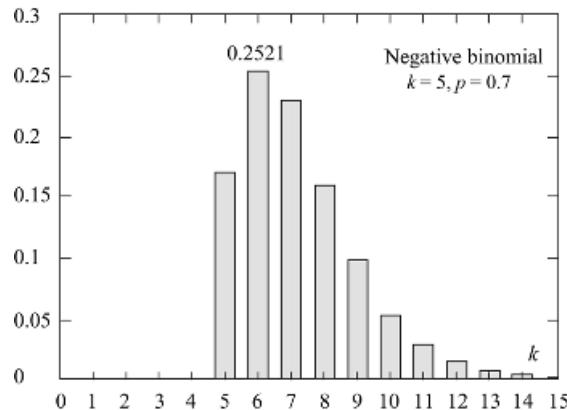


FIGURE 4.5.1

From the table and the figure we see that the maximum probability of 0.2521 occurs at  $k = 6$  and it is most likely that company will interview exactly 6 persons for the 5 jobs.

## 4.6 HYPERGEOMETRIC DISTRIBUTION

The binomial distribution is obtained while sampling with replacement. Let us select a sample of  $n$  microprocessor chips from a bin of  $N$  chips. It is known that  $K$  of these chips are defective. If we sample with replacement then the probability of any defective chip is  $K/N$ . The probability that  $k$  of them will be defective in a sample size of  $n$  is given by the binomial distribution since each drawing of a chip constitutes a Bernoulli trial. On the other hand, if we sample without replacement, then the probability for the first chip is  $K/N$ , the probability for the second chip is  $(K-1)/(N-1)$ , and for each succeeding sample the probability changes. Hence the drawings of the chips do not constitute Bernoulli trials and binomial distribution does not hold. Hence we have to use alternate methods to arrive at the probability of finding  $k$  defective chips in  $n$  trials.

The number of sample points in this sample space is choosing  $n$  chips out of the total  $N$  chips without regard to order. This can be done in  $\binom{N}{n}$  ways. Out of this sample size of  $n$ , there are  $k$  defective chips and  $(n-k)$  good chips. The number of ways of picking  $k$  defective chips out of the total of  $K$  defective chips is  $\binom{K}{k}$  and the number of ways of drawing  $(n-k)$  good chips out of the remaining  $(N-K)$  good chips is  $\binom{N-K}{n-k}$ . We denote the total number of ways of drawing a sample of  $n$  chips with  $k$  bad chips as the event  $E$ . Since these drawings are functionally independent the number of ways is  $\binom{K}{k} \binom{N-K}{n-k}$ . Hence the probability of the event  $E$  is given by

$$P\{E = k\} = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} \quad (4.6.1)$$

This is known as the *hypergeometric distribution*. We came across this distribution in Example 4.2.4. Since the summation over all  $k$  must be one, we must have the following in Eq. (4.6.1):

$$\sum_{k=0}^n \binom{K}{k} \binom{N-K}{n-k} = \binom{N}{n} \quad (4.6.2)$$

This result follows directly from Vandermonde's identity of Eq. (3.3.6) given by

$$\sum_{k=0}^n \binom{n}{k} \binom{m}{r-k} = \binom{n+m}{r} \quad (3.3.6)$$

with  $n = K$ ,  $m = N - K$ , and  $r = n$ .

**Example 4.6.1 (Mass Megabucks)** Massachusetts megabucks lottery is played as follows. A player has to choose 6 numbers from a slate of 42 numbers. Twice a week the lottery commission draws 6 numbers. If all the 6 numbers of the player match all the 6 numbers drawn by the commission, the player wins megabucks. There are smaller prizes for matching 5, 4, or 3 numbers. We have to calculate the probability of winning megabucks or the smaller prizes.

TABLE 4.6.1 Mass megabucks

$k$	$N(k)$	$C(k)$	$P(k)$
0	1947792	3	0.371306035
1	2261952	2	0.431194105
2	883575	6	0.168435197
3	142800	37	0.02722185
4	9450	555	0.001801446
5	216	24286	0.000041176
6	1	5,245,786	0.000000191

Let  $k = 0, 1, \dots, 6$ . Substituting  $N = 42$ ,  $n = 6$ , and  $K = 6$ , we can find from Eq. (4.6.1)  $P\{E = k\}$ :

$$P\{E = k\} = \frac{N(k)}{D(k)} = \frac{\binom{6}{k} \binom{42-K}{6-k}}{\binom{42}{6}}, \quad k = 0, 1, \dots, 6$$

The results are tabulated for  $k = 0, \dots, 6$  in Table 4.6.1. In this table, the second column  $N(k) = \binom{6}{k} \binom{42-K}{6-k}$  is the number of ways  $k$  digits can be formed from 6 digits. In the third column  $C(k) = D(k)/N(k)$  represents the chance of winning with  $k$  digits rounded to the nearest digit and in the fourth column  $P(k)$  is the probability of winning.

**Example 4.6.2** Hypergeometric distribution is used to estimate an unknown population from the data obtained. To estimate the population  $N$  of tigers in a wildlife sanctuary,  $K$  tigers are caught, tagged, and released. After the lapse of a few months, a new batch of  $n$  tigers is caught, and  $k$  of them are found to be tagged. It is assumed that the population of tigers does not change between the two catches. The total number of tigers  $N$  in the sanctuary is to be estimated.

In Eq. (4.6.1), the known quantities are  $K$ ,  $k$ , and  $n$ . The total number  $N$  has to be estimated in the following equation:

$$P_k\{N\} = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

We will find the value of  $N$  that maximizes the probability  $P_k\{N\}$ . Such an estimate is called the *maximum-likelihood estimate*, which will be discussed in Chapter 18. Since the numbers are very large, we investigate the following ratio:

$$\begin{aligned} \frac{P_k\{N\}}{P_k\{N-1\}} &= \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} \Big/ \frac{\binom{K}{k} \binom{N-1-K}{n-k}}{\binom{N-1}{n}} \\ &= \frac{\binom{N-K}{n-k} \binom{N-1}{n}}{\binom{N}{n} \binom{N-1-K}{n-k}} = \frac{N-n}{N} \frac{N-K}{N-K-n+k} \end{aligned}$$

The ratio  $(P_k\{N\})/(P_k\{N - 1\})$  is greater than 1 if  $(N - n)(N - K) > N(N - K - n + k)$ ,  $Nk < nK$ , or  $N < nK/k$ .

**Conclusion.** If  $N$  is less than  $nK/k$ , then  $P_k(N)$  is increasing, and if  $N$  is greater than  $nK/k$ , then  $P_k(N)$  is decreasing. Hence the maximum value of  $P_k(N)$  occurs when  $N$  is equal to the nearest integer not exceeding  $nK/k$ . Thus, if  $K = 100$ ,  $n = 50$ , and  $k = 20$ , then the population of tigers is  $100.50/20 = 250$ .

## **Approximation to Binomial Distribution**

If  $N$  and  $K$  are very large with respect to  $n$  and  $k$ , then the hypergeometric distribution can be approximated by the binomial distribution as shown below:

$$\begin{aligned}
P\{E = k\} &= \binom{K}{k} \binom{N-K}{n-k} / \binom{N}{n} \\
&= \frac{k!}{k!(K-k)!} \frac{(N-K)!}{(n-k)!(N-K-n+k)!} \frac{n!(N-n)!}{N!} \\
&= \frac{n!}{k!(n-k)!} \frac{K!}{(K-k)!} \frac{(N-n)!}{N!} \frac{(N-K)!}{(N-K-n+k)!} \\
&= \binom{n}{k} \frac{K(K-1)\cdots(K-k+1)}{N(N-1)\cdots(N-k+1)} \\
&\quad \times \frac{(N-K)(N-K-1)\cdots(N-K-n+k+1)}{(N-k)(N-K-1)\cdots(N-k-n+k+1)} \quad (4.6.3)
\end{aligned}$$

Since  $K \gg k$  and  $N \gg n$ , Eq. (3.3.7) can be approximated as

$$P\{E=k\} = \binom{n}{k} \frac{K \cdot K \cdots K}{N \cdot N \cdots N} \frac{(N-K)(N-K) \cdots (N-K)}{N \cdot N \cdots N} \quad (4.6.4)$$

Approximating  $p \cong K/N$  and  $q \cong (N - K)/n$  in Eq. (3.3.8), we obtain

$$P\{E = k\} \cong \binom{n}{k} p^k q^{n-k} \quad (4.6.5)$$

which is a binomial distribution.

## 4.7 POISSON DISTRIBUTION

Perhaps three of the most important distributions in probability theory are (1) Gaussian, (2) Poisson, and (3) binomial. We have already discussed the binomial and other allied distributions that are characterized by discrete time ( $n$ ) and discrete state, that is, the cumulative distribution function has jumps occurring only at discrete points. In the regular Poisson distribution the cumulative distribution has unit jumps at random times and not at discrete times. Thus, it is discrete in state but continuous in time.

The Poisson distribution takes values  $0, 1, \dots, k, \dots$  at random points with a parameter  $\lambda$  and is given by

$$p\{k ; \lambda\} = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \geq 0 \quad (4.7.1)$$

We will show that this is a probability function by summing  $k$  from 0 to  $\infty$  as shown below:

$$\sum_{k=0}^{\infty} p\{k ; \lambda\} = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1 \quad (4.7.2)$$

Similar to Eq. (4.2.6) we can derive a recursive form for evaluating the Poisson distribution as shown below:

$$\sum_{k=0}^{\infty} p\{k+1 ; \lambda\} = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^{k+1}}{(k+1)!} = \frac{\lambda}{k+1} \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = \frac{\lambda}{k+1} p\{k ; \lambda\} \quad (4.7.3)$$

This last equation is solved from the starting value  $p(0, \lambda) = e^{-\lambda}$ .

**Example 4.7.1** During a certain experiment the number of  $\alpha$  particles passing through a sensor in a 1-ms interval is 3.2. The Poisson distribution from Eq. (4.7.1) with  $\lambda = 3.2$  can be written as

$$p\{k ; 3.2\} = e^{-3.2} \frac{(3.2)^k}{k!}, \quad k \geq 0$$

This is graphed in Fig. 4.7.1.

We want to find the probability that

1. Six  $\alpha$  particles pass through the sensor in a given millisecond. Substituting  $\lambda = 3.2$  and  $k = 6$ , we obtain

$$p\{6 ; 3.2\} = e^{-3.2} \frac{(3.2)^6}{6!} = 0.06079$$

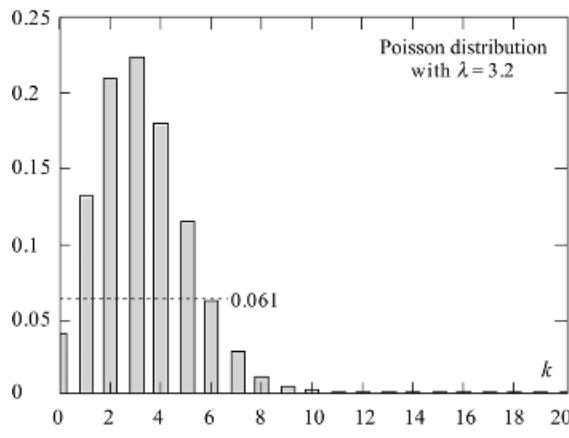


FIGURE 4.7.1

2. No more than three  $\alpha$  particles pass through the sensor. We have to find the cumulative distribution given by

$$\sum_{k=0}^3 p\{k ; 3.2\} = \sum_{k=0}^3 e^{-3.2} \frac{(3.2)^k}{k!} = 0.6025$$

**Example 4.7.2** The average number of customers arriving at the service counter in a supermarket is  $\lambda$ , and the probability of  $k$  customers arriving in the interval  $(0,t)$  minutes is given by the Poisson distribution  $p\{k ; \lambda\} = e^{-\lambda}[(\lambda t)^k/k!]$ . In the nonoverlapping intervals  $(0,t_1)$  and  $(t_1,t)$ , we can write the following Poisson law

$$\begin{aligned} & P\{n_1 \text{ customers in } (0,t_1) \text{ and } n_2 \text{ in } (t_1,t)\} \\ &= P\{n_1 \text{ customers in } (0,t_1)\} \cdot P\{n_2 \text{ customers in } (t_1,t)\} \end{aligned}$$

and the two events are independent in disjoint intervals. We will find the conditional probability,  $P\{n_1 \text{ customers in } (0,t_1) \mid (n_1 + n_2) \text{ customers in } (0,t)\}$ . From the definition of conditional probability, we can write

$$\begin{aligned} P\{n_1; (0,t_1) \mid (n_1 + n_2); (0,t)\} &= \frac{P\{[n_1; (0,t_1)][(n_1 + n_2); (0,t)]\}}{P\{(n_1 + n_2); (0, t)\}} \\ &= \frac{P\{[n_1; (0,t_1)][n_2; (t_1,t)]\}}{P\{(n_1 + n_2); (0, t)\}} \\ &= \frac{e^{-\lambda t_1} (\lambda t_1)^{n_1}}{n_1!} \frac{e^{-\lambda(t-t_1)} [\lambda(t-t_1)]^{n_2}}{n_2!} \left/ \frac{e^{-\lambda t} (\lambda t)^{n_1+n_2}}{(n_1 + n_2)!} \right. \\ &= \frac{(\lambda t_1)^{n_1} [\lambda(t-t_1)]^{n_2} (n_1 + n_2)!}{(\lambda t)^{n_1+n_2} n_1! n_2!} \\ &= \frac{t_1^{n_1} (t-t_1)^{n_2} (n_1 + n_2)!}{t^{n_1+n_2} n_1! n_2!} \end{aligned}$$

which yields a surprising result independent of  $\lambda$ !

Poisson distribution has been used in a number of applications such as the

1. Number of telephone calls coming into an exchange
2. Number of cars entering a highway ramp
3. Number of packets transmitted along wireless lines
4. Number of customers at the checkout counter
5. Number of misprints in a book
6. Number of photons emitted

### **Approximation to Binomial Distribution**

Poisson distribution is also used for approximating a binomial distribution with parameters  $(n,p)$  under the condition that  $n$  is large compared to  $k$  and  $p$  is small enough that the product  $np = \lambda$  is of moderate magnitude. As a rule of thumb,  $\lambda$  may

be a single digit. We shall now derive this approximation:

$$\begin{aligned}
 b(k;n,p) &= \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
 &= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
 &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}
 \end{aligned} \tag{4.7.4}$$

As  $n$  becomes very large and for moderate  $\lambda$ , the quantity  $(1 - \lambda/n)^n$  can be approximated by  $e^{-\lambda}$  and the quantity  $(1 - \lambda/n)^k$  can be approximated as 1. Finally, as  $k$  is much smaller than  $n$ , the quantity  $[n(n-1)\cdots(n-k+1)]/n^k$  can be approximated by 1. Thus the binomial distribution  $b(k;n,p)$  can be approximated as

$$b(k;n,p) \approx \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \geq 0 \tag{4.7.5}$$

**Example 4.7.3** A computer system consists of 1000 microchips, and the probability of failure of a microchip is 0.005. We have to find the probability of  $k$  microchips failing using both the binomial distribution and its Poisson approximation for  $k = 0, 1, \dots, 15$ . We will also find the percentage error between the two.

Using binomial distribution,  $k$  chips failing in 1000 trials given by  $b(k; 1000, 0.005) = b(k)$  is

$$b(k; 1000, 0.005) = b(k) = \frac{1000!}{k!(1000-k)!} 0.005^k 0.995^{1000-k}$$

The value  $b(k)$  can be solved using the recursive formula developed in Eq. (4.2.6), and the steps are

$$\begin{aligned}
 b(k+1) &= \frac{1000-k}{k+1} \frac{0.005}{0.995} b(k) \\
 &= \frac{1000-k}{k+1} \frac{1}{188} b(k)
 \end{aligned}$$

Using the Poisson approximation we have  $np = \lambda = 5$ , and substituting in the Poisson distribution, we have

$$p(k; \lambda) = p(k; 5) = p(k) = e^{-5} \left(\frac{5^k}{k!}\right)$$

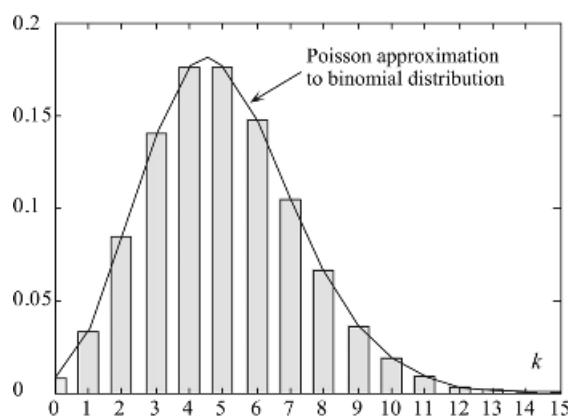
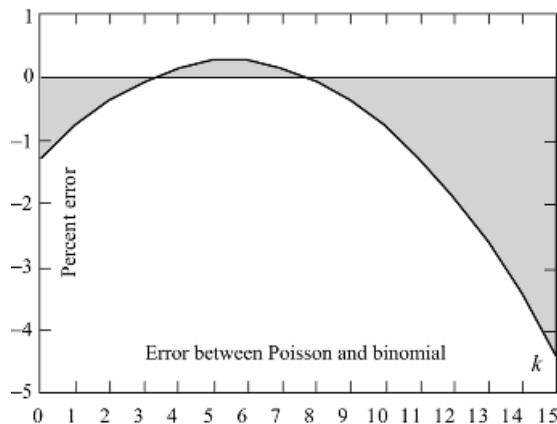
Both  $b(k)$  and  $p(k)$  are shown for  $k = 0, 1, \dots, 15$  in Table 4.7.1 along with the percentage error.

Table 4.7.1 is also graphed in Fig. 4.7.2 for the binomial distribution and the Poisson approximation. The figure shows that the approximation is very close. The error in the approximation is also graphed in Fig. 4.7.3.

From the table and the graphs we see that the approximation becomes worse as  $k$  is increased and the approximation is excellent in the vicinity of  $k = 5 = \lambda$ , about 0.05–0.25%.

**TABLE 4.7.1 Comparison of binomial with Poisson approximation**

$k$	Binomial	Poisson	%Error
0	0.006654	0.006738	-1.26208
1	0.033437	0.03369	-0.75577
2	0.083929	0.084224	-0.352343
3	0.140303	0.140374	-0.050683
4	0.175731	0.175467	0.150021
5	0.175908	0.175467	0.250272
6	0.14659	0.146223	0.250272
7	0.104602	0.104445	0.14992
8	0.065245	0.065278	-0.051188
9	0.036138	0.036266	-0.353762
10	0.017996	0.018133	-0.758823
11	0.008139	0.008242	-1.267706
12	0.003371	0.003434	-1.88207
13	0.001287	0.001321	-2.603907
14	0.000456	0.000472	-3.435549
15	0.000151	0.000157	-4.379687

**FIGURE 4.7.2****FIGURE 4.7.3**

**Example 4.7.4 (See Ref. 47)** We will revisit the two or more birthdays matching problem posed in Example 3.2.4, and solve it using the Poisson approximation and compare the results. In addition, we will use the Poisson approximation to find probability of matching three or more birthdays that would have been very difficult using the exact method.

Given the total numbers of persons  $n \leq 365$ , we will find, using the Poisson approximation, the probability of two or more birthdays matching. The probability of two birthdays matching is  $\frac{1}{365}$ , and this we will set equal to  $p$ . The number of ways we can obtain a sample of two without replacement and without ordering is  $\binom{n}{2}$ , each of which has the same probability of  $\frac{1}{365}$  of matching. Hence the  $\binom{n}{2}$  ways constitute Bernoulli trials and since  $p = \frac{1}{365}$  is very small and

$$np = \binom{n}{2} \frac{1}{365} = \frac{n(n-1)}{730} = \lambda$$

is less than 10, all the conditions for approximation by a Poisson distribution are satisfied. Hence

$$\begin{aligned} P\{2 \text{ or more people share the same birthday}\} \\ = 1 - P\{\text{no birthdays match}\} \\ = 1 - p(0; \lambda) = 1 - \exp\left(-\frac{n(n-1)}{730}\right) \frac{\lambda^0}{0!} \end{aligned}$$

and  $[1 - p(0; \lambda)]$  is shown in Table 4.7.2 for the same values of  $n$  as given in Table 3.2.3 with the corresponding percentage error shown in Table 4.7.3.

The percent error curve between the true value and the Poisson approximation for values of  $n = 1, \dots, 100$  is shown in Fig. 4.7.4.

An examination of the error curve reveals that the Poisson approximation underestimates the true values, and for values of  $n$  higher than 82, the error is almost zero. Even then the maximum error is only about 1.46% occurring at  $n = 26$ .

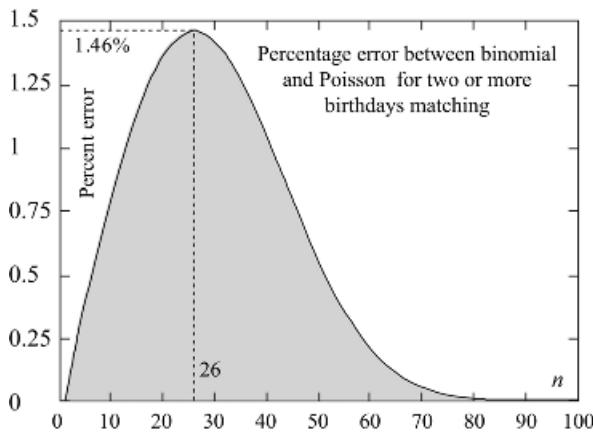
As mentioned before, this approximation can be extended to the case of three or more matches. Approaching, as before, the number of 3-combinations is  $\binom{n}{3} = \frac{n(n-1)(n-2)}{6}$ , and

**TABLE 4.7.2 Poisson approximation for two matches**

#	Probability	#	Probability	#	Probability
20	0.0405805	30	0.69632	44	0.925113
21	0.437488	31	0.720282	45	0.933618
22	0.468938	32	0.743058	46	0.941318
23	0.500002	33	0.764625	47	0.948266
24	0.530536	34	0.784972	48	0.954517
25	0.560412	35	0.804097	49	0.960121
26	0.589513	36	0.82201	50	0.965131

**TABLE 4.7.3 Error in Poisson approximation**

#	%Error	#	%Error	#	%Error
20	1.369162	30	1.415262	44	0.833107
21	1.397497	31	1.392672	45	0.781948
22	1.420488	32	1.365868	46	0.731364
23	1.438108	33	1.335125	47	0.681649
24	1.450358	34	1.300742	48	0.633067
25	1.457273	35	1.263039	49	0.58586
26	1.458918	36	1.222352	50	0.540238

**FIGURE 4.7.4**

the probability  $p$  of three matches is  $(\frac{1}{365})^2$ . Substituting in the Poisson formula we obtain

$$\begin{aligned}
 P\{3 \text{ or more share the same birthday}\} \\
 &= 1 - P\{\text{no 3 birthdays match}\} \\
 &= 1 - p(0; \lambda) = 1 - \exp\left(-\frac{n(n-1)(n-2)}{6 \times 365^2}\right) \frac{\lambda^0}{0!} \\
 &= 1 - \exp\left(-\frac{n(n-1)(n-2)}{799350}\right)
 \end{aligned}$$

This probability is graphed in Fig. 4.7.5 for values of  $n = 1, \dots, 150$  and tabulated in Table 4.7.4 for values of  $n = 80-86, 90-96$ , and  $104-110$ .

Examination of Fig. 4.7.3 and Table 4.7.4 reveals that for an even chance for at least three matches, there must be about 84 people and for a 70% chance there must be about 100 people compared to 23 and 30 for at least two matches. This approximation technique can be extended to any number of matches from 2 to 365, whereas exact computation becomes extremely cumbersome, if not impossible, for anything more than two matches.

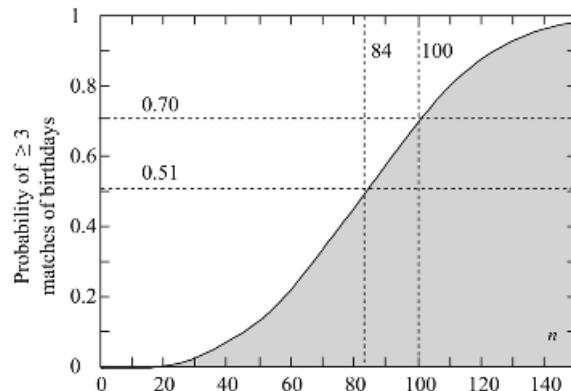


FIGURE 4.7.5

TABLE 4.7.4 Poisson approximation for three matches

#	Probability	#	Probability	#	Probability
80	0.460278	90	0.58597	104	0.745102
81	0.472929	91	0.598231	105	0.755146
82	0.485593	92	0.610393	106	0.764978
83	0.498257	93	0.622444	107	0.774593
84	0.510911	94	0.634375	108	0.783987
85	0.523543	95	0.646176	109	0.793155
86	0.536141	96	0.657838	110	0.802095

## 4.8 LOGARITHMIC DISTRIBUTION

### 4.8.1 Finite Law (Benford's Law)

#### *First Significant Digit*

Most physical data depend on units. Examples are rainfall in inches, land area in acres, blood pressure in millimeters of mercury, stock prices in dollars, distances in miles, and volume in cubic feet. If clusters of these quantities can be described by a probability distribution  $P(x)$ , then such a probability distribution must be invariant under a scale change, so that

$$P(kx) = KP(x) \quad (4.8.1)$$

where  $k$  and  $K$  are scale factors. Since  $\int P(x)dx = 1$ , we have  $\int P(kx)dx = 1/k$ , and hence  $K = 1/k$  is the normalization factor. Thus

$$P(kx) = \frac{1}{k} P(x) \quad \text{or} \quad kP(kx) = P(x) \quad (4.8.2)$$

Differentiating Eq. (4.8.2) with respect to  $k$ , we obtain

$$P(kx) + k \frac{d}{d(kx)} P(kx) \frac{d(kx)}{dk} = 0$$

Substituting  $k = 1$  in the equation above, we have

$$x \frac{d}{dx} P(x) = -P(x)$$

The solution for Eq. (4.8.3) is

$$P(x) = \frac{1}{x} \quad (4.8.4)$$

Equation (4.8.4) can be considered as a probability distribution only for finite values of  $x$  since the integral for infinite values of  $x$  diverges. We shall assume that the data are defined to some arbitrary base  $B$  rather than base 10. If  $X_1$  is the random variable corresponding to the occurrence of the first digit  $d$  to the base  $B$ , then the probability of occurrence of  $P_{X_1}(d)$  is given by

$$\begin{aligned} P_{X_1}(d) &= \frac{\int_d^{d+1} P(x) dx}{\int_1^B P(x) dx} = \frac{\int_d^{d+1} \frac{1}{x} dx}{\int_1^B \frac{1}{x} dx}; \quad d = 1, 2, \dots, B-1 \\ &= \frac{\ln(d+1) - \ln(d)}{\ln(B) - \ln(1)} = \frac{\ln\left(1 + \frac{1}{d}\right)}{\ln(B)} \end{aligned} \quad (4.8.5)$$

We can show that indeed Eq. (4.8.5) is a probability function as follows:

$$\begin{aligned} \sum_1^{B-1} \ln\left(1 + \frac{1}{d}\right) &= \ln\left(1 + \frac{1}{1}\right) + \ln\left(1 + \frac{1}{2}\right) + \dots + \ln\left(1 + \frac{1}{B-1}\right) \\ &= \ln \prod_1^{B-1} \left(1 + \frac{1}{d}\right) = \ln\left(\frac{2}{1} \frac{3}{2} \frac{4}{3} \dots \frac{B}{B-1}\right) = \ln(B) \end{aligned} \quad (4.8.6)$$

Thus  $P_{X_1}(d)$  over all digits 1 to  $B-1$  sums to 1.

If the base  $B$  is 10, the generic  $\ln$  becomes  $\log$ , and the probability of the first significant digit  $d$  from 1 to 9 is,

$$P_{X_1}(d) = \log\left(1 + \frac{1}{d}\right); \quad d = 1, 2, \dots, 9 \quad (4.8.7)$$

This equation is called *Benford's law*, first stated in 1938 by Benford [5]. It gives the probability of occurrence of the first significant digit in a cluster of distributions. Incidentally, 0 cannot be the first significant digit. It can be a significant digit only after the first one.

**Example 4.8.1** In a cluster of numbers from different scale factors, we want to find the occurrence of 1, 2, ..., 9 as the first significant digit. Using Eq. (4.8.7), we have

$$P_{X_1}(1) = \log\left(1 + \frac{1}{1}\right) = 0.30103$$

The probability of 1 as the first significant digit is 30.1% and not a uniform distribution of 11.1% as one would expect! The probability of 2 is 17.6%. The probabilities of other significant digits are tabulated in Table 4.8.1 and plotted in Fig. 4.8.1. They show that 9 as the first significant digit has the least probability of occurrence of 4.6%.

**Example 4.8.2** As a realistic example we have taken the values of physical constants (73 in number), in Chapter 17 of Ref. 18 and tabulated the significant digits from 1 to 9 in

**TABLE 4.8.1** Benford distribution

$d$	$f_D(d)$
1	0.30103
2	0.176091
3	0.124939
4	0.09691
5	0.079181
6	0.066947
7	0.057992
8	0.051153
9	0.045757

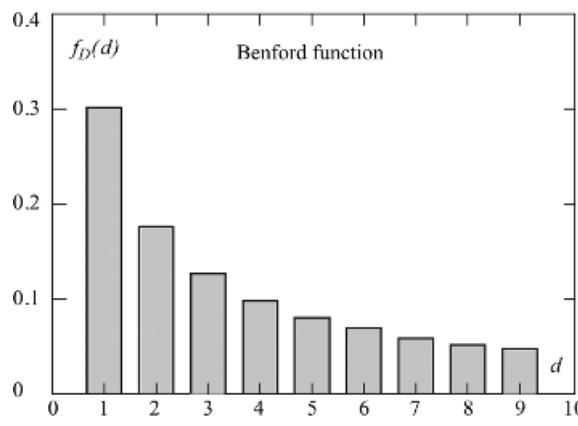
**FIGURE 4.8.1**

Table 4.8.2. The physical constants have totally different units. The probabilities of the significant digits are shown in Fig. 4.8.2 along with the graph of Benford's law.

The table and the graphs show that even for a small sample of 73, the probabilities follow the logarithmic distribution reasonably closely except for the number 3.

In conclusion, in a mixed population of different scale factors, the probabilities of the digits 1 to 9 are logarithmically distributed and are not uniform—a surprising result!

**TABLE 4.8.2**

$k$	$P(k)$
1	0.342
2	0.205
3	0.055
4	0.096
5	0.096
6	0.055
7	0.027
8	0.041
9	0.082

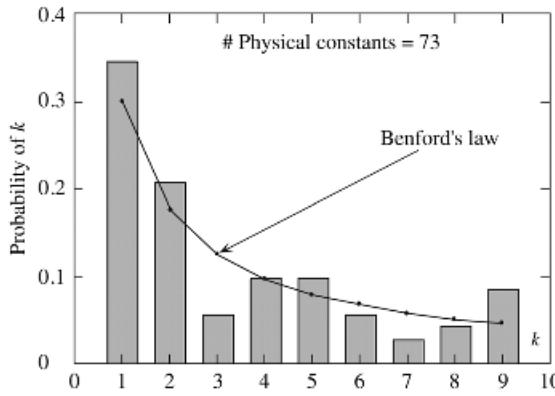


FIGURE 4.8.2

### Second Significant Digit

We shall now compute the occurrence of  $0, 1, \dots, B - 1$  as the second significant digit. Note that 0 can occur as a subsequent significant digit after the first. We will only consider numbers of only first-order magnitude since the pattern will be the same for any other order. With base  $B$  and the total number of digits being  $B - 1$ , the occurrences of 2 as the second significant digit is  $1.2, 2.2, \dots, (B - 1) \cdot 2$ . Figure 4.8.3 shows in logarithmic scale the arrangement of numbers for base 10.

Thus, the probability of 2 as the second significant digit occurring after the first digit 1 is given by an equation similar to Eq. (4.8.5):

$$P\{2 \text{ as second digit after first digit } 1\} = \frac{\ln(1.3) - \ln(1.2)}{\ln(B) - \ln(1)} = \frac{\ln\left(\frac{13}{12}\right)}{\ln(B)} \quad (4.8.8)$$

If  $X_2$  is the random variable corresponding to the second digit, then the probability of 2 occurring after all the  $B - 1$  first digits is given by

$$P_{X_2}(2) = \frac{\ln\left(\frac{13}{12}\right) + \ln\left(\frac{23}{22}\right) + \dots + \ln\left(\frac{(B-1)3}{(B-1)2}\right)}{\ln(B)} \quad (4.8.9)$$

Summing Eq. (4.8.10) over all the first digits  $d = 0, \dots, B - 1$ , we obtain the probability

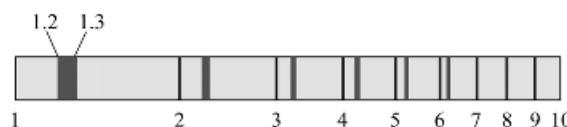


FIGURE 4.8.3

of occurrence of the second significant digit,  $P_{X2}(d)$ , as

$$\begin{aligned} P_{X2}(d) &= \frac{1}{\ln(B)} \sum_{k=1}^{B-1} \ln\left(1 + \frac{1}{kB + d}\right); \quad d = 0, 1, \dots, B-1 \\ &= \frac{1}{\ln(B)} \ln\left(\prod_{k=1}^{B-1} \left(1 + \frac{1}{kB + d}\right)\right) \end{aligned} \quad (4.8.10)$$

It can be shown that

$$\sum_{d=0}^{B-1} \ln\left[\prod_{k=1}^{B-1} \left(1 + \frac{1}{kB + d}\right)\right] = \ln\left[\prod_{d=0}^{B-1} \prod_{k=1}^{B-1} \left(1 + \frac{1}{kB + d}\right)\right] = \ln(B) \quad (4.8.11)$$

and hence,  $P_{X2}(d)$  is a probability function.

We can extend the preceding analysis to obtain a general probability function for the  $(n+1)$ st significant digit after the first one digit  $n = 0, 1, \dots, B-1$  as shown below.

The third digit after the first one can be obtained from Eq. (4.8.10). The occurrence of 3 as the third significant digit is 103, 113, ..., 1(B-1)3; 203, 213, ..., 2(B-1)3; ...; (B-1)03, (B-1)13, ..., (B-1)(B-1)3. Thus, there are  $B(B-1)$  possible ways in which the digit 3 or any other digit  $d = 0, 1, \dots, B-1$  can occur as the third significant number. An equation analogous to Eq. (4.8.9) can be written for the probability of 3 occurring as the third significant digit:

$$P_{X3}(3) = \frac{1}{\ln(B)} \sum_{k=B}^{B^2-1} \ln\left(1 + \frac{1}{kB + 3}\right) \quad (4.8.12)$$

Similar to Eq. (4.8.10), we can write the probability of any digit  $n$  occurring as the third significant digit:

$$\begin{aligned} P_{X3}(d) &= \frac{1}{\ln(B)} \sum_{k=B}^{B^2-1} \ln\left(1 + \frac{1}{kB + d}\right) \\ &= \frac{1}{\ln(B)} \cdot \ln\left(\prod_{k=B}^{B^2-1} 1 + \frac{1}{kB + d}\right); \quad d = 0, \dots, B-1 \end{aligned} \quad (4.8.13)$$

We can extend Eq. (4.8.13) to all  $(n+1)$  significant digits after the first with  $n = 1, 2, \dots, B-1$ :

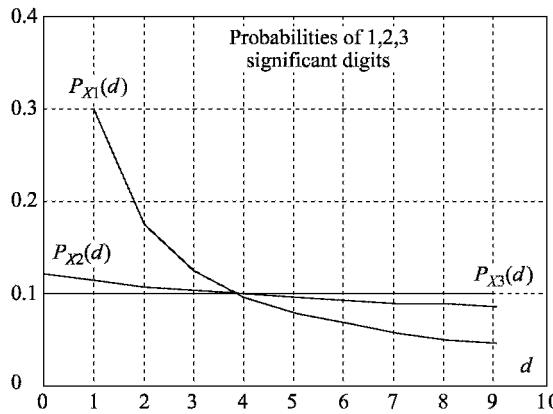
$$\begin{aligned} P_{X(n+1)}(d) &= \frac{1}{\ln(B)} \sum_{k=B^{n-1}}^{B^n-1} \ln\left(1 + \frac{1}{kB + d}\right) \\ n &= 1, 2, \dots, B-1; \quad d = 0, \dots, B-1 \end{aligned} \quad (4.8.14)$$

**Example 4.8.3** Using Eq. (4.8.14) with the base  $B = 10$ , we will find the occurrence of the digits 0, 1, ..., 9 occurring as the first, second, ..., tenth significant digit, noting that the probability of occurrence of 0 as the first significant digit is 0. Equation (4.8.14) becomes

$$\begin{aligned} P_{X(n+1)}(d) &= \frac{1}{\ln(10)} \sum_{k=10^{n-1}}^{10^n-1} \ln\left(1 + \frac{1}{k \cdot 10 + d}\right) \\ n &= 0, 1, \dots, 9; \quad d = 0, 1, \dots, 9 \end{aligned}$$

**TABLE 4.8.3 Benford distribution for 1–9 significant digits**

$d$	$P_{X1}(d)$	$P_{X2}(d)$	$P_{X3}(d)$	$P_{X4}(d)$	$P_{X5}(d)$	$P_{X6}(d)$	$P_{X7}(d)$
0	0	0.119679	0.101784	0.100176	0.100018	0.100002	0.1
1	0.30103	0.11389	0.101376	0.100137	0.100014	0.100001	0.1
2	0.176091	0.108821	0.100972	0.100098	0.10001	0.100001	0.1
3	0.124939	0.10433	0.100573	0.100059	0.100006	0.100001	0.1
4	0.09691	0.100308	0.100178	0.100019	0.100002	0.1	0.1
5	0.079181	0.096677	0.099788	0.09998	0.099998	0.1	0.1
6	0.066947	0.093375	0.099401	0.099941	0.099994	0.099999	0.1
7	0.057992	0.090352	0.099019	0.099902	0.09999	0.099999	0.1
8	0.051153	0.08757	0.098641	0.099863	0.099986	0.099999	0.1
9	0.045757	0.084997	0.098267	0.099824	0.099982	0.099998	0.1

**FIGURE 4.8.4**

The values of  $P_{X(n+1)}(d)$  have been tabulated in Table 4.8.3 for values of  $d = 0, \dots, 9$  and  $n = 1, \dots, 6$ .

It is interesting to note that the first significant digit is considerably different from uniform distribution. As the level of significance increases, the distribution tends to be more and more uniform and the significant digit is practically uniform. In fact, the third significant digit is almost a uniform distribution as shown in Fig. 4.8.4.

Benford's law is being used widely by the Internal Revenue Service for detecting fraudulent tax returns.

## 4.8.2 Infinite Law

We will also present two versions of a logarithmic distribution where the number of outcomes are infinite:

*Version 1.* The first version is defined by

$$p(k, \theta) = \frac{-\theta^k}{k \ln(1 - \theta)}; \quad 0 < \theta < 1, \quad k \geq 1 \quad (4.8.15)$$

TABLE 4.8.4

$k$	$P(\theta, k)$
1	0.390865
2	0.175889
3	0.105534
4	0.071235
5	0.051289
6	0.038467
7	0.029675
8	0.023369
9	0.018695
10	0.015143

where  $\theta$  is the shape parameter and  $k$  are the infinite outcomes. Here  $\theta$  is between 0 and 1, and  $k$  is greater than or equal to 1. We will show that this is a probability distribution by summing over all  $k$ :

$$\sum_{k=1}^{\infty} p(k, \theta) = \sum_{k=1}^{\infty} \frac{-\theta^k}{k \ln(1-\theta)} = \frac{1}{\ln(1-\theta)} \sum_{k=1}^{\infty} -\frac{\theta^k}{k} = \frac{\ln(1-\theta)}{\ln(1-\theta)} = 1 \quad (4.8.16)$$

In Eq. (4.8.15), as  $\theta$  tends to zero,  $p(k, \theta)$  tends to 1 as shown below:

$$\lim_{\theta \rightarrow 0} p(k, \theta) = \lim_{\theta \rightarrow 0} \frac{-\theta^k}{k \ln(1-\theta)} = \lim_{\theta \rightarrow 0} \frac{k\theta^{k-1}(1-\theta)}{k} = 1 \quad (4.8.17)$$

The probability mass function (pmf) given by Eq. (4.8.15) is tabulated for  $k = 1, \dots, 10$  and  $\theta = 0.9$  in Table 4.8.4 and shown in Fig. 4.8.5. This logarithmic distribution is used to model the number of items purchased by a customer in a supermarket in a given amount of time.

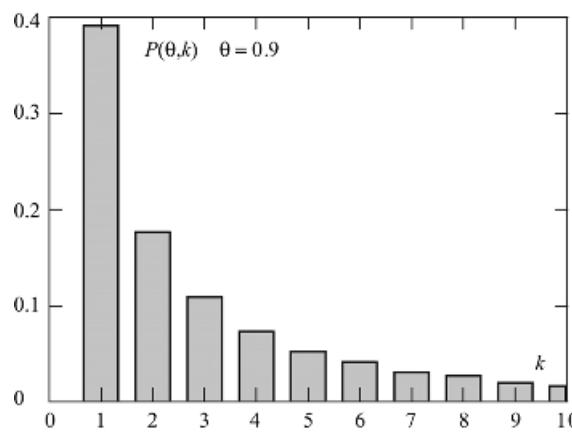


FIGURE 4.8.5

*Version 2.* The second version of this logarithmic distribution is obtained by substituting  $\lambda = (1 - \theta)$  in Eq. (4.8.15) to obtain

$$p(k,\lambda) = \frac{-(1 - \lambda)^k}{k \ln(\lambda)} : 0 < \lambda < 1, k \geq 1 \quad (4.8.18)$$

Here, as  $\lambda$  tends to 1,  $p(k,\lambda)$  tends to 1 as shown below:

$$\lim_{\theta \rightarrow 0} p(k,\lambda) = \lim_{\theta \rightarrow 0} \frac{-(1 - \lambda)^k}{k \ln(\lambda)} = \lim_{\theta \rightarrow 0} \frac{k \lambda^{k-1} \lambda}{k} = 1 \quad (4.8.19)$$

## 4.9 SUMMARY OF DISCRETE DISTRIBUTIONS

Assumptions:  $P\{\text{success}\} = p$ ;  $P\{\text{failure}\} = q$ ;  $p + q = 1$ .

Distribution	Type	Definition
Bernoulli		$b(p,q) = \begin{cases} p; & k = 1 \\ q; & k = 0 \\ 0; & \text{otherwise} \end{cases}$
Binomial		$b(k; n, p) = \begin{cases} \binom{n}{k} p^k q^{n-k}; & k = 0, 1, \dots, n \\ 0; & \text{otherwise} \end{cases}$
Multinomial		$p(k_1, k_2, \dots, k_m) = \binom{n}{k_1, k_2, \dots, k_m} p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}$ $\sum_{i=1}^m k_i = n; \quad \sum_{i=1}^m p_i = 1$
Geometric		$g(k, p) = \begin{cases} q^{k-1}; & k = 1, 2, \dots \\ 0; & k \leq 0 \end{cases}$
Negative binomial: Pascal		$nb(p) = \begin{cases} \binom{n-1}{k-1} p^k q^{n-k}; & k = k, k+1, \dots \\ 0; & \text{otherwise} \end{cases}$
Hypergeometric		$p(n, k) = \begin{cases} \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}; & k = 0, 1, \dots, n \\ 0; & \text{otherwise} \end{cases}$

(Continued)

Distribution Type	Definition
Poisson	$p(k, \lambda) = \begin{cases} e^{-\lambda} \frac{\lambda^k}{k!}; & k = 0, 1, \dots \\ 0; & \text{otherwise} \end{cases}$
Logarithmic–Benford	$p(d) = \frac{1}{N} \log \left( 1 + \frac{1}{d} \right); \quad d = 1, \dots, 10^N - 1$
Logarithmic 1—semiinfinite	$p(k, \theta) = \frac{-\theta^k}{k \ln(1 - \theta)}; \quad 0 < \theta < 1, k \geq 1$
Logarithmic 2—semiinfinite	$p(k, \theta) = \frac{-(1 - \theta)^k}{k \ln(\theta)}; \quad 0 < \theta < 1, k \geq 1$

# Random Variables

## 5.1 DEFINITION OF RANDOM VARIABLES

What is a random variable? An outlandish definition would be that it is neither random nor a variable! The real definition of a random variable is that it is a *function*  $X$  that assigns a rule of correspondence for *every* point  $\xi$  in the sample space  $S$  called the *domain*, a unique value  $X(\xi)$  on the real line  $R$  called the *range*. Let  $\mathcal{F}$  be the field associated with the sample space and  $\mathcal{F}_X$  be the field associated with the real line. The random variable  $X$  induces a probability measure  $P_X$  in  $R$  and hence  $X$  is a mapping of the probability space  $\{S, \mathcal{F}, P\}$  to the probability space  $\{R, \mathcal{F}_X, P_X\}$  as shown below:

$$X: \{S, \mathcal{F}, P\} \longrightarrow \{R, \mathcal{F}_X, P_X\} \quad (5.1.1)$$

Consider an event  $D \subset S \in \mathcal{F}$ . The random variable  $X$  maps every point  $\xi_i$  in the event  $D$  to points  $x_i$  in the event  $I_x$ , called the *image* of  $D$  under  $X$ , where  $I_x \subset R \in \mathcal{F}_X$ .  $X$  can be a random variable only if the *inverse image*  $X^{-1}(I_x)$  belongs to the field  $\mathcal{F}$  of subsets of  $S$ , and hence it must be an event. The mapping and the inverse mapping are shown in Fig. 5.1.1. With this restriction we should be able to find the induced probability measure  $P_X$  in terms of the probability measure  $P$  as follows:

$$P_X\{I_x\} = P\{X^{-1}(I_x)\} = P\{D\} = P\{\xi : X(\xi) \in I_x\} \quad (5.1.2)$$

Since  $I_x$  belongs to the field  $\mathcal{F}_X$ , on the real line it may consist of sets of the form  $\{-\infty < \eta \leq x\}$ ,  $\{x < \eta < \infty\}$ ,  $\{x_1 < \eta \leq x_2\}$ ,  $\{\eta = x\}$ . Out of these sets we define  $I_x$  by

$$I_x = \{-\infty < \eta \leq x\} \quad (5.1.3)$$

All the other sets  $\{x < \eta < \infty\}$ ,  $\{x_1 < \eta \leq x_2\}$ ,  $\{\eta = x\}$  can be expressed in terms of the defined  $I_x$  as follows:

$$\begin{aligned} \bar{I}_x &= \{x < \eta \leq \infty\} \\ I_{x_2} - I_{x_1} &= \{-\infty < \eta \leq x_2\} - \{-\infty < \eta \leq x_1\} = \{x_1 < \eta \leq x_2\} \\ \lim_{\Delta x \rightarrow 0+} \{I_{x+\Delta x} - I_x\} &= \{\eta = x\} \end{aligned}$$

With the definition of  $I_x$  given by Eq. (5.1.3), we can write Eq. (5.1.2) as follows:

$$P_X\{I_x\} = P\{\xi : X(\xi) \in I_x\} = P\{\xi : X(\xi) \leq x\} \quad (5.1.4)$$

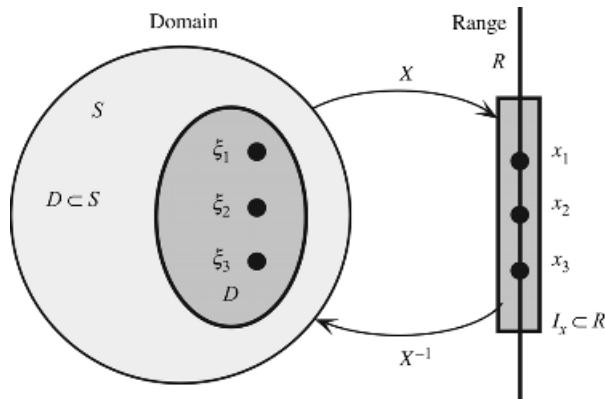


FIGURE 5.1.1

We now define the quantity  $P\{\xi : X(\xi) \leq x\}$  as the *cumulative distribution function* (cdf)  $F_X(x)$ , and rewrite Eq. (5.1.4) in abbreviated notation as

$$P\{\xi : X(\xi) \in I_x\} = P\{X \leq x\} = F_X(x) \quad (5.1.5)$$

Equation (5.1.5) converts from a cumbersome set function  $P\{X \leq x\}$  into a convenient point function  $F_X(x)$ . The value  $x$  scans the entire real line, that is,  $-\infty < x < \infty$ , and  $F_X(x)$  must be defined for all  $x$  on the real line.

Note that  $\{X \leq x\}$  in Eq. (5.1.5) is meaningless unless it is defined carefully, because  $X$  is a function and a function without an argument cannot be less than a number! For example, we cannot have an exponential less than 2, unless we specify the argument of the exponential. The abbreviated notation in Eq. (5.1.5) is to be defined as *the probability of the set of all points  $\xi \in S$  such that the number  $X(\xi)$  is less than or equal to the number  $x$* . The random variable  $X$  should not be confused with the value  $x$  that the random variable  $X$  can take any value, for example,  $F_X(y) = P\{X \leq y\}$ .

Having defined the cumulative distribution function (cdf), we now define the *probability density function* (pdf) as the derivative of the cdf  $F_X(x)$  with respect to  $x$ :

$$\begin{aligned} f_X(x) &= \lim_{\Delta x \rightarrow \infty} \frac{P(X \leq x + \Delta x) - P(X \leq x)}{\Delta x} \\ &= \lim_{\Delta x \rightarrow \infty} \frac{F_X(x + \Delta x) - F_X(x)}{\Delta x} = \frac{d}{dx} F_X(x) \end{aligned} \quad (5.1.6)$$

The cdf  $F_X(x)$  can be given in terms of the pdf  $f_X(x)$  as follows:

$$F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi \quad (5.1.7)$$

Whereas the distribution function  $F_X(x)$  is a probability measure, the density function  $f_X(x)$  is not a probability measure unless it is multiplied by the infinitesimal  $\Delta x$  to yield

$$f_X(x)\Delta x = P(x < X \leq x + \Delta x) \quad (5.1.8)$$

## 5.2 DETERMINATION OF DISTRIBUTION AND DENSITY FUNCTIONS

Most of the difficulty in determining distribution functions for random variables can be avoided if the following step by step procedure is followed.

*Step 1.* Since the random variable is a mapping from the sample space  $S$  to the real line, the mapping diagram is drawn connecting every point  $\xi$  in the sample space to the points on the real line  $R$ .

*Step 2.* The regions for  $x$  on the entire real line from the mapped values are determined.

The regions may correspond to the different subsets of  $S$  that map to the real line  $R$ . The set of points  $I_x$  will be equal to the right closed set  $(-\infty, x]$  or,  $I_x = \{-\infty < \eta \leq x\}$ . Note that whereas  $I_x$  is a right closed subset, the region of definition of  $x$  is a subset of the real line  $R$ .

*Step 3.* On the sample space  $S$ , the probability of the set of all points that map into  $I_x$  for every region of  $x$  in step 2 is determined.

*Step 4.* The cdf  $F_X(x) = P\{X \leq x\}$  is graphed for all the regions of  $x$  in step 2 on the entire real line.

*Step 5.* The derivative of  $F_X(x)$  determines the density function  $f_X(x)$ .

**Example 5.2.1** A die is tossed, and the random variable  $X$  is defined by the amount won (+) or lost (-) on the face of the die as shown in Table 5.2.1. We have to find the cdf  $F_X(x)$  and the pdf  $f_X(x)$ .

We will follow the steps to find the probability functions.

*Step 1.* The mapping diagram (Fig. 5.2.1) is drawn with positive numbers indicating win and negative numbers indicating loss.

*Step 2.* From the mapping diagram the regions of  $x$  are (a)  $x \leq -9$ , (b)  $-9 < x \leq -4$ , (c)  $-4 < x \leq 5$ , (d)  $5 < x \leq 8$ , and (e)  $x > 8$ . The corresponding set  $I_x$  on the real line for all the 5 regions is  $I_x = \{-\infty < \eta \leq x\}$ .

*Step 3.* We will find the all the points in the sample space that map into  $I_x$  for every region of  $x$ :

- $x \leq -9$ : Since no points in  $S$  map into  $I_x$  (Fig. 5.2.1a), we have  $F_X(x) = 0$ .
- $-9 < x \leq -4$ : In this region (Fig. 5.2.1b), only one point {3} maps into  $I_x$ . Since  $P\{3\} = \frac{1}{6}$ ,  $F_X(x) = \frac{1}{6}$ .
- $-4 < x \leq 5$ : Here the points {3,2,6} from  $S$  map into  $I_x$  (Fig. 5.2.1c). Hence  $F_X(x) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$ .
- $5 < x \leq 8$ : In this region (Fig. 5.2.1d) the points {3,2,6,5} in  $S$  map in to  $I_x$ . Hence  $F_X(x) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$

TABLE 5.2.1

Pips on the Die	Win or Loss \$
1,4	+8
2,6	-4
3	-9
5	+5

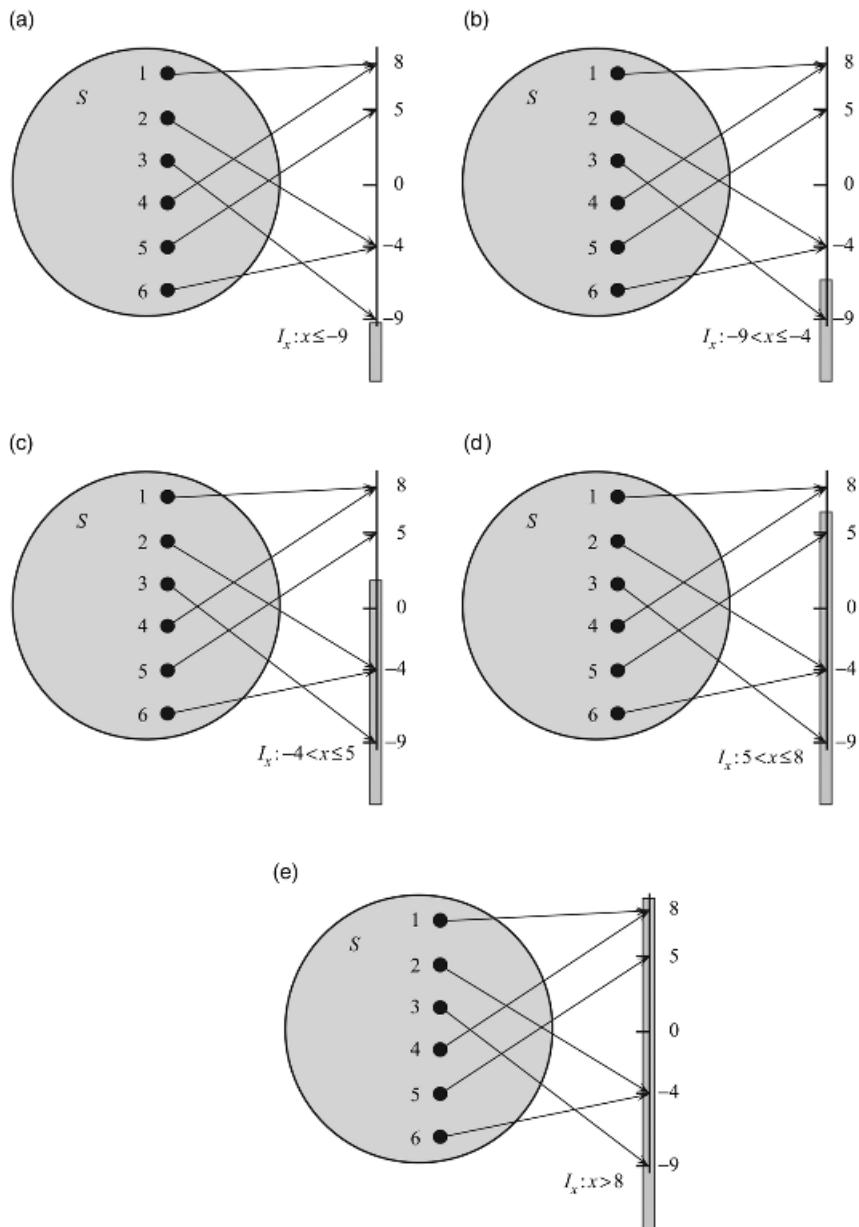


FIGURE 5.2.1

- (e)  $x > 8$ : In this region (Fig. 5.2.1e) all six points, that is, the entire sample space  $S$ , map into  $I_x$ . Hence  $F_X(x) = 1$ .

In terms of unit step functions we can write the cdf  $F_X(x)$  as follows:

$$F_X(x) = \frac{1}{6}u(x+9) + \frac{1}{3}u(x+4) + \frac{1}{6}u(x-5) + \frac{1}{3}u(x-8)$$

*Step 4.* The cdf  $F_X(x)$  can be graphed as shown in Fig. 5.2.2.

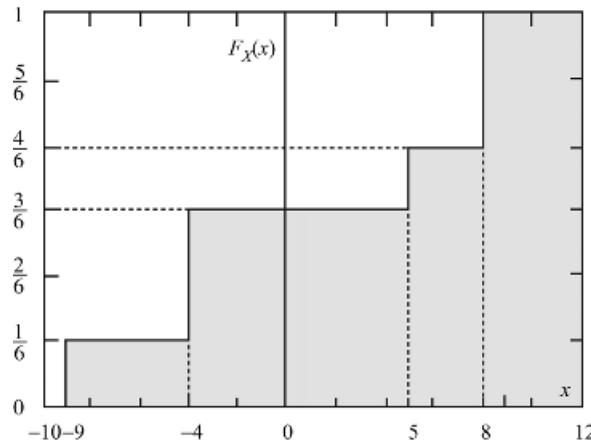


FIGURE 5.2.2

*Step 5.* We can now find the density function  $f_X(x)$  by differentiating the distribution function  $F_X(x)$ , bearing in mind that the differentiation of the unit step is the Dirac delta function. Performing the indicated differentiation, we obtain

$$f_X(x) = \frac{1}{6}\delta(x + 9) + \frac{1}{3}\delta(x + 4) + \frac{1}{6}\delta(x - 5) + \frac{1}{3}\delta(x - 8)$$

Note that the sum of the strengths of the impulse function adds to 1. The pdf  $f_X(x)$  is shown in Fig. 5.2.3.

From Fig. 5.2.2 we can write the following probabilities:

- $P\{X = -4\} = F_X(-4+) - F_X(-4-) = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}$
- $P\{X \leq -4\} = F_X(-4+) - F_X(-\infty) = \frac{1}{2} - 0 = \frac{1}{2}$
- $P\{-5 < X \leq 4\} = F_X(-4+) - F_X(-5-) = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}$
- $P\{-4 < X \leq 4\} = F_X(-4+) - F_X(-4-) = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}$
- $P\{5 \leq X < 8\} = F_X(8-) - F_X(5+)$  (Note closure on the left)  
 $= 12/3 - 2/3 = 0$

The example above shows that the mapping is from a discrete sample space  $S$  to discrete points on the real line  $R$ .

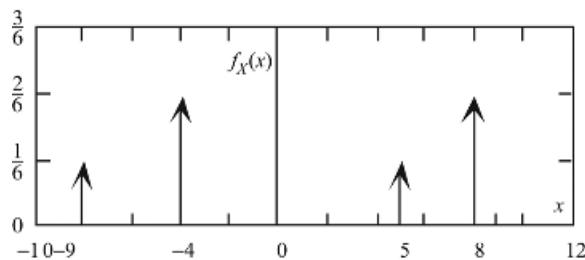


FIGURE 5.2.3

TABLE 5.2.2

Time Interval	Win or Loss (\$)
6:00–6:15	+2
6:15–6:45	-4
6:45–7:15	-4
7:15–8:00	+6

**Example 5.2.2** A telephone call occurs at random between 6:00 p.m. and 8:00 p.m. Depending on the interval of arrival, the win/loss table is shown in Table 5.2.2. We have to find the cdf  $F_X(x)$  and the pdf  $f_X(x)$ .

Since the phone call is uniformly distributed in 120 min, the probability of a phone call in the region  $(0, x]$  is given by  $x/120$ . We follow the steps as outlined in Example 5.2.1:

*Step 1.* The mapping diagram is shown in Fig. 5.2.4. Note that the mapping is from the real line  $R$  to the real line  $R$ .

*Step 2.* From the mapping diagram the regions of  $x$  are (a)  $x \leq -4$ , (b)  $-4 < x \leq 2$ , (c)  $2 < x \leq 6$ , (d)  $x > 6$ .

*Step 3.*

- (a)  $x \leq -4$ : No points map  $F_X(x) = 0$ .
- (b)  $-4 < x \leq 2$ : The interval that maps is 30 min; hence  $F_X(x) = \frac{60}{120} = \frac{1}{2}$ .
- (c)  $2 < x \leq 6$ : The interval that maps is 45 min; hence  $F_X(x) = \frac{75}{120} = \frac{5}{8}$ .
- (d)  $x > 6$ : The entire sample space of 120 min maps; hence  $F_X(x) = 1$ .

In terms of unit step functions, the cdf  $F_X(x)$  is

$$F_X(x) = \frac{1}{2}u(x + 4) + \frac{1}{8}u(x - 2) + \frac{3}{8}u(x - 6)$$

*Step 4.* The cdf  $F_X(x)$  is graphed in Fig. 5.2.5

*Step 5.* The pdf  $f_X(x)$  is obtained by differentiating  $F_X(x)$ , and the result is

$$f_X(x) = \frac{1}{2}\delta(x + 4) + \frac{1}{8}\delta(x - 2) + \frac{3}{8}\delta(x - 6)$$

$f_X(x)$  is graphed in Fig. 5.2.6.

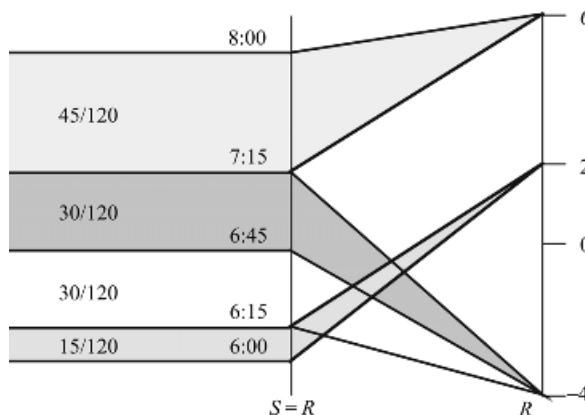


FIGURE 5.2.4

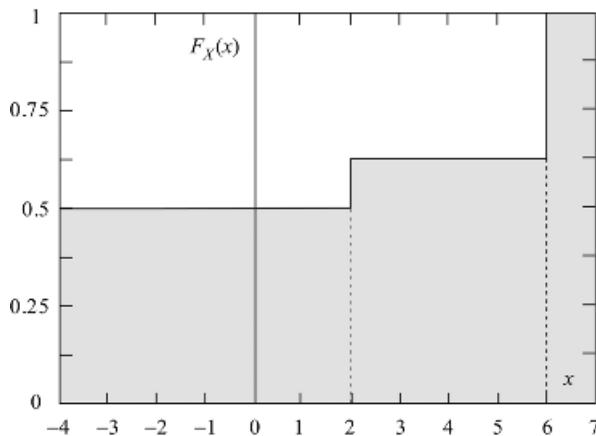


FIGURE 5.2.5

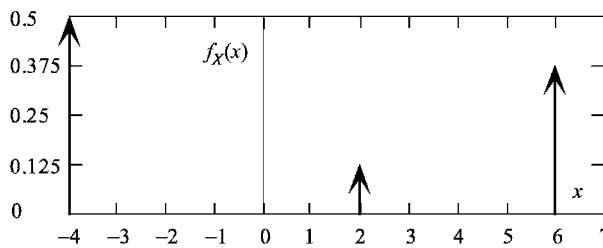


FIGURE 5.2.6

In the preceding example, the sample space is continuous whereas the range space is discrete. Since the random variable takes discrete values, we have a discrete random variable.

**Example 5.2.3** A telephone call occurs at random between 6:00 p.m. and 8:00 p.m. We want to find the cdf  $F_X(x)$  and the pdf  $f_X(x)$ . Here we have a continuous mapping from the real line to the real line and  $X$  is a continuous random variable.

*Step 1.* See the mapping diagram in Fig. 5.2.7.

*Step 2.* From the mapping diagram the regions of  $x$  are (a)  $x \leq 0$ , (b)  $0 < x \leq 120$ , (c)  $x > 120$ , and  $I_x = \{-\infty < \eta \leq x\}$ .

*Step 3*

- (a)  $x \leq 0$ : No points in  $S$  map into  $I_x$ . Hence  $F_X(x) = 0$ .
- (b)  $0 < x \leq 120$ : The call is uniformly distributed in the interval  $(0, 120]$ , and since  $(0, x]$  is an interval mapping into  $I_x$ , we have  $F_X(x) = x/120$ .
- (c)  $x > 120$ : Here the entire sample space maps into  $I_x$ , and hence  $F_X(x) = 1$ .

*Step 4.* The cdf  $F_X(x)$  is graphed in Fig. 5.2.8.

*Step 5.* The density function  $f_X(x)$  is obtained by taking the derivative of  $F_X(x)$ , or  $f_X(x) = \frac{1}{120}$ ,  $0 \leq x \leq 120$ . This is sketched in Fig. 5.2.9.

We found the cdf and pdf for discrete random variables in Examples 5.2.1 and 5.2.2 and for continuous random variables in Example 5.2.3. We will now give an example of a mixed random variable that has both continuous and discrete components.

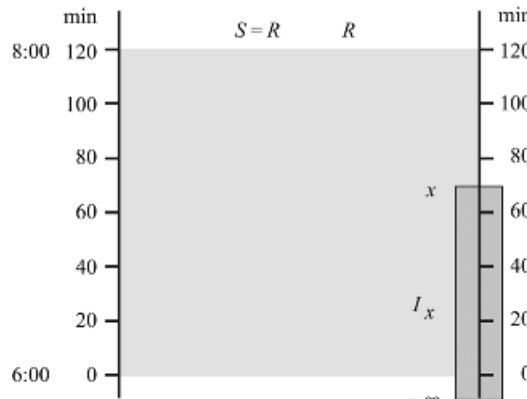


FIGURE 5.2.7

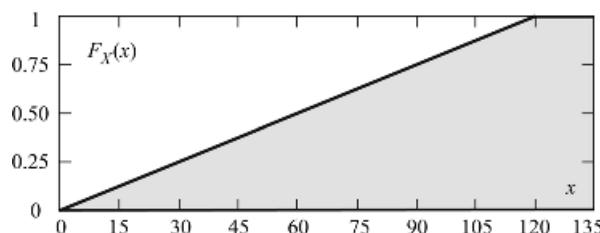


FIGURE 5.2.8

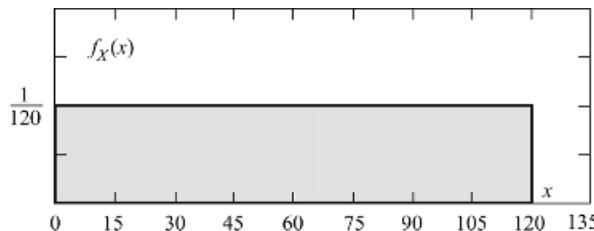


FIGURE 5.2.9

**Example 5.2.4** A telephone call occurs at random between 6:00 p.m. and 8:00 p.m. However, between 6:20 and to 6:40 p.m. calls accumulate and are released exactly at 6:40 p.m. with a similar situation between 7:00 and 7:20 p.m. and are released at exactly 7:20 p.m. We want to find the cdf  $F_X(x)$  and the pdf  $f_X(x)$ .

*Step 1.* The mapping diagram shown in Fig. 5.2.10 is similar to Fig. 5.2.7 except that there are deadspots between 6:20 and 6:40 p.m. and 7:00 and 7:20 p.m.

*Step 2.* From the mapping diagram the regions of  $x$  are (a)  $x \leq 0$ , (b)  $0 < x \leq 20$ , (c)  $20 < x < 40$ , (d)  $x = 40$ , (e)  $40 < x \leq 60$ , (f)  $60 < x < 80$ , (g)  $x = 80$ , (h)  $80 < x \leq 120$ , (i)  $x > 120$ .

*Step 3*

(a)  $x \leq 0$ : No points in  $S$  map into  $I_x$ ; hence  $F_X(x) = 0$ .

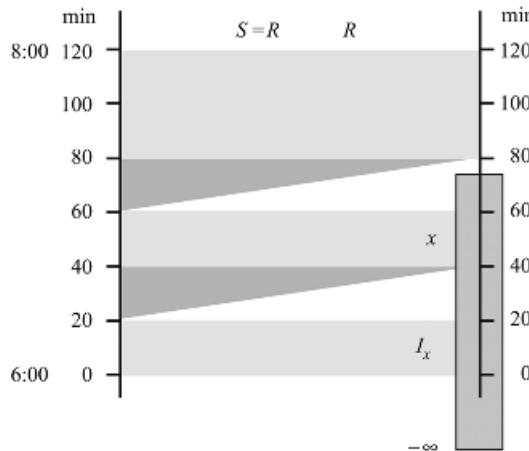


FIGURE 5.2.10

- (b)  $0 < x \leq 20$ : The call is uniformly distributed in the interval  $(0, 120]$ ; hence  $F_X(x) = x/120$ .
- (c)  $20 < x < 40$ : There is a deadspot in this region; hence  $F_X(x) = F_X(20) = \frac{20}{120} = \frac{1}{6}$ .
- (d)  $x = 40$ : There is a jump at this point equal to  $P\{X = 40\} = \frac{1}{6}$ ; hence  $F_X(40) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$ .
- (e)  $40 < x \leq 60$ : Here also  $(0, x]$  is an interval in  $I_x$  and we have  $F_X(x) = x/120$  and  $F_X(40) < \frac{1}{3}$ .
- (f)  $60 < x < 80$ : Here again we have a deadspot; and hence  $F_X(x) = F_X(60) = \frac{60}{120} = \frac{1}{2}$ .
- (g)  $x = 80$ : There is again a jump at this point equal to  $P\{X = 80\} = \frac{1}{6}$ ; hence  $F_X(80) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$ .
- (h)  $80 < x \leq 120$ : Here  $(0, x]$  is an interval in  $I_x$  and we have  $F_X(x) = x/120$  and  $F_X(80) = \frac{2}{3}$ .
- (i)  $x > 120$ : The entire sample space maps into  $I_x$ ; hence  $F_X(x) = 1$

Step 4. The cdf  $F_X(x)$  is graphed in Fig. 5.2.11.

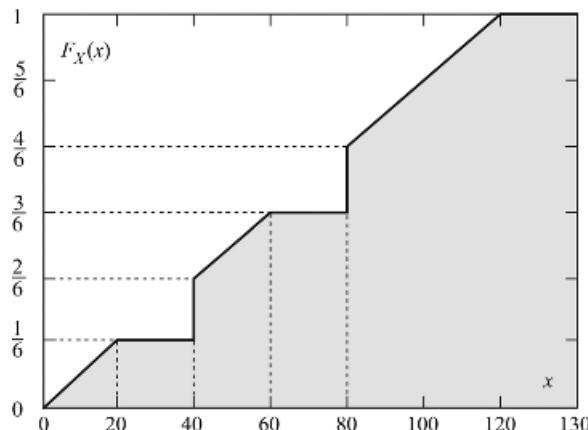


FIGURE 5.2.11

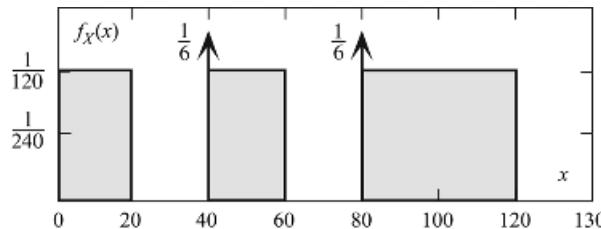


FIGURE 5.2.12

Step 5. The pdf  $f_X(x)$  is the derivative of  $F_X(x)$  and is graphed in Fig. 5.2.12.

From the graph of the cdf  $F_X(x)$  of Fig. 5.2.11, we can obtain the following probabilities:

1.  $P\{X = 40\} = P\{X = 80\} = \frac{1}{6}$ .
2.  $P\{X < 80\} = \frac{1}{2}$ .
3.  $P\{X \leq 80\} = \frac{2}{3}$ .
4.  $P\{40 \leq X \leq 80\} = \frac{1}{3}$ .
5.  $P\{X = 100\} = 0$ .
6.  $P\{50 < X \leq 100\} = \frac{5}{6} - \frac{5}{12} = \frac{5}{12}$ .

## 5.3 PROPERTIES OF DISTRIBUTION AND DENSITY FUNCTIONS

### Cumulative Distribution Function $F_X(x)$

1. From the three examples given in Section 5.2 the cdf of a discrete random variable is a staircase function with jumps.
2.  $F_X(x) \geq 0$ , that is, nonnegative.
3.  $F_X(-\infty) = 0$ ,  $F_X(\infty) = 1$ .
4. If  $x_1 < x_2$ , then  $F_X(x_1) \leq F_X(x_2)$ , that is,  $F_X(x)$  is a nondecreasing function (monotone increasing). For a continuous random variable, if  $x_1 < x_2$ , then  $F_X(x_1) < F_X(x_2)$ , that is,  $F_X(x)$  is an increasing function (strict monotone increasing).
5.  $F_X(x)$  is right-continuous:

$$F_X(x) = \lim_{\varepsilon \rightarrow 0} F_X(x + \varepsilon) \quad \text{for } \varepsilon > 0$$

From this definition, the value that  $F_X(x)$  takes at the point of a jump, is to the right of the jump, and if there is an event  $A$  in the interval  $\{x_1 < X \leq x_2\}$ , then the interval is closed on the righthand side so that we can write  $P\{x_1 < X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1\} = F_X(x_2) - F_X(x_1)$ . This ensures that the point  $x_1$  is not counted twice. Figure 5.3.1 will clarify this point.

$$6. \quad F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi = P\{X \leq x\} \quad (5.3.1)$$

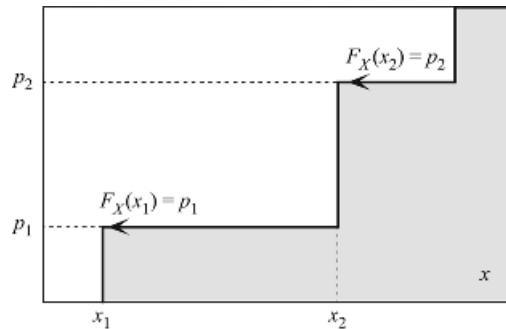


FIGURE 5.3.1

7. At the point  $x = a$  of the jump

$$F_X(x) = \int_{a-}^{a+} f_X(\xi) d\xi = F_X(a+) - F_X(a-) = P\{X = a\} \quad (5.3.2)$$

#### Probability Density Function $f_X(x)$

1.  $f_X(x) \geq 0$ , nonnegative.
2.  $f_X(x)$  consists of the sum of Dirac delta functions for a discrete random variable.
3.  $f_X(x)$  is not a probability measure, but  $f_X(x)\Delta x$  is a probability measure as shown:

$$f_X(x)\Delta x \geq P\{x < X \leq x + \Delta x\} \quad (5.3.3)$$

$$4. f_X(x) \geq \lim_{\Delta x \rightarrow 0} \frac{P\{x < X \leq x + \Delta x\}}{\Delta x} = \frac{d}{dx} F_X(x) \quad (5.3.4)$$

$$5. \int_a^b f_X(\xi) d\xi - F_X(b) - F_X(a) = P\{a < X \leq b\} \quad (5.3.5)$$

#### Probability Mass Functions (pmf's)

We have already come across probability mass functions in connection with discrete random variables. We will now formalize this concept. The probability of a random variable equal to a number is called the *probability mass function* (pmf). In Eq. (5.3.2),  $P\{X = a\}$  is the pmf.

1.  $P\{X = a\}$  is the probability of the jump at the point  $x = a$  in the cdf  $F_X(x)$ . Or,  $P\{X = a\} = F_X(a) - F_X(a-)$ .
2. The pmf  $P\{X = a\}$  is nonzero for discrete random variables at  $x = a$ , where they jump.
3.  $P\{X = a\} = 0$  for continuous random variables, since at any point of continuity  $a$ ,  $F_X(a) - F_X(a-) = 0$ .
4. The density function  $f_X(x)$  for discrete random variables involves summation of the pmf's with Dirac delta functions multiplying the pmf at every point in the range. For example, the pdf of the binomial distribution given by the

probability mass function  $\binom{n}{k} p^k q^{n-k}$  is

$$f_X(x) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \delta(x - k) \quad (5.3.6)$$

The pdf of the Poisson distribution given by the pmf  $e^{-\lambda}(\lambda^k/k!)$  is

$$f_X(x) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta(x - k) \quad (5.3.7)$$

## 5.4 DISTRIBUTION FUNCTIONS FROM DENSITY FUNCTIONS

Given any density function (pdf), the distribution function (cdf) can be found by integrating the pdf over all  $x$  on the real line as given in Eq. (5.3.1). Care must be taken in integrating over various regions along the  $x$  axis.

**Example 5.4.1** We will start with a simple example. The density function  $f_X(x)$  is expressed as

$$f_X(x) = \begin{cases} \frac{k}{4}, & 0 < x \leq 2 \\ \frac{1}{2}, & 2 < x \leq 3 \end{cases}$$

and is shown in Fig. 5.4.1. We have to first find  $k$  and then the distribution function  $F_X(x)$  for all  $x$ .

Integrating  $f_X(x)$  over the range of definition of  $x$ , we have

$$\begin{aligned} \int_0^3 f_X(x) dx &= \int_0^2 \frac{k}{4} dx + \int_2^3 \frac{1}{2} dx = \frac{k}{4} + \frac{1}{2} = 1, \quad \therefore k = 1 \\ f_X(x) &= \frac{1}{4}u(x) + \frac{1}{4}u(x-2) \end{aligned}$$

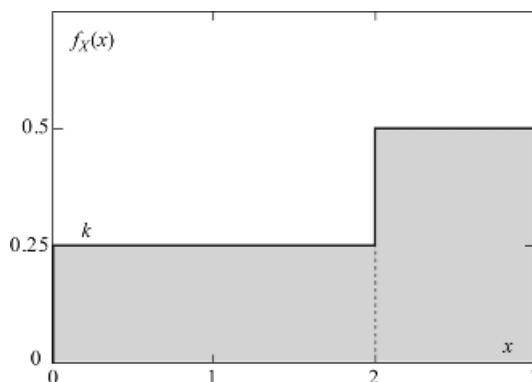


FIGURE 5.4.1

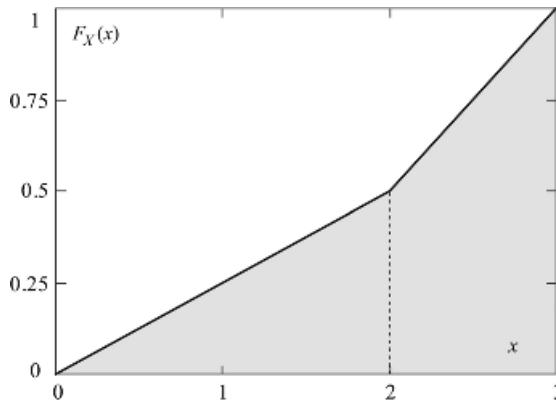


FIGURE 5.4.2

We note that there are four natural ranges for  $x$  given by (1)  $x \leq 0$ , (2)  $0 < x \leq 2$ , (3)  $2 < x \leq 3$ , (4)  $x > 3$ . We will evaluate  $F_X(x)$  in all these ranges:

1.  $x \leq 0$ : Since  $f_X(x)$  is zero,  $F_X(x) = 0$ .
2.  $0 < x \leq 2$ : Using Eq. (5.3.5), we can write

$$F_X(x) = \int_0^x \frac{1}{4} d\xi = \frac{x}{4}$$

3.  $2 < x \leq 3$ : Again using Eq. (5.3.5) and taking care to carry over  $F_X(2)$ , we have

$$F_X(x) = \int_0^2 \frac{1}{4} d\xi + \int_2^x \frac{1}{2} d\xi = \frac{1}{2} + \frac{1}{2}(x - 2) = \frac{1}{2}(x - 1)$$

4.  $x > 3$ : Since  $f_X(x)$  is zero in this range,  $F_X(x) = F_X(3) = 1$ .

$F_X(x)$  is graphed in Fig. 5.4.2.

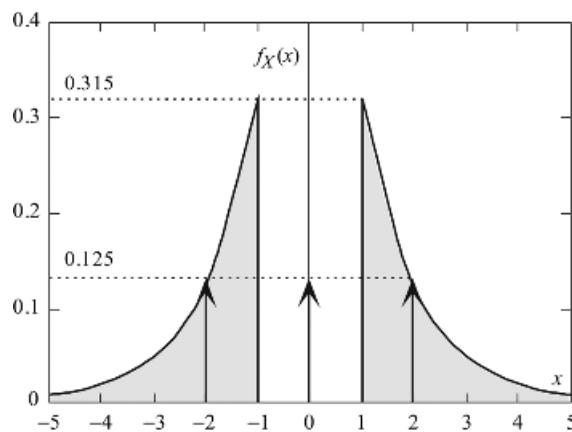


FIGURE 5.4.3

**Example 5.4.2** This is a more involved example that characterizes a mixed random variable. The density function is given by

$$f_X(x) = K\{e^x[u(-x-1) - u(-x-5)] + e^{-x}[u(x-1) - u(x-5)]\} \\ + \frac{1}{8}[\delta(x+2) + \delta(x) + \delta(x-2)]$$

(see Fig. 5.4.3). We will find  $K$  and graph the distribution function  $F_X(x)$  for all  $x$ .

We integrate  $f_X(x)$  over the range  $(-5, 5]$  to evaluate the constant  $K$ . Thus

$$K\left[\int_{-5}^5 e^x dx + \int_1^5 e^{-x} dx\right] + \frac{1}{8} + \frac{1}{8} = 1 \\ 2K[e^{-1} - e^{-5}] + \frac{3}{8} = 1 \quad \therefore K = \frac{5}{16(e^{-1} - e^{-5})}$$

The ranges for  $x$  for integrating  $f_X(x)$  are given by (1)  $x \leq -5$ , (2)  $-5 < x \leq -2$ , (3)  $-2 < x \leq -1$ , (4)  $-1 < x \leq 0$ , (5)  $0 < x \leq 1$ , (6)  $1 < x \leq 2$ , (7)  $2 < x \leq 5$ , (8)  $x > 5$ . We shall now evaluate  $F_X(x)$  in each of these ranges:

1.  $x \leq -5$ : Since  $f_X(x) = 0$ , we have  $F_X(x) = 0$

2.  $-5 < x \leq -2$ :  $F_X(x) = \frac{5}{16(e^{-1} - e^{-5})} \int_{-5}^x e^{\xi} d\xi = \frac{5(e^x - e^{-5})}{16(e^{-1} - e^{-5})}$

3.  $-2 < x \leq -1$ :  $F_X(x) = \frac{5(e^x - e^{-5})}{16(e^{-1} - e^{-5})} + \frac{1}{8}$

4.  $-1 < x \leq 0$ :  $F_X(x) = \frac{5(e^{-1} - e^{-5})}{16(e^{-1} - e^{-5})} + \frac{1}{8} = \frac{7}{16}$

5.  $0 < x \leq 1$ :  $F_X(x) = \frac{7}{16} + \frac{1}{8} = \frac{9}{16}$

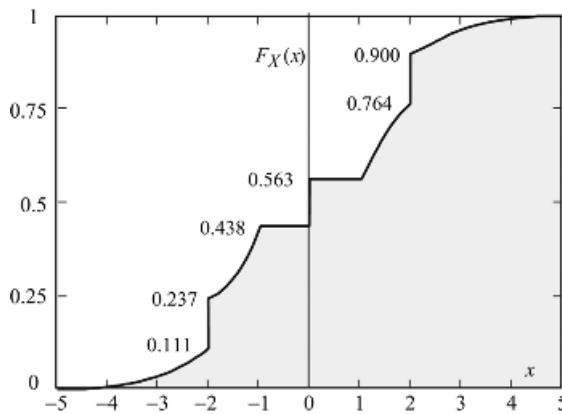


FIGURE 5.4.4

$$6. \quad 1 < x \leq 2: \quad F_X(x) = \frac{9}{16} + \int_1^x e^{-\xi} d\xi = \frac{9}{16} + \frac{5(e^{-1} - e^{-x})}{16(e^{-1} - e^{-5})}$$

$$7. \quad 2 < x \leq 5: \quad F_X(x) = \frac{9}{16} + \frac{5(e^{-1} - e^{-x})}{16(e^{-1} - e^{-5})} + \frac{1}{8}$$

$$8. \quad x > 5: \quad F_X(x) = \frac{9}{16} + \frac{5(e^{-1} - e^{-5})}{16(e^{-1} - e^{-5})} + \frac{1}{8} = 1$$

The graph of  $F_X(x)$  is shown in Fig. 5.4.4.

In the next chapter we will discuss the various continuous distributions.

# Continuous Random Variables and Basic Distributions

## 6.1 INTRODUCTION

In the last chapter we defined a continuous random variable as the mapping from a continuous sample space  $S$  to continuous points on the real line  $R$ . We also discussed some of the properties of continuous pdf  $f_X(x)$  and the continuous cdf,  $F_X(x)$ . In this chapter we will discuss three of the basic continuous distributions: uniform, exponential, and Gaussian. Other distributions will be discussed in the next chapter.

## 6.2 UNIFORM DISTRIBUTION

The basic continuous distributions are uniform, exponential, and Gaussian. We have already seen (in Example 5.2.3) the distribution of a telephone call occurring at random in a given interval. The pdf  $f_X(x)$  of a random variable uniformly distributed in the interval  $(a, b]$ , (notation  $X \sim U(a, b]$ ), is given by

$$f_X(x) = \begin{cases} 0 & x \leq a \\ \frac{1}{b-a} & a < x \leq b \\ 0 & x > b \end{cases} \quad (6.2.1)$$

and the corresponding distribution function  $F_X(x)$  is

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ 1 & x > b \end{cases} \quad (6.2.2)$$

Such distributions occur in transmission of sinusoids where the phase angle will be uniformly distributed in  $(0, 2\pi)$ . The density function is shown in Fig. 6.2.1, and the distribution function is shown in Fig. 6.2.2.

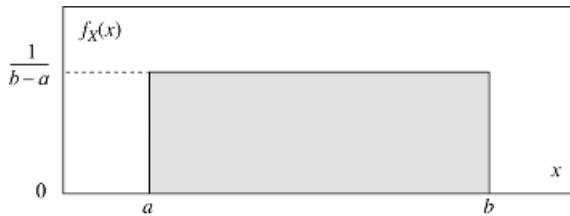


FIGURE 6.2.1

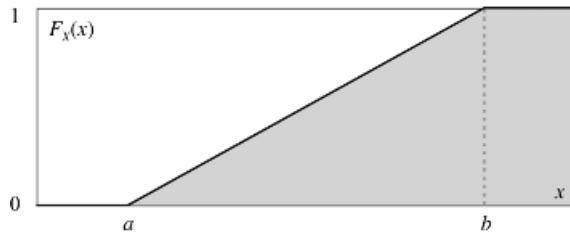


FIGURE 6.2.2

**Example 6.2.1** A sinusoidal signal with random phase angle  $\Theta$ , uniformly distributed between  $(0,2\pi]$  is sent along a communications channel. We want to find the probability that the phase angle is between  $(\pi/4, 3\pi/4]$ .

Since it is uniformly distributed, the pdf is  $1/2\pi$  in  $(0,2\pi]$ , and hence the  $P\{\pi/4, 3\pi/4\}$  is given by

$$P\left\{\frac{\pi}{4} < \theta \leq \frac{3\pi}{4}\right\} = \frac{1}{2\pi} \frac{3\pi - \pi}{4} = \frac{1}{4}$$

### 6.3 EXPONENTIAL DISTRIBUTION

Exponential distributions find wide applicability in queuing theory, reliability theory, and communication theory. It is the continuous analog of the geometric distribution discussed in Chapter 4. The pdf of an exponential distribution is given by

$$f_X(x) = \begin{cases} 0 & x \leq 0 \\ \lambda e^{-\lambda x} & x > 0 \end{cases} \quad (6.3.1)$$

where the parameter  $\lambda$  is the average rate at which events occur. The corresponding cdf  $F_X(x)$  is given by

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases} \quad (6.3.2)$$

The pdf is shown in Fig. 6.3.1, and the cdf is shown in Fig. 6.3.2 for values of  $\lambda = 0.5, 1, 2, 3, 4$ .

The exponential distribution belongs to a class of distributions possessing *memoryless* (*Markov*) properties that we will presently show. A random variable  $X$  is called *memoryless* if it satisfies the conditional probability

$$P\{X > x + x_0 \mid X > x_0\} = P\{X > x\} \quad (6.3.3)$$

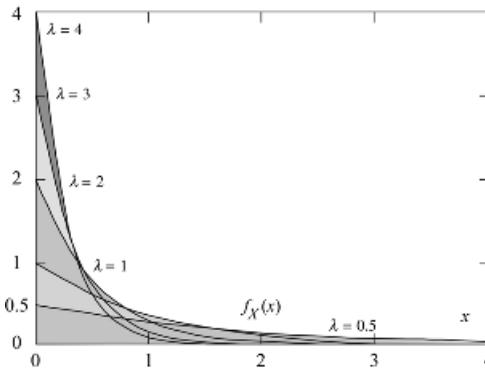


FIGURE 6.3.1

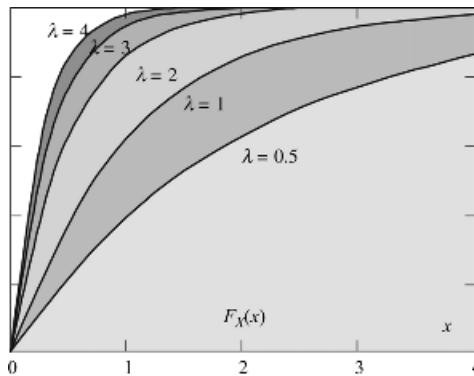


FIGURE 6.3.2

The condition given by Eq. (6.3.3) is equivalent to

$$\frac{P\{X > x + x_0, X > x_0\}}{P\{X > x_0\}} = P\{X > x\} \quad (6.3.4)$$

or

$$P\{X > x + x_0\} = P\{X > x\}P\{X > x_0\}$$

To demonstrate the memoryless property, let  $X$  be a random variable representing the time to failure of a computer with cdf given by Eq. (6.3.2). We are given that the computer has lasted  $x_0$  hours. We want to find the conditional probability that the computer will fail at time  $x$  beyond  $x_0$  given that it has lasted for  $x_0$  hours. We proceed as follows:

$$\begin{aligned} P\{X \leq x + x_0 \mid X > x_0\} &= F_X(x + x_0 \mid X > x_0) \\ &= \frac{P\{X \leq x + x_0, X > x_0\}}{P\{X > x_0\}} = \frac{P\{x_0 < X \leq x + x_0\}}{1 - F_X(x_0)} \\ &= \frac{F_X(x + x_0) - F_X(x_0)}{1 - F_X(x_0)} = \frac{1 - e^{-\lambda(x+x_0)} - 1 + e^{-\lambda x_0}}{1 - 1 + e^{-\lambda x_0}} \\ &= \frac{e^{-\lambda x_0}(1 - e^{\lambda x})}{e^{-\lambda x_0}} = 1 - e^{\lambda x}, \quad x > x_0 \end{aligned} \quad (6.3.5)$$

The conditional distribution  $F_X(x + x_0 \mid X > x_0)$  is the same as  $F_X(x)$ ! The computer “forgets” that it has been operating for  $x_0$  hours and has no memory of previous hours of operation. Hence the exponential distribution is called a *memoryless distribution*.

### **Poisson Arrival Process**

The number  $k$  of events occurring in an interval  $(0, t]$  is Poisson-distributed with  $p(k, \lambda) = e^{-\lambda t}[(\lambda t)^k / k!]$ , and the waiting times for the occurrence from one event to the next are independent and are exponentially distributed as  $e^{-\lambda t}$ .

**Example 6.3.1** A computer company leases its computers. The lifetime of one of the computers is exponentially distributed with  $\lambda = \frac{1}{50,000}$  h. If a person leases the computer for 10,000 h, we have to find the probability that the computer company will have to replace the computer before the lease runs out.

By the memoryless property derived in Eq. (6.3.5), it is immaterial how long it has been leased before, and hence we have

$$P\{\text{computer fails in } t \leq 10,000\} = 1 - e^{-10000\lambda} = 1 - e^{-(1/5)} = 0.1813$$

and the probability that the computer will not be replaced before the lease runs out is

$$P\{\text{lifetime} > 10,000\} = 1 - 0.1813 = 0.8187$$

However, if the failure time  $F_X(x)$  is not exponentially distributed, then the probability of replacement has to be computed, taking into account the prior use of  $x_0$  hours of the computer. This is given by the conditional probability

$$P\{\text{lifetime} > x_0 + 10,000 \mid \text{lifetime} > 10000\} = \frac{1 - F_X(x_0 + 1000)}{1 - F_X(x_0)}$$

which means that we have to know  $x_0$ , the number of hours the computer was operational before the lease.

### **Hazard Rate**

Given that an item has not failed upto time  $t$ , the conditional probability that it will fail in the next small interval of time  $\Delta t$  is determined as follows

$$\begin{aligned} P\{t < X \leq t + \Delta t \mid X > t\} &= \frac{P\{t < X \leq t + \Delta t, X > t\}}{P\{X > t\}} \\ &= \frac{P\{t < X \leq t + \Delta t\}}{P\{X > t\}} = \frac{f_X(t)\Delta t}{1 - F_X(t)} = \beta(t)\Delta t \end{aligned} \quad (6.3.6)$$

where  $F_X(t) = P\{X \leq t\}$  represents the failure distribution and the quantity  $\beta(t) = f_X(t) / [1 - F_X(t)]$  is called the *instantaneous hazard rate*. It is not a probability function as determined by the following analysis:

$$\int_0^t \beta(\xi) d\xi = \int_0^t \frac{f_X(\xi)}{1 - F_X(\xi)} d\xi = \int_0^t -\frac{df_X(x)}{1 - F_X(\xi)} = -\ln[1 - F_X(t)]$$

Since  $\int_0^t \beta(\xi) d\xi \rightarrow \infty$  as  $t \rightarrow \infty$ , we conclude that  $\beta(t)$  is not a probability function. However, solving for  $F_X(t)$ , we obtain

$$-\ln[1 - F_X(t)] = \int_0^t \beta(\xi) d\xi \quad (6.3.7)$$

or

$$1 - F_X(t) = \exp\left[-\int_0^t \beta(\xi) d\xi\right] = 0 \text{ as } t \rightarrow \infty$$

From Eq. (6.3.7) we can obtain the cdf  $F_X(t)$ , given by

$$F_X(t) = 1 - \exp\left[-\int_0^t \beta(\xi) d\xi\right] \quad (6.3.8)$$

Thus,  $\beta(t)$  uniquely determines the failure probability distribution function  $F_X(t)$ . While  $\beta(t)$  is not a probability density function, it is to be noted that the function  $\beta(x)$ , given by

$$\beta(x) = \begin{cases} \frac{f_X(x)}{1 - F_X(t)}, & x > t \\ 0, & x \leq t \end{cases} \quad (6.3.9)$$

is the conditional failure density conditioned on  $x > t$  and  $\beta(t)$  is the value of  $\beta(x)$  at  $x = t$ .

**Example 6.3.2** Physicians are of the opinion that the death rate of a person  $E$  exercising regularly is half that of a person  $S$  who has sedentary habits. We want to compare the probabilities of survival of  $E$  and  $S$  given that the death rate of an exercising person  $E$  is  $\beta_E(t)$  and the rate for a sedentary person  $S$  is  $\beta_S(t)$  with  $\beta_E(t) = \frac{1}{2}\beta_S(t)$ .

Since the rates are conditional failure rates, we will assume that both persons have lived upto the age  $T_1$ , and we want to compare the probabilities of their reaching the age  $T_2$  with  $T_2 > T_1$ . Thus we need to find

$$P\{X_i > T_2 | X_i > T_1\} = \frac{P\{X_i > T_2\}}{P\{X_i > T_1\}} = \frac{1 - F_{X_i}(T_2)}{1 - F_{X_i}(T_1)}$$

where  $X_i = E$  or  $S$ . Substituting from Eq. (6.3.8), we have

$$\frac{1 - F_{X_i}(T_2)}{1 - F_{X_i}(T_1)} = \frac{\exp\left[-\int_0^{T_2} \beta_{X_i}(\xi) d\xi\right]}{\exp\left[-\int_0^{T_1} \beta_{X_i}(\xi) d\xi\right]} = \exp\left[-\int_{T_1}^{T_2} \beta_{X_i}(\xi) d\xi\right], \quad i = E \text{ or } S$$

Simplifying, we obtain

$$\begin{aligned} P\{E > T_2 | E > T_1\} &= \exp\left[-\frac{1}{2}\int_{T_1}^{T_2} \beta_S(\xi) d\xi\right] \\ P\{S > T_2 | S > T_1\} &= \exp\left[-\int_{T_1}^{T_2} \beta_S(\xi) d\xi\right] \end{aligned}$$

Thus, the probability of an exercising person who has lived upto  $T_1$  years living to  $T_2$  years is the square root of the probability of the sedentary person. If the death rate of a sedentary person,  $\beta_S(t) = 0.001t$ ,  $T_1 = 60$  years and  $T_2 = 70$  years, then the following probabilities

can be evaluated:

$$P\{S > 70 \mid S > 60\} = \exp\left[-\int_{60}^{70} 0.001\xi d\xi\right] = 0.522$$

$$P\{E > 70 \mid E > 60\} = \exp\left[-\frac{1}{2}\int_{60}^{70} 0.001\xi d\xi\right] = \sqrt{0.522} = 0.7225$$

The probability of a 60 year old sedentary person reaching 70 years is 0.522, whereas the same person exercising regularly will reach 70 years with a probability of 0.7225.

## 6.4 NORMAL OR GAUSSIAN DISTRIBUTION

The most important distribution in probability theory is the Gaussian distribution. Under some general conditions, a random variable  $X$  consisting of a number of components, each with a general distribution tends to a normal distribution as the number becomes very large.

The probability density  $f_X(x)$  of a Gaussian distribution is given by

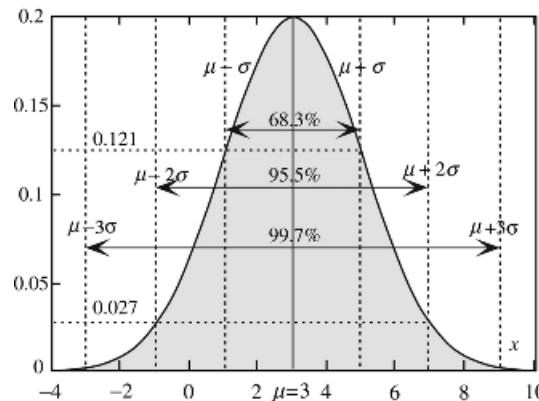
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(x-\mu)/\sigma]^2} \quad (6.4.1)$$

where  $\mu$  is called the *mean* and  $\sigma^2$  is called the *variance*. These quantities will be defined later. The cumulative distribution function  $F_X(x)$  is given by

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(\xi-\mu)/\sigma]^2} d\xi \quad (6.4.2)$$

The Gaussian density function  $f_X(x)$  given by Eq. (6.4.1) is shown in Fig. 6.4.1 for  $\mu = 3$  and  $\sigma^2 = 4$ . It also shows that the area under the Gaussian pdf between  $(\mu - \sigma, \mu + \sigma]$  is 68.3%, between  $(\mu - 2\sigma, \mu + 2\sigma]$  is 95.5%, and between  $(\mu - 3\sigma, \mu + 3\sigma]$  is 99.7%. Thus, almost the entire pdf (99.7%) is between the  $\mu \pm 3\sigma$  points of the Gaussian curve.

The Gaussian distribution function,  $F_X(x)$  is shown in Fig. 6.4.2 along with  $(\mu - \sigma, \mu + \sigma]$ ,  $(\mu - 2\sigma, \mu + 2\sigma]$ , and  $(\mu - 3\sigma, \mu + 3\sigma]$  points.



**FIGURE 6.4.1**

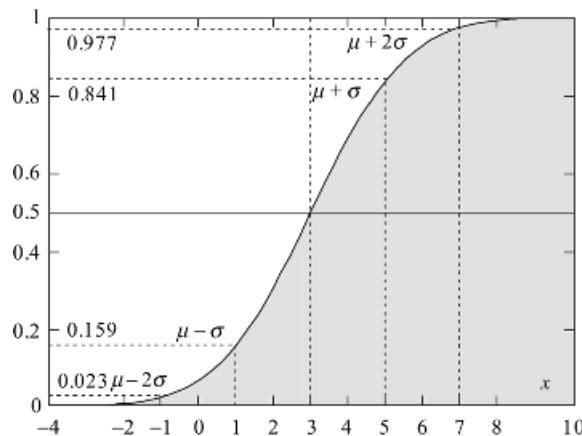


FIGURE 6.4.2

A *standard Gaussian* is defined as the random variable with  $\mu = 0$  and  $\sigma^2 = 1$ , given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)x^2} \quad (6.4.3)$$

and the corresponding cumulative distribution function is denoted by  $F_X(x)$ , given by

$$\Phi_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-(1/2)\xi^2} d\xi \quad (6.4.4)$$

There is no closed-form solution for Eq. (6.4.4). Hence,  $F_X(x)$  is tabulated for values of  $x$  between  $-5$  and  $5$  in Appendix B. The standard Gaussian pdf  $f_X(x)$  is shown in Fig. 6.4.3 with the 68.3% points between  $(-1,1]$ , 95.5% points between  $(-2,2]$ , and the 99.7% points between  $(-3,3]$ .

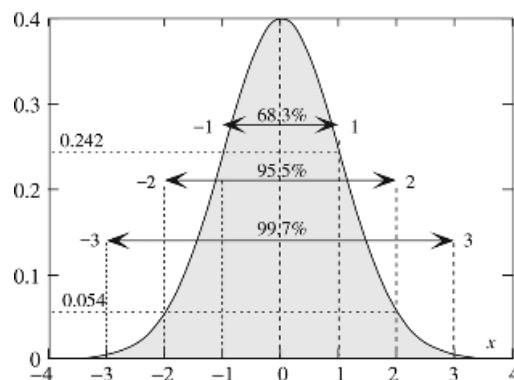


FIGURE 6.4.3

Given a random variable  $X$  with any  $\mu$  and  $\sigma^2$ , the random variable

$$Z = \frac{X - \mu_X}{\sigma_X}$$

is a standard Gaussian. Hence, the distribution function  $F_X(x)$  can be given in terms of the standard Gaussian  $F_Z(x)$  as follows:

$$F_X(x_0) = P\{X \leq x_0\} = P\left\{\frac{X - \mu_X}{\sigma_X} \leq \frac{x_0 - \mu_X}{\sigma_X}\right\} = \Phi_X\left(\frac{x_0 - \mu_X}{\sigma_X}\right) \quad (6.4.5)$$

We shall now prove that the  $f_X(x)$  is a probability density function by showing that if

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1/2)x^2} dx$$

then  $I = 1$ . Direct evaluation of the integral is difficult. Hence we square  $I$  and write

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-(1/2)x^2} dx \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1/2)y^2} dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(1/2)(x^2+y^2)} dx dy \end{aligned} \quad (6.4.6)$$

Substituting  $x = r \cos(\theta)$  and  $y = r \sin(\theta)$  in Eq. (6.4.5) we obtain,

$$I^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} r e^{-(1/2)r^2} dr d\theta = 1 \quad (6.4.7)$$

and the result is proven.

**Example 6.4.1** The number of malfunctioning computers is Gaussian-distributed with  $\mu = 4$  and  $\sigma = 3$ . We want to find the probability that the number of bad computers is between 3 and 5.

Denoting the malfunctioning computers by the random variable  $X$ , we have to find  $P\{3 < X \leq 5\}$ . Using Eq. (6.4.5) and referring to the Gaussian tables, we have

$$\begin{aligned} P\{3 < X \leq 5\} &= P\left\{\frac{3 - 4}{3} < \frac{X - 4}{3} \leq \frac{5 - 4}{3}\right\} = P\left\{-\frac{1}{3} < Z \leq \frac{1}{3}\right\} \\ &= \Phi_X\left(\frac{1}{3}\right) - \Phi_X\left(-\frac{1}{3}\right) = 0.2611 \end{aligned}$$

**Example 6.4.2** We shall study what grading on the curve in examinations means. If the average of the class in a particular examination is  $\mu$ , then one method of grading is to assign  $A$  to those receiving over  $\mu + 2\sigma$ ,  $AB$  to those receiving between  $\mu + \sigma$  and  $\mu + 2\sigma$ ,  $B$  to those receiving between  $\mu$  and  $\mu + \sigma$ ,  $C$  to those receiving between  $\mu - \sigma$  and  $\mu$ ,  $D$  to those receiving between  $\mu - \sigma$  and  $\mu - 2\sigma$ , and  $F$  to those receiving less than  $\mu - 2\sigma$ . We want to calculate the percentage of students receiving the grades  $A$ ,  $AB$ ,  $B$ ,  $C$ ,  $D$ , and  $F$ .

We can calculate these probabilities from Eq. (6.4.5) as follows:

$$A: P\{X > \mu + \sigma\} = P\left\{\frac{X - \mu}{\sigma} > 2\right\} = 1 - \Phi(2) = 0.0228 = 2.28\%$$

$$AB: P\{\mu + \sigma < X \leq \mu + 2\sigma\} = P\left\{1 < \frac{X - \mu}{\sigma} \leq 2\right\} = \Phi(2) - \Phi(1) \\ = 0.1359 = 13.59\%$$

$$B: P\{\mu < X \leq \mu + \sigma\} = P\left\{0 < \frac{X - \mu}{\sigma} \leq 1\right\} = \Phi(1) - \Phi(0)$$

$$= 0.3413 = 34.13\%$$

$$C: P\{\mu - \sigma < X \leq \mu\} = P\left\{-1 < \frac{X - \mu}{\sigma} \leq 0\right\} = \Phi(0) - \Phi(-1)$$

$$= 0.3413 = 34.13\%$$

$$D: P\{\mu - 2\sigma < X \leq \mu - \sigma\} = P\left\{-2 < \frac{X - \mu}{\sigma} \leq -1\right\} = \Phi(-1) - \Phi(-2)$$

$$= 0.1359 = 13.59\%$$

$$F: P\{X \leq \mu - 2\sigma\} = P\left\{\frac{X - \mu}{\sigma} \leq -2\right\} = \Phi(-2) = 0.0228 = 2.28\%$$

These grades are shown in Fig. 6.4.4.

### Properties of Standard Gaussian

#### Density and Distribution

1. The pdf  $f_X(x)$  is an even function.
2.  $f_X(x)$  attains its maximum value of  $1/\sqrt{2\pi}$  at  $x = 0$ .
3. The cdf  $F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi$ .

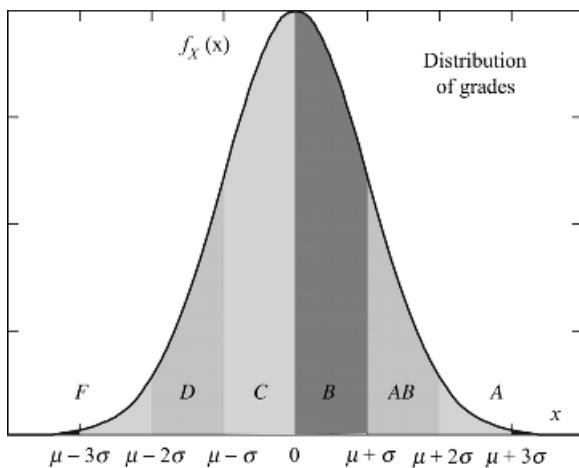


FIGURE 6.4.4

4.  $\Phi_X(0) = \frac{1}{2}$ .

5. The error function  $\text{erf}(x)$  is defined as

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} d\xi \quad (6.4.8)$$

6. The complement of the error function  $\text{erfc}(x)$  is defined as

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-\xi^2} d\xi = 1 - \text{erf}(x) \quad (6.4.9)$$

7. The tails of the distribution  $Q(x)$  is used in describing the probability of error in transmission of communication signals. It is defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-(\xi^2/2)} d\xi \quad (6.4.10)$$

We will discuss this function in greater detail in the next section.

8.  $Q(x)$  in terms of error function complement  $\text{erfc}(x)$  is

$$Q(x) = \frac{1}{2} \text{erfc}\left(\frac{x}{\sqrt{2}}\right) \quad (6.4.11)$$

9. The general distribution function  $F_X(x)$  in terms of  $\text{erf}(x)$  and  $\text{erfc}(x)$  is

$$\begin{aligned} F_X(x) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-(1/2)[(\xi-\mu)/\sigma]^2} d\xi = \frac{1}{2} \frac{2}{\sqrt{\pi}} \int_{-\infty}^{(x-\mu)/\sqrt{2\sigma}} e^{-\xi^2} d\xi \\ &= \frac{1}{2} + \frac{1}{2} \text{erf}\left(\frac{x-\mu}{\sqrt{2\sigma}}\right) = 1 - \frac{1}{2} \text{erfc}\left(\frac{x-\mu}{\sqrt{2\sigma}}\right) \end{aligned} \quad (6.4.12)$$

10.  $F_X(\mu) = \frac{1}{2}$  (6.4.13)

### **Gaussian Tails $Q(x)$**

As mentioned in property 7 [Eq. (6.4.10)]  $Q(x)$  is used to express the bit error probability in transmission of communication signals. Since  $Q(x)$  does not have a closed-form solution, bounds can be established.

The distribution  $Q(x)$  can be expanded into an asymptotic series [1] given by

$$Q(x) = \frac{e^{-x^2}}{\sqrt{2\pi}x} \left\{ 1 - \frac{1}{x^2} + \frac{1.3}{x^4} + \cdots + (-1)^n \frac{1.3 \cdots (2n-1)}{x^{2n}} \right\} \quad (6.4.14)$$

and for large  $x$ ,  $Q(x)$  can be approximated by

$$Q(x) \approx \frac{e^{-x^2}}{\sqrt{2\pi}x} \quad (6.4.15)$$

For small  $x$ ,  $Q(x)$  diverges, and hence this is not a very good bound. A better bound is

$$Q(x) \approx QA(x) = \frac{e^{-x^2}}{\sqrt{2\pi}} \frac{1}{\sqrt{1+x^2}} \quad (6.4.16)$$

Even this bound, although better than Eq. (6.4.15), is not very good for small values of  $x$ . However, a tighter bound has been established by Börjesson and Sundberg [8] by first integrating  $Q(x)$  by parts as shown below:

$$\begin{aligned} Q(x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-(\xi^2/2)} d\xi = 1\sqrt{2\pi} \int_x^\infty \frac{1}{\xi} \xi e^{-(\xi^2/2)} d\xi \\ &= \frac{1}{\sqrt{2\pi}} \left\{ \frac{1}{x} e^{-(x^2/2)} - \int_x^\infty \frac{1}{\xi^2} e^{-(\xi^2/2)} d\xi \right\} \end{aligned}$$

and the expression for  $Q(x)$  can be manipulated to obtain

$$Q(x) = \frac{1}{(1-\alpha)x + \alpha\sqrt{x^2 + \beta}} e^{-(x^2/2)} \quad (6.4.17)$$

Substituting the values  $\alpha = 1$  and  $\beta = 1$  in Eq. (6.4.17), we obtain Eq. (6.4.16). Better values for  $\alpha$  and  $\beta$  can be obtained by minimizing the absolute relative error function  $\varepsilon(x)$ , given by

$$\varepsilon(x) = \left| \frac{Q_0(x) - Q(x)}{Q_0(x)} \right| \quad (6.4.18)$$

and constraining  $\varepsilon(0) = 0$ . In Eq. (6.4.18)  $Q_0(x)$  is the true value given by Eq. (6.4.10). As a consequence, the upper bound  $QU(x)$  is obtained for values of  $\alpha = 0.33919$  and  $\beta = 5.533425$ , and the lower bound  $QL(x)$  is obtained for values of  $\alpha = 1/\pi$  and  $\beta = 2\pi$ . Thus the upper bound is

$$QU(x) = \frac{1}{0.66081x + 0.33919\sqrt{x^2 + 5.533425}} \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} \quad (6.4.19)$$

and the lower bound is,

$$QL(x) = \frac{1}{(1 - 1/\pi) + 1/\pi\sqrt{x^2 + 2\pi}} \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} \quad (6.4.20)$$

The functions  $Q(x)$ ,  $QU(x)$ ,  $QL(x)$ , and  $QA(x)$  are shown in Fig. 6.4.5 and tabulated in Table 6.4.1 showing the upper and lower bounds are for all  $x > 0$ . The bound  $QA(x)$  is not good for low values of  $x$ , whereas  $QU(x)$  and  $QL(x)$  practically lie on  $Q(x)$ .

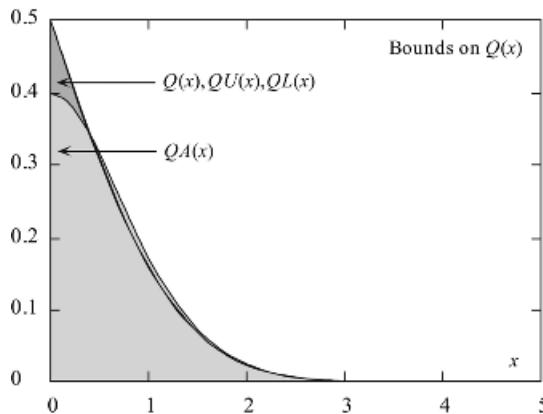


FIGURE 6.4.5

**TABLE 6.4.1**

$x$	Lower Bound $QL(x)$	True Value $Q(x)$	Upper Bound $QU(x)$	Approx. Value $QA(x)$
0	0.5	0.5	0.5	0.39894228
0.5	0.30496416	0.30853754	0.3071816	0.3148968
1	0.15704991	0.15865525	0.15837866	0.17109914
1.5	0.06633866	0.0668072	0.06684731	0.07184344
2	0.0226461	0.02275013	0.02279147	0.02414549
2.5	0.00619131	0.00620967	0.00622374	0.00650985
3	0.00134729	0.0013499	0.00135302	0.00140147
3.5	0.00023233	0.00023263	0.00023314	0.00023974
4	0.00003164	0.00003167	0.00003173	0.00003246
4.5	0.0000034	0.0000034	0.0000034	0.00000347
5	0.00000029	0.00000029	0.00000029	0.00000029

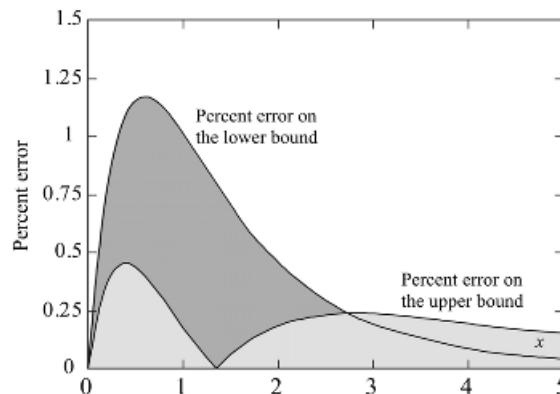
The absolute percentage errors for the upper  $QU(x)$  and lower bounds  $QL(x)$  are shown in Fig. 6.4.6. Table 6.4.2 shows in addition to the upper- and lower-bound errors the absolute percentage error for the approximate value  $QA(x)$ . Since these errors are large, they are not shown in Fig. 6.4.6.

The errors for the approximate value are high (20.2%). The upper-bound errors are better for smaller values of  $x$ , whereas the lower-bound is better for higher values of  $x$ . Even then, these bounds clearly illustrate that the approximation for all  $x$  is excellent with the maximum error of about 1.16% occurring at low values of  $x$ . This is much better than the usual approximation  $QA(x)$ , where the maximum error is 20.2% for low values. For high values of  $x$ ,  $QU(x)$ ,  $QL(x)$ , and  $QA(x)$  provide reasonably good approximation.

### Properties of $Q$ Function

1. The limit as  $x \rightarrow \infty$  is zero:

$$\lim_{x \rightarrow \infty} Q(x) = 0 \quad (6.4.21)$$

**FIGURE 6.4.6**

**TABLE 6.4.2 Error percentages**

$x$	Lower Bound	Upper Bound	Approximate Value
0	0	0.00000176	20.21154392
0.5	1.15816594	0.43947331	2.06109849
1	1.01184414	0.17433835	7.8433496
1.5	0.70133395	0.06003046	7.53846047
2	0.45727327	0.1816945	6.13342513
2.5	0.29559943	0.226622	4.83410996
3	0.19344425	0.23152986	3.82069533
3.5	0.12922798	0.21848937	3.05864395
4	0.08836646	0.19884525	2.48602699
4.5	0.06185728	0.1779578	2.05102577
5	0.04427826	0.15815582	1.71571131

2. The limit as  $x \rightarrow -\infty$  is 1:

$$\lim_{x \rightarrow -\infty} Q(x) = 1 \quad (6.4.22)$$

3. The value of  $Q(x)$  at  $x = 0$  is half:

$$Q(0) = \frac{1}{2} \quad (6.4.23)$$

4.  $Q(x)$  is even:

$$Q(-x) = Q(x) \quad (6.4.24)$$

5.  $Q(x)$  can be expressed in terms of error function as

$$Q(x) = \frac{1}{2} \left[ 1 - \operatorname{erf} \left( \frac{x}{\sqrt{2}} \right) \right] = \frac{1}{2} \operatorname{erfc} \left( \frac{x}{\sqrt{2}} \right), \quad x \geq 0 \quad (6.4.25)$$

6. If  $\Phi(x)$  is a standard Gaussian, then

$$Q(x) = 1 - \Phi(x) \quad (6.4.26)$$

7. For any Gaussian distributed random variable  $X$  with mean  $\mu$  and variance  $\sigma^2$

$$P\{X > \alpha\} = 1 - F_X(\alpha) = Q\left(\frac{\alpha - \mu}{\sigma}\right) \quad (6.4.27)$$

### Gaussian Approximation to Binomial Distribution

In Example 4.7.3 we saw that the binomial distribution can be approximated by a Poisson distribution under the condition that  $n$  is large compared to  $k$  and  $p$  is small enough that the product  $np = \lambda$  is of moderate magnitude. On the other hand, if  $npq$  is quite large, the Poisson approximation does not fare very well. In this case we use the Gaussian approximation, which is called the *DeMoivre–Laplace limit theorem*, which states that, if  $S_n$  is the total number of successes in  $n$  Bernoulli trials, each with probability of success  $p$  and failure  $q$  with  $(p + q = 1)$  with pmf  $b(k;n,p)$ , then, as  $n \rightarrow \infty$ , the standardized random variable  $Z = (S_n - np)/\sqrt{npq}$  converges to a standard normal distribution function where  $\mu = np$  is the mean and  $\sigma^2 = npq$  is the variance. Or, if

$n_1 < n_2$ , then

$$\lim_{n \rightarrow \infty} Z = P\left\{n_1 < \frac{S_n - np}{\sqrt{npq}} \leq n_2\right\} \rightarrow \Phi(n_2) - \Phi(n_1) \quad (6.4.28)$$

and the probability mass function  $b(k; n, p)$  converges to the density function  $f_Z(z)$ . Or

$$\lim_{n \rightarrow \infty} \frac{n!}{k!(n-k)!} p^k q^{n-k} \rightarrow \frac{1}{\sqrt{2\pi npq}} e^{-(1/2)[(x-np)/npq]^2} \quad (6.4.29)$$

The Demoivre–Laplace limit theorem is a special case of the central limit theorem presented in a later chapter.

**Example 6.4.3 (Gaussian and Poisson Approximation)** We will find the Gaussian approximation for the binomial distribution given in Example 4.7.3, where the Poisson approximation has already been found. In the binomial distribution  $b(k; n, p)$  of that example,  $n = 1000$ ,  $p = 0.005$  with  $np = 5$ , and  $npq = 4.975$ .

The approximate Gaussian pdf will be given by

$$f_X(k) = \frac{1}{\sqrt{2\pi \cdot 4.975}} e^{-(1/2)[(k-5)^2/4.975]} \approx b(k; 1000, 0.005)$$

For the Poisson approximation  $np = \lambda = 5$ , and this is given in Example 4.7.3 as

$$p_X(k; \lambda) = p_X(k; 5) = p_X(k) = e^{-5} \left( \frac{5^k}{k!} \right)$$

Since  $\lambda = 5$  is a single digit and  $npq = 4.975$  is not very large, we would reasonably infer that the Poisson approximation will be better than the Gaussian. The binomial distribution with the Poisson and Gaussian approximations are shown in Fig. 6.4.7 for  $k = 0, \dots, 15$ .

The true value of  $b(5; 1000, 0.005) = 0.17591$ , the Gaussian approximation  $f_X(5) = 0.17886$  with an absolute error of 1.678% and the Poisson approximation  $p_X(5; 5) = 0.17547$  with an error of 0.25%, confirming our inference that the Poisson approximation is better. Figure 6.4.7 clearly shows that the Poisson approximation follows the binomial more closely than the Gaussian.

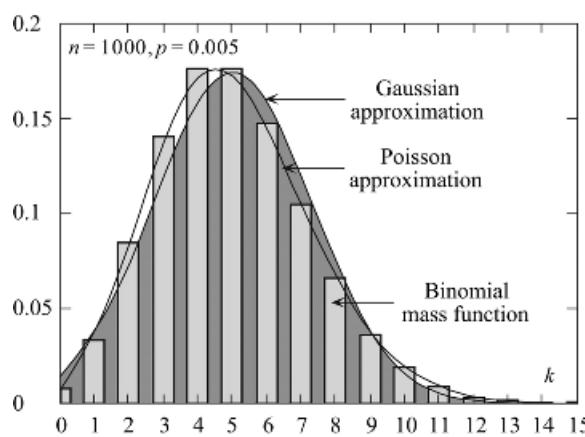


FIGURE 6.4.7

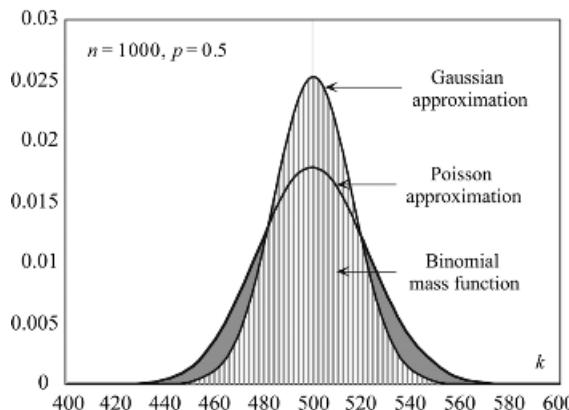


FIGURE 6.4.8

**Example 6.4.4** We shall now modify Example 6.4.3 by keeping  $n = 1000$  and changing  $p = 0.5$ , yielding  $np = \lambda = 500$ , and  $npq = 250$ . Here both  $\lambda$  and  $npq$  are quite large, so we can infer that the Gaussian will be a better approximation than the Poisson. The approximate Gaussian pdf is given by

$$f_X(k) = \frac{1}{\sqrt{2\pi \cdot 250}} e^{-(1/2)[(k-500)^2/250]} \approx b(k; 1000, 0.5)$$

and the corresponding Poisson approximation is

$$P_X(k; \lambda) = P_X(k; 500) = P_X(k) = e^{-500} \left( \frac{500^k}{k!} \right)$$

The binomial distribution with the Poisson and Gaussian approximations for this case are shown in Fig. 6.4.8.

TABLE 6.4.3 Comparison of binomial with Gaussian and Poisson approximations

$k$	Binomial	Gaussian	Poisson	Error $\varepsilon_{BG}$	Error $\varepsilon_{BP}$
$n = 1000, p = 0.005, np = 5, npq = 4.975$					
2	0.083929	0.072391	0.084224	13.74699	0.352343
3	0.140303	0.119653	0.140374	14.718176	0.50683
4	0.175731	0.161758	0.175467	7.951344	0.150021
5	0.175908	0.17886	0.175467	1.678441	0.250272
6	0.14659	0.161758	0.146223	10.347485	0.250272
7	0.104602	0.119653	0.104445	14.38896	0.14992
8	0.035245	0.072391	0.065278	10.953113	0.051188
$n = 1000, p = 0.5, np = 500, npq = 250$					
497	0.024775	0.024781	0.017731	0.023389	28.431591
498	0.025024	0.02503	0.017803	0.02438	28.85844
499	0.25175	0.025181	0.017838	0.024978	29.141872
500	0.25225	0.025231	0.017838	0.025177	29.283306
501	0.025175	0.025181	0.017803	0.024978	29.283306
502	0.025024	0.02503	0.017732	0.02438	29.141589
503	0.024775	0.024781	0.017626	0.023389	28.857017

The true value for  $k = 500$  for binomial  $b(500,1000,0.5) = 0.025225$ , the Gaussian approximation  $f_X(k) = 0.025231$  with an absolute error of 0.025% and the Poisson approximation  $p_X(k) = 0.017838$  with an error of 29.28%. In this case, the Gaussian approximation, as expected, fares very much better than the Poisson one. The results graphed in Fig. 6.4.8 for  $k = 400, \dots, 600$  clearly show that the Gaussian approximation follows the binomial much more closely than does the Poisson, unlike the scenario described in Example 6.4.3.

The percentage absolute errors for the Gaussian and Poisson approximations to the binomial are defined by

$$\text{Gaussian: } \varepsilon_{BG} = \left| \frac{b(k;n,p) - f_X(k;np)}{b(k;n,p)} \right| \times 100$$

$$\text{Poisson: } \varepsilon_{BP} = \left| \frac{b(k;n,p) - p_X(k;\lambda)}{b(k;n,p)} \right| \times 100$$

These errors are tabulated for  $p = 0.005$  and 0.5 in Table 6.4.3 in the neighborhoods of  $np = \lambda = 5$  and 500, respectively. The absolute error percentages also show that for large  $n$  and for values of moderate  $np$  and  $npq$ , the Poisson approximation is better and for large values of  $np$  and  $npq$ , the Gaussian approximation is better.

# Other Continuous Distributions

## 7.1 INTRODUCTION

We will now discuss some of the other important continuous distributions that occur in communications, signal processing, queuing, statistical inference, and other probabilistic situations. This list is by no means exhaustive. For a more complete list of distributions, refer to Web references W1 and W2 (last two entries in References list at end of the book). Most of these distributions are derivable from the three basic distributions discussed in Chapter 6.

## 7.2 TRIANGULAR DISTRIBUTION

If two independent random variables  $X$  and  $Y$  are uniformly distributed in the interval  $(a, b]$ , then  $Z = X + Y$  has a triangular density defined by

$$f_Z(z) = \begin{cases} \frac{z-a}{(b-a)^2} & a < z \leq b \\ \frac{-(z-2b+a)}{(b-a)^2} & b < z \leq 2b-a \\ 0 & \text{otherwise} \end{cases} \quad (7.2.1)$$

The density  $f_Z(z)$  is shown in Fig. 7.2.1 with the corresponding uniform densities of  $X$  and  $Y$  for  $a = 1$  and  $b = 5$ .

The distribution function  $F_Z(z)$  is given by

$$F_Z(z) = \begin{cases} 0 & z \leq a \\ \frac{1}{2} \left( \frac{z-a}{b-a} \right)^2 & a < z \leq b \\ 1 - \frac{1}{2} \left( \frac{z+a-2b}{b-a} \right)^2 & b < z \leq 2b-a \\ 1 & z > 2b-a \end{cases} \quad (7.2.2)$$

The distribution function  $F_Z(z)$  is shown for  $a = 1$  and  $b = 5$  in Fig. 7.2.2.

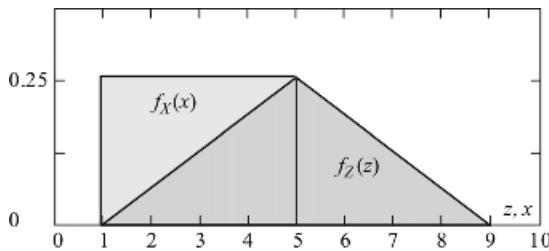


FIGURE 7.2.1

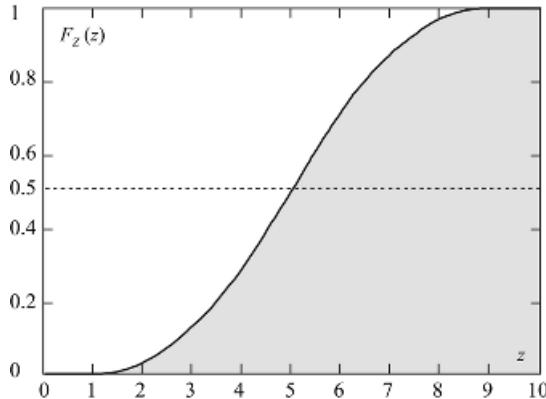


FIGURE 7.2.2

### 7.3 LAPLACE DISTRIBUTION

The Laplace density, also called a *double-exponential density*, is used to model navigation and pilot errors, speech, and image processing. As the term suggests, double-exponential density is defined by

$$f_X(x) = \begin{cases} \frac{\lambda}{2} e^{-\lambda|x|} & \lambda > 0 \end{cases} \quad (7.3.1)$$

The cdf  $F_X(x)$  is given by

$$F_X(x) = \begin{cases} \frac{1}{2} e^{-\lambda x} & x \leq 0 \\ 1 - \frac{1}{2} e^{-\lambda x} & x > 0 \end{cases} \quad (7.3.2)$$

It is an even density function, as shown for  $\lambda = 0.5$  in Fig. 7.3.1.

The cdf  $F_X(x)$  is shown in Fig. 7.3.2.

An alternate form of Laplace density in terms of the variance  $\sigma^2$  is defined by

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2}\sigma} e^{-\sqrt{2}(|x|/\sigma)} \end{cases} \quad (7.3.3)$$

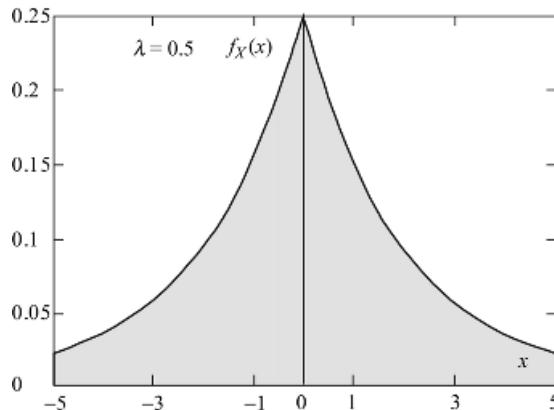


FIGURE 7.3.1

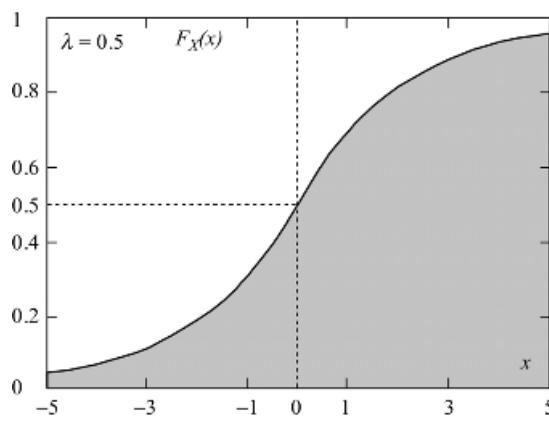


FIGURE 7.3.2

with the corresponding cdf  $F_X(x)$  given by

$$F_X(x) = \begin{cases} \frac{1}{2} e^{\sqrt{2}(x/\sigma)} & x \leq 0 \\ 1 - \frac{1}{2} e^{\sqrt{2}(x/\sigma)} & x > 0 \end{cases} \quad (7.3.4)$$

where  $\sigma^2$  is the variance of the Laplace distribution.

## 7.4 ERLANG DISTRIBUTION

We already saw that the waiting time between successive events of a Poisson arrival process is exponentially distributed with parameter  $\lambda$ . The distribution of the waiting time for  $n$  successive independent events of a Poisson arrival process is given by the Erlang distribution with  $n$  degrees of freedom and is defined for positive integral values of  $n$  by

$$f_X(x) = \frac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!}, \quad \text{where } n \text{ is an integer} \quad (7.4.1)$$

The Erlang distribution is obtained by the  $n$ -fold convolution of  $n$  independent exponential distributions. The cumulative distribution corresponding to Eq. (7.4.1) can be derived as follows. Let  $X_n$  denote the time at which the  $n$ th event occurs. Let  $N(x)$  be the number of events in the interval  $(0, x]$ . Hence the event  $\{X_n \leq x\}$  can occur only if the number of events occurring in  $(0, x]$  is at least  $n$ . Since  $N(x)$  is Poisson-distributed, the cdf  $F_{X_n}(x)$  can be given by

$$\begin{aligned} P\{X_n \leq x\} &= F_{X_n}(x) = P\{N(x) > n\} \\ &= \sum_{k=n}^{\infty} e^{-\lambda x} \frac{(\lambda x)^k}{k!} = 1 - \sum_{k=0}^{n-1} e^{-\lambda x} \frac{(\lambda x)^k}{k!} \end{aligned} \quad (7.4.2)$$

We can obtain the pdf  $f_{X_n}(x)$  by differentiating Eq. (7.4.2) with respect to  $x$ . Performing the indicated differentiation, we can write

$$\begin{aligned} \frac{d}{dx} \{F_{X_n}(x)\} &= \frac{d}{dx} \left\{ 1 - \sum_{k=0}^{n-1} e^{-\lambda x} \frac{(\lambda x)^k}{k!} \right\} \\ &= -e^{-\lambda x} \sum_{k=0}^{n-1} \frac{k(\lambda x)^{k-1} \lambda}{k!} + \lambda e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} \\ &= \lambda e^{-\lambda x} \left\{ \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} - \sum_{k=1}^{n-1} \frac{(\lambda x)^{k-1}}{(k-1)!} \right\} \\ &= \lambda e^{-\lambda x} \left\{ \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} - \sum_{k=0}^{n-2} \frac{(\lambda x)^k}{k!} \right\} \end{aligned}$$

or

$$\frac{d}{dx} \{F_{X_n}(x)\} = f_{X_n}(x) = \frac{\lambda e^{-\lambda x} (\lambda x)^{n-1}}{(n-1)!} \quad (7.4.3)$$

and Eq. (7.4.3) is the Erlang distribution. This distribution is shown in Fig. 7.4.1 for  $n = 10$  and  $\lambda = 2, 2.5, 3, 3.5, 4, 5$ .

Another family of Erlang distributions is shown for  $\lambda = 4$  and  $n = 2, 3, 4, 5, 6$  in Fig. 7.4.2.

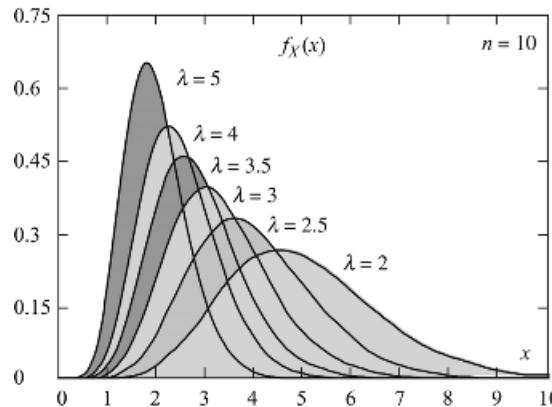


FIGURE 7.4.1

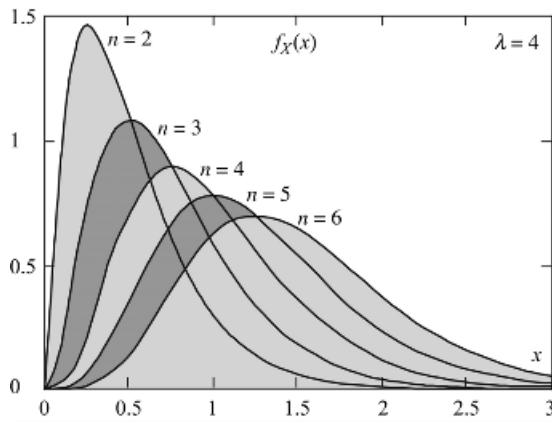


FIGURE 7.4.2

In both these cases the Erlang distributions reduce to an exponential distribution for  $n = 1$ . In cases where the data may not be modeled as a simple exponential, the Erlang family provides greater flexibility.

## 7.5 GAMMA DISTRIBUTION

A gamma distribution is a generalization of the Erlang distribution where  $n = \alpha$ , which may not be an integer. It finds wide use in statistics. Gamma density is a two-parameter family with parameters  $\alpha > 0$ , termed the *shape parameter*, and  $\lambda > 0$ , called the *scale parameter*, and is defined as

$$f_X(x) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad \alpha, \lambda, x > 0 \quad (7.5.1)$$

where  $\Gamma(\alpha)$  is the gamma function defined by the integral

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \alpha > 0 \quad (7.5.2)$$

Integrating Eq. (7.5.2) by parts with  $u = x^{\alpha-1}$  and  $dv = e^{-x} dx$ , we have

$$\begin{aligned} \Gamma(\alpha) &= \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \alpha > 0 = -x^{\alpha-1} e^{-x} \Big|_0^\infty + (\alpha - 1) \int_0^\infty x^{\alpha-2} e^{-x} dx \\ &= (\alpha - 1) \Gamma(\alpha - 1) \end{aligned} \quad (7.5.3)$$

If  $\alpha$  is an integer  $n$ , then  $\Gamma(n) = (n-1)\Gamma(n-1)$  and continued expansion yields  $\Gamma(n) = (n-1)!$  Hence the gamma function can be regarded as the generalization of the factorial function for all positive real numbers. The most useful fractional argument for the gamma function is  $\frac{1}{2}$  with

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \quad (7.5.4)$$

The *incomplete gamma function*  $\Gamma_x(\alpha)$  is defined by

$$\Gamma_x(\alpha) = \int_0^x \xi^{\alpha-1} e^{-\xi} d\xi \quad (7.5.5)$$

We will use Eq. (7.5.5) later for the  $\chi^2$  distribution function  $F_X(x)$  in Section 7.7. We shall now show that the density function given by Eq. (7.5.1) indeed integrates to 1. If we define the integral  $I_x$  by

$$I_x = \int_0^x \frac{\lambda(\lambda\eta)^{\alpha-1} e^{-\lambda\eta}}{\Gamma(\alpha)} d\eta \quad (7.5.6)$$

Then, substituting  $\lambda\eta = \xi$  in Eq. (7.5.6), we can rewrite

$$I_x = \frac{1}{\Gamma(\alpha)} \int_0^{\lambda x} \xi^{\alpha-1} e^{-\xi} d\xi \quad (7.5.7)$$

Applying Eq. (7.5.5) to Eq. (7.5.7), we have

$$I_x = \frac{\Gamma_{\lambda x}(\alpha)}{\Gamma(\alpha)} \quad \text{or} \quad \lim_{x \rightarrow \infty} I_x = \frac{\Gamma(\alpha)}{\Gamma(\alpha)} = 1$$

thus showing the validity of Eq. (7.5.1) as a proper density function. If  $\alpha$  is an integer, then we have the *Erlang* distribution.

The gamma density family is shown in Fig. 7.5.1 with the parameter  $\lambda = 2$  and  $\alpha = 0.5, 1, 1.5, 2.5, 3.5, 4.5, 5.5$ .

The cumulative distribution function  $F_X(x)$  is given by

$$F_X(x) = \frac{1}{\Gamma(\alpha)} \int_0^x \lambda(\lambda\xi)^{\alpha-1} e^{-\lambda\xi} d\xi = \frac{\Gamma_x(\alpha)}{\Gamma(\alpha)} \quad (7.5.8)$$

This integral in Eq. (7.5.8) has no closed-form solution unless  $\alpha$  is an integer, in which case it is the Erlang distribution given by Eq. (7.4.2). However, the integral given by Eq. (7.5.8) has been extensively tabulated.

### Reproductive Property of Gamma Distribution

If  $X_1, X_2, \dots, X_n$  are  $n$  independent gamma-distributed random variables with shape parameters  $\alpha_1, \alpha_2, \dots, \alpha_n$  and having the same scale parameter  $\lambda$ , then random variable

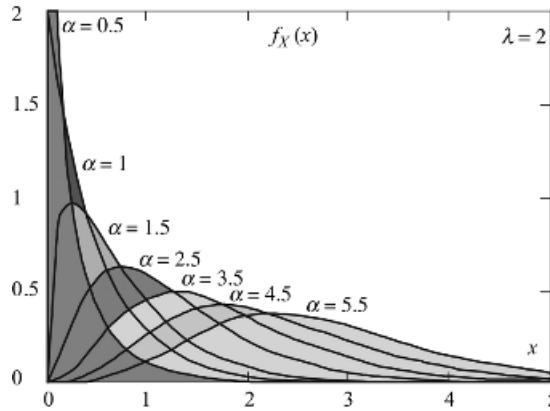


FIGURE 7.5.1

$Z = X_1 + X_2 + \dots + X_n$  is also gamma-distributed with shape parameter  $\beta = \alpha_1 + \alpha_2 + \dots + \alpha_n$ . As a consequence of this property, since an exponential random variable is gamma-distributed with parameters  $(1, \lambda)$ , the sum of  $n$  exponential random variables is gamma-distributed with parameters  $(n, \lambda)$ , which is an Erlang distribution given by Eq. (7.4.1).

## 7.6 WEIBULL DISTRIBUTION

We will define the Weibull distribution as a three-parameter family of distribution functions, in which the parameter  $\lambda > 0$  is the *scale* parameter,  $\alpha > 0$  is the *shape* parameter, and  $\mu$  is the *location* parameter. It is defined by

$$f_X(x) = \begin{cases} \alpha \lambda^\alpha (x - \mu)^{\alpha-1} e^{-[\lambda(x-\mu)]^\alpha} & x > \mu \\ 0 & x \leq \mu \end{cases} \quad (7.6.1)$$

We will now find the cdf  $F_X(x)$  of the Weibull distribution. We can write

$$F_X(x) = \int_{\mu}^x \alpha \lambda^\alpha (\xi - \mu)^{\alpha-1} e^{-[\lambda(\xi-\mu)]^\alpha} d\xi \quad (7.6.2)$$

Substituting  $\eta = [\lambda(\xi - \mu)]^\alpha$  and hence  $d\eta = [\lambda(\xi - \mu)]^{\alpha-1} \alpha \lambda d\xi$ , we can rewrite Eq. (7.6.2) as follows:

$$F_X(x) = \int_0^{[\lambda(x-\mu)]^\alpha} e^{-\eta} d\eta = -e^{-\eta} \Big|_0^{[\lambda(x-\mu)]^\alpha}, \quad x > \mu \quad (7.6.3)$$

Substitution of the limits in Eq. (7.6.3) yields

$$F_X(x) = \begin{cases} 1 - e^{-[\lambda(x-\mu)]^\alpha} & x > \mu \\ 0 & x \leq \mu \end{cases} \quad \text{and} \quad F_X(\infty) = 1 \quad (7.6.4)$$

Two families of Weibull distributions are shown for  $\mu = 0$  and  $\alpha = 2$ ,  $\lambda = 0.5, 1, 2, 3, 4$  (Fig. 7.6.1) and for  $\lambda = 2$  and  $\alpha = 0.5, 1, 2, 3, 4, 5$  (Fig. 7.6.2).

When  $\mu = 0$  and  $\lambda = 1$  in Eqs. (7.6.1) and (7.6.4), then the resulting distribution is called the *standard Weibull*.

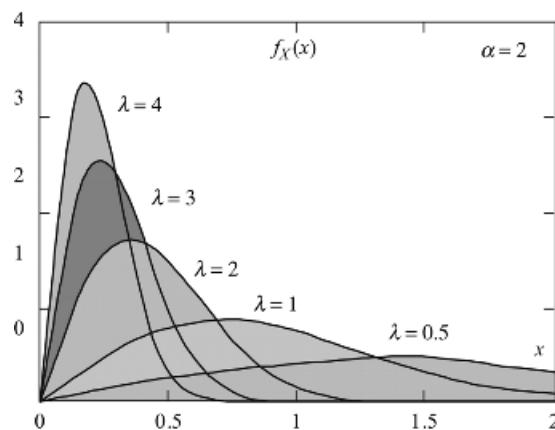


FIGURE 7.6.1

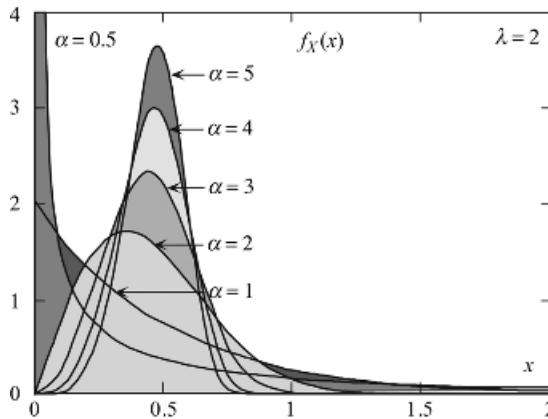


FIGURE 7.6.2

The Weibull distribution was originally used for modeling fatigue data and is at present used extensively in engineering problems for modeling the distribution of the lifetime of an object that consists of several parts and that fails if any component part fails.

## 7.7 CHI-SQUARE DISTRIBUTION

The Laplace, Erlang, gamma, and Weibull distributions are all based on exponential distributions. The chi-square distribution straddles both the exponential and Gaussian distributions. The general gamma distribution has parameters  $(\alpha, \lambda)$ . In the gamma distribution, if we substitute  $\alpha = n/2$  and  $\lambda = \frac{1}{2}$ , we obtain the standard chi-square distribution, defined by

$$f_X(x) = \begin{cases} \frac{x^{(n/2)-1} e^{-(x/2)}}{2^{n/2} \Gamma(n/2)}, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (7.7.1)$$

where  $n$  is the degrees of freedom for the  $\chi^2$  distribution. We can introduce another parameter  $\sigma^2$  and rewrite Eq. (7.7.1) as

$$f_X(x) = \begin{cases} \frac{x^{(n/2)-1} e^{-(x/2\sigma^2)}}{(2\sigma^2)^{n/2} \Gamma(n/2)}, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (7.7.2)$$

If we substitute  $x = \chi^2$  in Eq. (7.7.2), we obtain the familiar form of the chi-square distribution, given by

$$f_X(\chi^2) = \frac{\chi^{2[(n/2)-1]} e^{-(\chi^2/2\sigma^2)}}{(2\sigma^2)^{n/2} \Gamma(n/2)} \quad (7.7.3)$$

where  $\Gamma(x)$  is the gamma function defined in Eq. (7.5.2). If  $n$  is even in Eq. (7.7.2), then  $\Gamma(n/2) = (n/2 - 1)!$ , and if  $n$  is odd, then  $\Gamma(n/2)$  in Eqs. (7.7.1)–(7.7.3) is obtained by iterating  $(n/2 - 1)\Gamma(n/2 - 1)$  with the final iterant  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ .

Equation (7.7.3) can also be obtained from the following. If  $X_1, X_2, \dots, X_n$  are  $n$  i.i.d. Gaussian random variables with zero mean and variance  $\sigma^2$ , then  $X = \sum_{k=1}^n X_k^2$  is  $\chi^2$ -distributed with  $n$  degrees of freedom.

The cumulative chi-square distribution function  $F_X(x)$  is given by integrating Eq. (7.7.2):

$$F_X(x) = \begin{cases} \int_0^x \frac{\xi^{n/2-1} e^{-(\xi/2\sigma^2)}}{(2\sigma^2)^{n/2} \Gamma(n/2)} d\xi & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (7.7.4)$$

This equation can be given in terms of the incomplete gamma function  $\Gamma_x(\alpha)$  given by Eq. (7.5.5). Substituting  $\eta = \xi/2\sigma^2$  and  $d\xi = 2\sigma^2 d\eta$  in Eq. (7.7.4), we obtain

$$\begin{aligned} F_X(x) &= \int_0^x \frac{\xi^{n/2-1} e^{-(\xi/2\sigma^2)}}{(2\sigma^2)^{n/2-1} \Gamma(n/2)} 2\sigma^2 d\xi \\ &= \int_0^{x/2\sigma^2} \frac{\eta^{n/2-1} e^{-(\eta)}}{\Gamma(n/2)} d\eta = \frac{\Gamma_{x/2\sigma^2}(n/2)}{\Gamma(n/2)} \end{aligned} \quad (7.7.5)$$

When  $x \rightarrow \infty$  then  $\Gamma_{x/2\sigma^2}(n/2) \rightarrow \Gamma(n/2)$  and hence  $F_X(x) = 1$ . The cdf  $F_X(x)$  in Eq. (7.7.5) has been extensively tabulated.

The chi-square distribution is shown in Fig. (7.7.1) for the number of degrees of freedom  $n = 2, 3, 4, 5, 6, 10$ .

The chi-square distribution is used directly or indirectly in many tests of hypotheses. The most common use of the chi-square distribution is to test the goodness of a hypothesized fit. The goodness-of-fit test is performed to determine whether an observed value of a statistic differs enough from a hypothesized value of a distribution to draw the inference whether the hypothesized quantity is the true distribution. Its utility is shown in Chapter 15.

**Example 7.7.1** The density function of  $f_Z(z)$  of  $Z = X^2 + Y^2$ , where  $X$  and  $Y$  are independent random variables distributed as  $N(0, \sigma^2)$ , can be found by substituting  $n = 2$  in Eq. (7.7.2), yielding

$$F_Z(z) = \begin{cases} \frac{e^{-(z/2\sigma^2)}}{2\sigma^2} & z > 0 \\ 0 & z \leq 0 \end{cases}$$

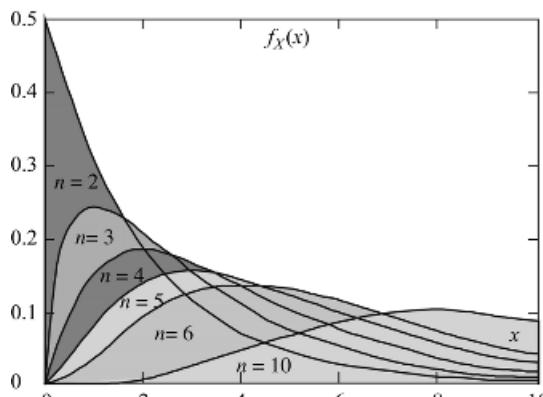


FIGURE 7.7.1

which is a  $\chi^2$  density with 2 degrees of freedom. The corresponding cdf,  $F_Z(z)$ , is given by

$$F_Z(z) = \begin{cases} 1 - e^{-(z/2\sigma^2)} & z > 0 \\ 0 & z \leq 0 \end{cases}$$

## 7.8 CHI AND OTHER ALLIED DISTRIBUTIONS

The Rayleigh and Maxwell distributions are special cases of the chi distribution. Just as the sum of the squares of  $n$  zero mean Gaussian random variables with variance  $\sigma^2$  is chi-square-distributed with  $n$  degrees of freedom, we can consider that the sum of the square roots of the sum of the squares of  $n$  zero mean Gaussian random variables with variance  $\sigma^2$  is chi-distributed with  $n$  degrees of freedom. Accordingly, let  $Z_n$  be the random variable given by

$$Z_n = \sqrt{X_1^2 + X_2^2 + \cdots + X_n^2} \quad (7.8.1)$$

where  $\{X_i : i = 1, \dots, n\}$  are zero mean Gaussian random variables with variance  $\sigma^2$ . Then the  $\chi$ -density  $f_Z(z)$  is given by

$$f_Z(z) = \begin{cases} \frac{1}{2^{n/2-1}\Gamma(n/2)} \frac{z^{n-1}}{\sigma^n} e^{-(1/2)(z/\sigma)^2} & z > 0 \\ 0 & z \leq 0 \end{cases} \quad (7.8.2)$$

where  $\Gamma(n/2)$  is the gamma function defined in Eq. (7.5.2).

A family of chi distributions is shown in Fig. 7.8.1 for  $\sigma = 1.5$  and  $n = 1, 2, 3, 5, 7, 9, 11, 13, 15$ .

The distribution function  $F_Z(z)$  is given by integrating Eq. (7.8.2):

$$F_Z(z) = \int_0^z \frac{1}{2^{n/2-1}\Gamma(n/2)} \frac{\zeta^{n-1}}{\sigma^n} e^{-(1/2)(\zeta/\sigma)^2} d\zeta \quad z > 0 \quad (7.8.3)$$

There is no closed-form solution for Eq. (7.8.3), but it can be expressed in terms of the incomplete gamma function  $\Gamma_Z(z)$  defined in Eq. (7.5.5). By substituting  $y = \zeta^2/2\sigma^2$  and

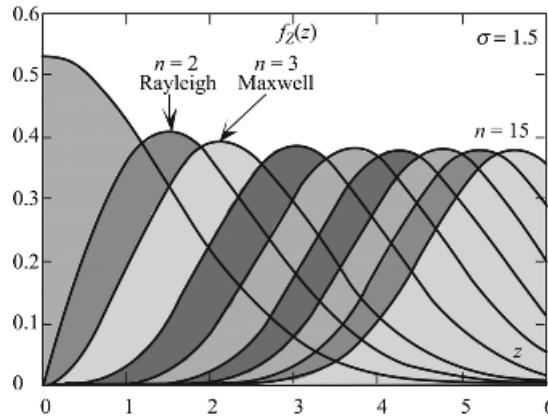


FIGURE 7.8.1

$d\xi = (\sigma^2 dy)/\zeta$  in the in Eq. (7.8.3), we can write

$$\begin{aligned} F_Z(z) &= \frac{2^{n/2-1}}{2^{n/2-1}\Gamma(n/2)} \int_0^{z^2/2\sigma^2} y^{(n/2-1)} e^{-y} dy \\ &= \frac{2^{n/2-1}}{2^{n/2-1}\Gamma(n/2)} \Gamma_{(z^2/2\sigma^2)}\left(\frac{n}{2}\right) = \frac{\Gamma_{(z^2/2\sigma^2)}(n/2)}{\Gamma(n/2)} \end{aligned} \quad (7.8.4)$$

where  $\Gamma_{(z^2/2\sigma^2)}(n/2)$  is the incomplete gamma function defined in Eq. (7.5.5). Clearly, as  $z \rightarrow \infty$ ,  $\Gamma_{(z^2/2\sigma^2)}(n/2) \rightarrow \Gamma(n/2)$ , thus showing that  $F_Z(\infty) = 1$ .

There are two very important special cases of the chi distribution when  $n = 2$  and  $n = 3$ .

### Rayleigh Distribution ( $n = 2$ )

When  $n = 2$  in the chi distribution of Eq. (7.8.2), we have

$$f_Z(z) = \begin{cases} \frac{z}{\sigma^2} e^{-(1/2)(z/\sigma)^2} & z > 0 \\ 0 & z \leq 0 \end{cases} \quad (7.8.5)$$

This is called the *Rayleigh density* and finds wide application in communications fading channels. If we have two independent zero mean Gaussian random variables  $X$  and  $Y$  with variance  $\sigma^2$ , and if we make the transformation  $X = Z \cos(\Theta)$  and  $Y = Z \sin(\Theta)$ , then  $Z = \sqrt{X^2 + Y^2}$  is Rayleigh-distributed, which is a chi distribution with 2 degrees of freedom.  $\Theta$  is uniformly distributed in the interval  $(0, 2\pi]$ . The Rayleigh density is shown in Fig. 7.8.2 for  $\sigma = \sqrt{10}, 4, 5$ .

The cdf corresponding to the Rayleigh density can be expressed in a closed form as follows:

$$\begin{aligned} F_Z(z) &= \int_0^z \frac{\xi}{\sigma^2} e^{-\xi^2/2\sigma^2} d\xi \quad \text{or,} \\ &= \begin{cases} 1 - e^{-z^2/2\sigma^2} & z > 0 \\ 0 & z \leq 0 \end{cases} \end{aligned} \quad (7.8.6)$$

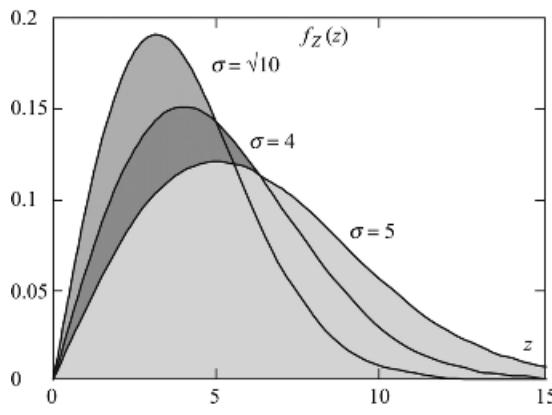


FIGURE 7.8.2

**Maxwell Density ( $n = 3$ )**

This is one of the earliest distributions derived by Maxwell and Boltzmann from statistical mechanics considerations. They derived that the velocity  $v_i$  of gas molecules in any direction  $i$ , based on the assumptions that the molecules are identical and distinguishable, and can occupy any energy state, follow the probability density function

$$f_V(v_i) = \frac{e^{-(\epsilon_i/kT)}}{Z} \quad (7.8.7)$$

where  $\epsilon_i$  is the energy function given by  $\epsilon_i = mv_i^2/2$  with  $m$  as the mass of the particle,  $k$  is the Boltzmann constant, and  $T$  is the absolute temperature.  $Z$  is the normalization constant found from integrating Eq. (7.8.7) over all  $v_i$ , or

$$\frac{1}{Z} \int_{-\infty}^{\infty} e^{-(mv_i^2/2kT)} dv_i = 1 \quad \text{and} \quad Z = \sqrt{\frac{2\pi kT}{m}}$$

so that the velocity density function along any direction follows a Gaussian distribution given by

$$f_V(v_i) = \sqrt{\frac{m}{2\pi kT}} e^{-(mv_i^2/2kT)}, \quad i = x, y, z \quad (7.8.8)$$

From Eq. (7.8.8) the velocity vector  $\mathbf{V} = \{v_x, v_y, v_z\}$  will be distributed with density function,

$$f_{\mathbf{V}}(v_x, v_y, v_z) = \left( \frac{m}{2\pi kT} \right)^{3/2} \exp \left[ -\frac{m(v_x^2 + v_y^2 + v_z^2)}{2kT} \right] \quad (7.8.9)$$

Eq. (7.8.9) is a chi-square density with 3 degrees of freedom. The speed  $x$  of gas molecules will be defined by

$$x = \sqrt{v_x^2 + v_y^2 + v_z^2}$$

Hence  $x$  is a distributed as a chi-density with three degrees of freedom. Its probability density can be derived as

$$f(x) = 4\pi \left[ \frac{m}{2\pi kT} \right]^{3/2} x^2 e^{-mx^2/2kT}, \quad x > 0 \quad (7.8.10)$$

If we substitute  $kT/m = \sigma^2$  in the Maxwell–Boltzmann distribution of Eq. (7.8.10), we obtain

$$f_X(x) = \sqrt{\frac{2}{\pi}} \frac{x^2}{\sigma^3} e^{-x^2/2\sigma^2}, \quad x > 0 \quad (7.8.11)$$

Equation (7.8.11) is a special case of the chi density of Eq. (7.8.2) with  $n = 3$  degrees of freedom. The pdf values  $f_X(x)$  for the Maxwell and Rayleigh densities are compared in Fig. 7.8.3 for values of  $\sigma = \sqrt{10}$ , 4, and 5.

The cdf  $F_X(x)$  corresponding to the Maxwell density  $f_X(x)$  is obtained by integrating Eq. (7.8.11) and is given by

$$F_X(x) = \sqrt{\frac{2}{\pi}} \int_0^x \frac{\xi^2}{\sigma^3} e^{-\xi^2/2\sigma^2} d\xi \quad (7.8.12)$$

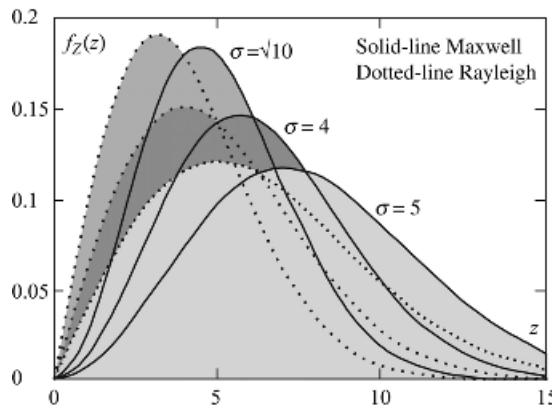


FIGURE 7.8.3

There is no closed form solution for Eq. (7.8.12). However, integrating Eq. (7.8.12) by parts, it can be expressed as

$$\begin{aligned} F_X(x) &= \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \int_0^x e^{-\xi^2/2\sigma^2} d\xi - \sqrt{\frac{2}{\pi}} \frac{x}{\sigma} e^{-x^2/2\sigma^2} \\ &= \operatorname{erf}\left(\frac{x}{\sqrt{2}\sigma}\right) - \sqrt{\frac{2}{\pi}} \frac{x}{\sigma} e^{-x^2/2\sigma^2} \quad x > 0 \end{aligned} \quad (7.8.13)$$

where  $\operatorname{erf}(x)$  is as defined in Eq. (6.4.6).

### Rice's Density

In the Rayleigh distribution discussed earlier, the random variables  $X$  and  $Y$  must be zero mean with the same variances  $\sigma^2$ . In many communication problems involving fading channels,  $X$  and  $Y$  will not be zero mean and will have mean values equal to  $\mu_X$  and  $\mu_Y$ . If the random variable  $Z$  is given by  $Z = \sqrt{X^2 + Y^2}$ , then  $f_Z(z)$  has a Rice density [49,50] with a noncentrality parameter  $m = \sqrt{\mu_X^2 + \mu_Y^2}$ , given by

$$f_Z(z) = \frac{z}{\sigma^2} \exp\left[\frac{-(z^2 + m^2)}{2\sigma^2}\right] I_0\left(\frac{mz}{\sigma^2}\right) \quad (7.8.14)$$

where  $I_0(z)$  is the zeroth-order modified Bessel function of the first kind. The general  $v$ th-order modified Bessel function is given by the series

$$I_v(z) = \left(\frac{z}{2}\right)^v \sum_{k=0}^{\infty} \frac{(z^2/4)^k}{k! \Gamma(v+k+1)} \quad (7.8.15)$$

If  $v$  is an integer, then  $I_v(z)$  can also be given by

$$I_v(z) = \frac{1}{\pi} \int_0^{\pi} e^{z \cos \theta} \cos(v\theta) d\theta \quad (7.8.16)$$

and  $I_0(z)$  the zeroth-order modified Bessel function by

$$I_0(z) = \frac{1}{\pi} \int_0^{\pi} e^{z \cos \theta} d\theta = \sum_{k=0}^{\infty} \frac{(z^2/4)^k}{(k!)^2} \quad (7.8.17)$$

and  $I_0(z)$ ,  $I_1(z)$ ,  $I_2(z)$ ,  $I_3(z)$  are shown in Fig. 7.8.4.

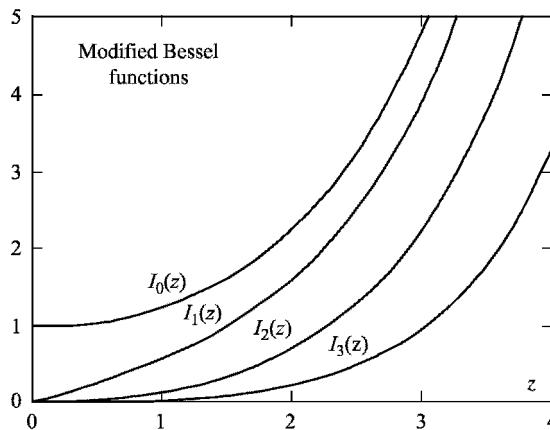


FIGURE 7.8.4

From Fig. 7.8.4 we see that  $I_0(0) = 1$ , and hence when  $m = 0$  (zero mean), then  $f_Z(z)$  is Rayleigh-distributed. Rice's density is shown in Fig. 7.8.5 for  $m = 0, 2, 4, 6, 8, 10, 12, 14$ , and  $m = 0$  corresponds to the Rayleigh density.

The distribution function  $F_Z(z)$  corresponding to Eq. (7.8.14) is given by integrating  $f_Z(z)$ :

$$F_Z(z) = \int_0^z \frac{\xi}{\sigma^2} \exp\left[\frac{-(\xi^2 + m^2)}{2\sigma^2}\right] I_0\left(\frac{m\xi}{\sigma^2}\right) d\xi \quad (7.8.18)$$

There is no closed-form solution for this integral.

### Nakagami Density

The Nakagami distribution [49], like the Rayleigh density and unlike Rice's distribution, also belongs to the class of central chi-square distributions. It is used for modeling data from multipath fading channels and has been shown to fit empirical results more generally than the Rayleigh distributions. Sometimes, the pdf of the amplitude of a mobile signal can also be described by the Nakagami  $m$  distribution.

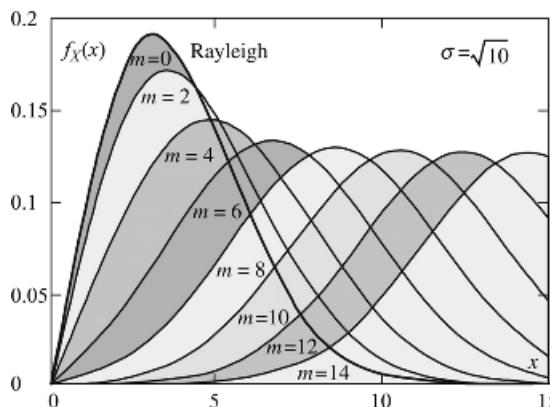


FIGURE 7.8.5

The Nakagami distribution may be a reasonable means to characterize the back scattered echo from breast tissues to automate a scheme for separating benign and malignant breast masses. The shape parameter  $m$  has been shown to be useful in tissue characterization. Chi-square tests showed that this distribution is a better fit to the envelope than the Rayleigh distribution. Two parameters,  $\mu$  (effective number) and  $\Omega$  (effective cross section), associated with the Nakagami distribution are used for the effective classification of breast masses.

The Nakagami  $m$  density is defined by

$$f_X(x) = \frac{2}{\Gamma(m)} \left[ \frac{m}{\Omega} \right]^m x^{2m-1} e^{-mx^2/\Omega} \quad (7.8.19)$$

where  $m$  is the shape parameter and  $\Omega$  controls the spread of the distribution. The interesting feature of this density is that, unlike the Rice distribution, it is not dependent on the modified Bessel function. The Nakagami family is shown in Fig. 7.8.6 for  $\Omega = 1$  and for  $m = 3, 2, 1.5, 1, 0.75, 0.5$ .

The cdf  $F_X(x)$  is given by

$$F_X(x) = \int_0^x \frac{2}{\Gamma(m)} \left[ \frac{m}{\Omega} \right]^m \xi^{2m-1} e^{-m\xi^2/\Omega} d\xi \quad (7.8.20)$$

This has no closed form solution. However, it can be given in terms of the incomplete gamma function by substituting  $y = m\xi^2/\Omega$  and  $d\xi = (\Omega/2m\xi) dy$  in Eq. (7.8.20). This results in

$$\begin{aligned} F_X(x) &= \int_0^x \frac{2}{\Gamma(m)} \left[ \frac{m\xi^2}{\Omega} \right]^m \frac{1}{\xi} e^{-m\xi^2/\Omega} d\xi \\ &= \int_0^{mx^2/\Omega} \frac{2}{\Gamma(m)} \left[ \frac{m\xi^2}{\Omega} \right]^m e^{-m\xi^2/\Omega} \frac{1}{2(m\xi^2/\Omega)} dy \\ &= \int_0^{mx^2/\Omega} \frac{2}{\Gamma(m)} y^m e^{-y} \frac{1}{2y} dy \\ &= \int_0^{mx^2/\Omega} \frac{1}{\Gamma(m)} y^{m-1} e^{-y} dy = \frac{\Gamma_{mx^2/\Omega}(m)}{\Gamma(m)} \end{aligned} \quad (7.8.21)$$

Since  $\Gamma_{mx^2/\Omega}(m) \rightarrow \Gamma(m)$  as  $x \rightarrow \infty$  we have  $F_X(\infty) = 1$ .

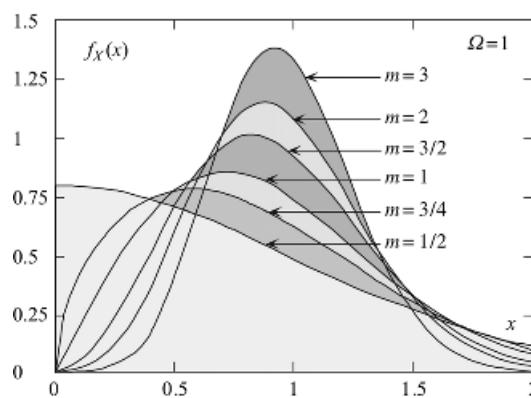


FIGURE 7.8.6

## 7.9 STUDENT-*t* DENSITY

The *t* distribution first developed by William Gosset under the pseudonym “Student” is used to estimate the confidence intervals for the unknown population mean where the variance is also unknown. In this case the statistic used for determining the confidence interval is given by

$$T = \frac{\hat{\mu}_X - \mu_X}{\hat{\sigma}/\sqrt{v}} \quad (7.9.1)$$

and the distribution of  $T$  is governed by the Student-*t* distribution with density given by

$$f_T(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{t^2}{v}\right)^{-[(v+1)/2]} \quad -\infty < t < \infty \quad (7.9.2)$$

where  $v$  is the number of degrees of freedom of  $T$ . It will be shown in Example 13.4.5 that if  $Z$  is a standard Gaussian random variable,  $W_v$  is a chi-square random variable with  $v$  degrees of freedom, and  $Z$  and  $W_v$  are independent, then the random variable  $T = Z/\sqrt{W_v/v}$  is *t*-distributed with  $v$  degrees of freedom.

The Student-*t* distribution with  $v = 2$  is compared to a standard Gaussian in Fig. 7.9.1.

After integrating the density function, we can write the distribution function  $F_T(t)$  as

$$F_T(t) = \int_{-\infty}^t \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{\tau^2}{v}\right)^{-[(v+1)/2]} d\tau \quad (7.9.3)$$

This is shown in Fig. 7.9.2.

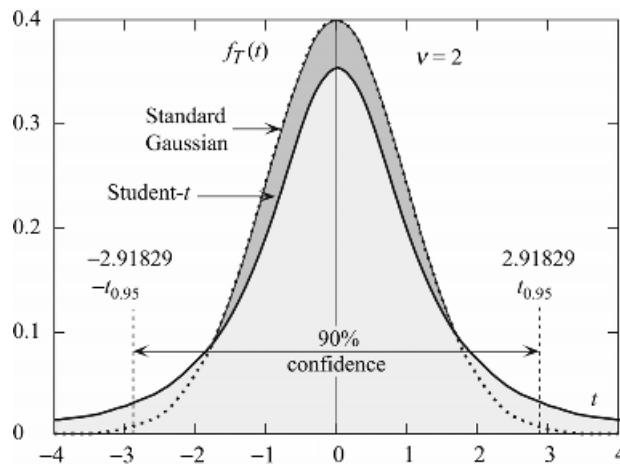


FIGURE 7.9.1

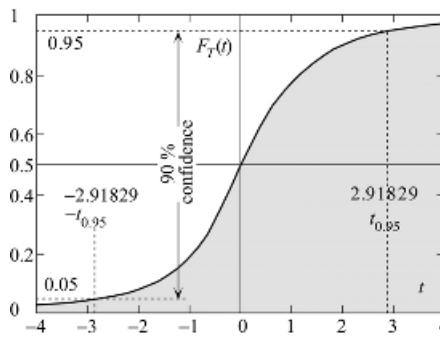


FIGURE 7.9.2

### Limit of Student-*t* as $v \rightarrow \infty$

We shall find the limit as  $v \rightarrow \infty$  in Eq. 7.9.2. We have

$$\begin{aligned} \lim_{v \rightarrow \infty} \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2} &= e^{-t^2/2} \\ \lim_{v \rightarrow \infty} \frac{\Gamma\left(v + \frac{1}{2}\right)}{\sqrt{v\pi} \Gamma\left(\frac{v}{2}\right)} &= \frac{1}{\sqrt{2\pi}} \end{aligned} \quad (7.9.4)$$

and hence

$$\lim_{v \rightarrow \infty} \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi} \Gamma(\frac{v}{2})} \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2} = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \quad (7.9.5)$$

that is, as the number of degrees of freedom  $v$  goes to infinity, the Student-*t* tends to a standard Gaussian.

If  $X_1, X_2, \dots, X_n$  are  $n$  independent identically distributed Gaussian random variables with mean  $\mu_X$  and standard deviation  $\sigma_X$ , then  $\hat{\mu}_X = (1/n) \sum_{i=1}^n X_i$  is also Gaussian-distributed with mean  $\mu_X$  and variance  $\sigma_X^2/n$ . Since we do not know  $\sigma_X^2$ , we use the sample variance given by  $\hat{\sigma}_X^2 = 1/(n-1) \sum_{i=1}^n (X_i - \hat{\mu}_X)^2$  and  $\hat{\sigma}_X^2$  is chi-square-distributed with  $n-1$  degrees of freedom. Under these conditions the random variable

$$T = \frac{(\hat{\mu}_X - \mu_X)}{\hat{\sigma}_X / \sqrt{n}} \quad (7.9.6)$$

has a *t* distribution with  $n-1$  degrees of freedom.

## 7.10 SNEDECOR *F* DISTRIBUTION

If  $X_1$  and  $X_2$  are two independent chi-squared-distributed random variables with  $v_1$  and  $v_2$  degrees of freedom, then the density of the random variable  $F = (X_1/v_1)/(X_2/v_2)$  is

$$f_{v_1, v_2}(x) = \frac{\Gamma\left(\frac{v_1 + v_2}{2}\right) \left(\frac{v_1}{v_2}\right)^{v_1/2}}{\Gamma\left(\frac{v_1}{2}\right) \Gamma\left(\frac{v_2}{2}\right) \left(\frac{1+v_1}{v_2}x\right)^{(v_1+v_2)/2}} x^{(v_1-2)/2}, \quad x > 0 \quad (7.10.1)$$

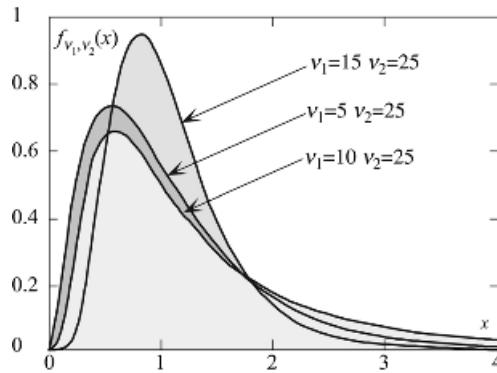


FIGURE 7.10.1

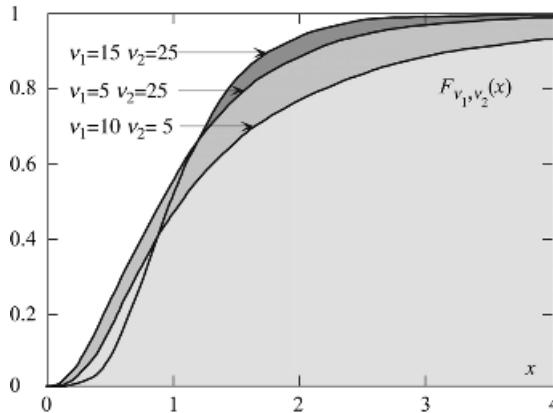


FIGURE 7.10.2

This density is shown in Fig. 7.10.1 for three cases:  $v_1 = 15, v_2 = 25$ ;  $v_1 = 5, v_2 = 25$ ;  $v_1 = 10, v_2 = 5$ . Note that in general  $f_{v_1, v_2}(x)$  is not equal to  $f_{v_2, v_1}(x)$ .

The corresponding distribution function is given by

$$F_{v_2, v_1}(x) = \int_0^x \frac{\Gamma\left(\frac{v_1 + v_2}{2}\right) \left(\frac{v_1}{v_2}\right)^{v_1/2}}{\Gamma\left(\frac{v_1}{2}\right) \Gamma\left(\frac{v_2}{2}\right) \left(\frac{1 + v_1}{v_2} \xi\right)^{(v_1+v_2)/2}} \xi^{(v_1-2)/2} d\xi \quad (7.10.2)$$

Equation (7.10.2) is shown in Fig. 7.10.2 for the same values of  $v_1$  and  $v_2$  as in Fig. 7.10.1. As  $v_1$  and  $v_2$  tend to infinity, then the density function becomes more sharply spiked around 1. The  $F$  distribution is used to draw inferences about the population variances in the case of two-sample situations.

## 7.11 LOGNORMAL DISTRIBUTION

A random variable  $X$  has lognormal density with parameters  $\alpha$ ,  $\mu$ , and  $\sigma^2$  if the transformed random variable  $Y = \ln(X)$  has a normal distribution with mean  $\mu$  and variance

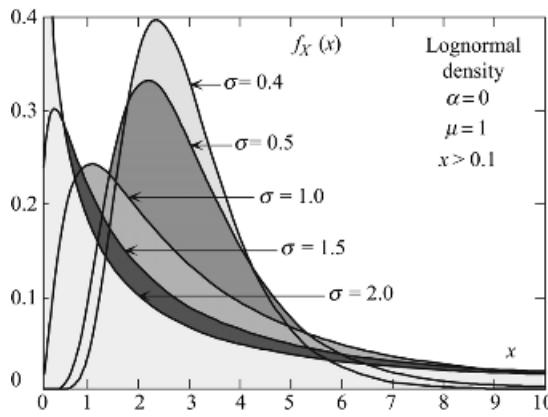


FIGURE 7.11.1

$\sigma^2$ . This density, defined only for positive values of  $x$ , is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma(x-\alpha)} \exp\left\{-\frac{1}{2}\left[\frac{\ln(x-\alpha)-\mu}{\sigma}\right]^2\right\}, \quad x > \alpha; \mu, \sigma > 0 \quad (7.11.1)$$

where  $\alpha$  is the location parameter,  $\mu$  is the scale parameter, and  $\sigma$  is the shape parameter. When  $\alpha$  and  $\mu$  are zero,  $f_X(x)$  is called the *standard lognormal distribution* with parameter  $\sigma$ . The case where  $\alpha = 0$  and  $\mu = 1$  is shown in Fig. 7.11.1 for  $\sigma = 0.4, 0.5, 1.0, 1.5, 2$  for values of  $x > 0.1$ .

The cumulative distribution function  $F_X(x)$  is obtained by integrating Eq. (7.11.1) and is given by

$$\begin{aligned} F_X(x) &= \int_0^x \frac{1}{\sqrt{2\pi}\sigma(\xi-\alpha)} e^{-(\ln(\xi-\alpha)-\mu)^2/2\sigma^2} d\xi \\ &= \int_0^{\frac{\ln(x-\alpha)-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\eta^2/2} d\eta = \frac{1}{2} \operatorname{erf}\left(\frac{\ln(x-\alpha)-\mu}{\sqrt{2}\sigma}\right) \end{aligned} \quad (7.11.2)$$

where

$$\int_0^x \frac{1}{\sqrt{2\pi}} e^{-\eta^2/2} d\eta = \frac{1}{2} \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)$$

The cdf  $F_X(x)$  is shown in Fig. 7.11.2 for  $\alpha = 0$ ,  $\mu = 1$  and for the same values of  $\sigma$  as the density function.

In finance, prices for primary instruments such as stocks and bonds are often assumed to be lognormally distributed. A standard assumption for the evolution of a stock price is that in an interval  $\Delta t$  the fractional increase in the stock price,  $\Delta S/S$ , is normally distributed with mean  $\mu\Delta t$  and variance  $\sigma^2\Delta t$ . The lognormal distribution is also used to model the lifetime of units whose failures are of a fatigue-stress nature. Consequently, the lognormal distribution is an adjunct to the Weibull distribution when attempting to model these types of units.

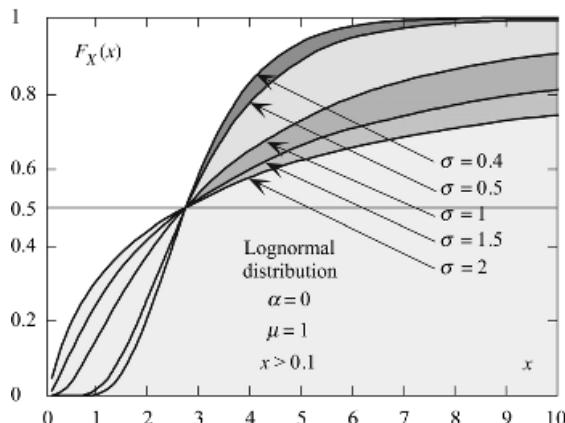


FIGURE 7.11.2

## 7.12 BETA DISTRIBUTION

The beta distribution is a two-parameter  $\alpha$  and  $\beta$  family of density functions  $f_X(x)$  for  $0 \leq x \leq 1$ . It is often used to represent proportions and percentages. It has the following probability density function:

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 \leq x \leq 1 \quad (7.12.1)$$

The choices of the two parameters  $\alpha$  and  $\beta$  makes it amenable to a variety of shapes. For example, for  $\alpha = \beta = 1$ ,  $f_X(x)$  becomes the rectangular distribution. The triangular distribution arises when one parameter has the value 2 and the other is unity. If both parameters are less than unity and positive, the density becomes U-shaped, and if only one of the parameters is less than unity, it becomes J-shaped. The distribution is unimodal when both parameters are positive, and symmetric when they are equal. The pdf curves are shown in Fig. 7.12.1 for the cases described above. The beta distribution is related to the Student- $t$  distribution [Eq. (7.9.2)] by the following transformation. If  $Z$  is a  $t$ -distributed

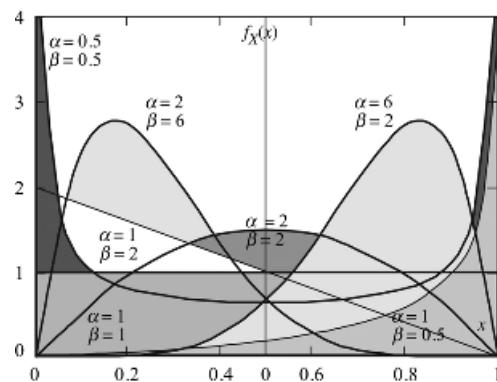


FIGURE 7.12.1

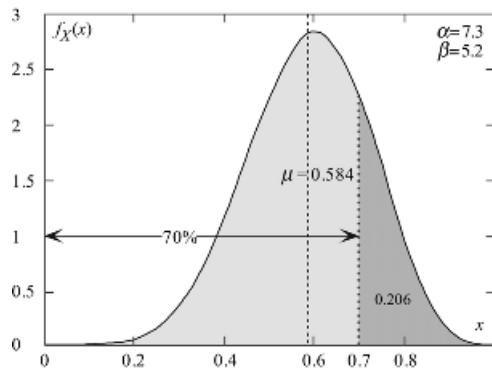


FIGURE 7.12.2

random variable with  $v$  degrees of freedom, then the transformation

$$X = \frac{1}{2} + \frac{Z}{2\sqrt{v + Z^2}}$$

is beta-distributed with  $\alpha = v/2$  and  $\beta = v/2$ .

**Example 7.12.1** We will give an example of how the beta distribution is used in stock-market analysis. The analyst from his model of stock performances predicts before the start of each day that the proportion of the listed stocks has a beta distribution with parameters  $\alpha = 7.3$  and  $\beta = 5.2$ . The average value is  $\alpha/(\alpha + \beta) = 0.584$ . We want to find the probability that more than 70% of the stocks increase in value and this is given by

$$P\{X > 0.7\} = 1 - \int_0^{0.7} f_X(x) dx = 1 - 0.794 = 0.206$$

This probability is 20.6%, as illustrated in Fig. 7.12.2.

## 7.13 CAUCHY DISTRIBUTION

Cauchy distribution resembles a normal distribution in appearance but has much heavier tails. However, in studying hypothesis tests that assume normality, seeing how the tests perform on data from a Cauchy distribution is a good indicator of how sensitive the tests are to heavy-tail departures from normality. The general Cauchy density is given by

$$f_X(x) = \frac{\beta}{\pi[(x - \alpha)^2 + \beta^2]} \quad (7.13.1)$$

where  $\alpha$  is the location parameter and  $\beta$  is the shape parameter. It is an example of a pathological case in the sense the moments do not exist even though the first moment can be evaluated to be equal to 0 using the Cauchy principle value method. The Cauchy density is plotted for  $\alpha = 5$  and  $\beta = 2, 3, 4, 5$  in Fig. 7.13.1.

Since the moments do not exist for the Cauchy distribution, it is characterized by the width at half the maximum value equal to  $2\beta$ , called *full width half maximum* (FWHM). For  $\beta = 2$  the maximum is 0.159154 and the half-maximum is 0.079577. This is shown in Fig. 7.13.2. For comparison, a Gaussian distribution is superposed on the same figure for the same FWHM of 4 that is, a Gaussian with mean 5 and  $\sigma = 4/2.35482 = 1.69864$ .

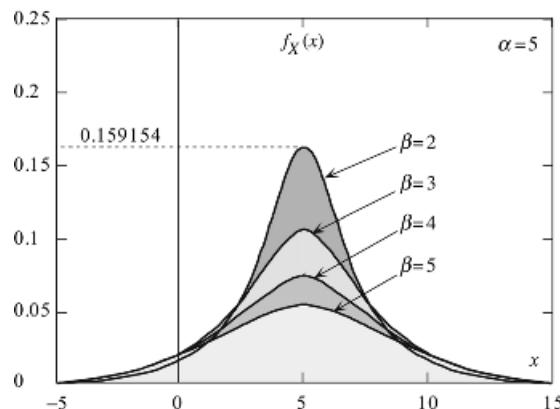


FIGURE 7.13.1

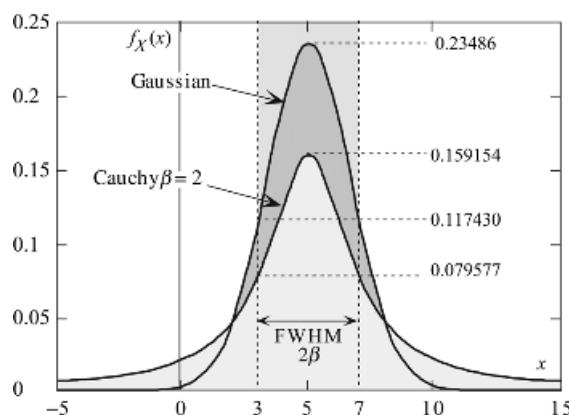


FIGURE 7.13.2

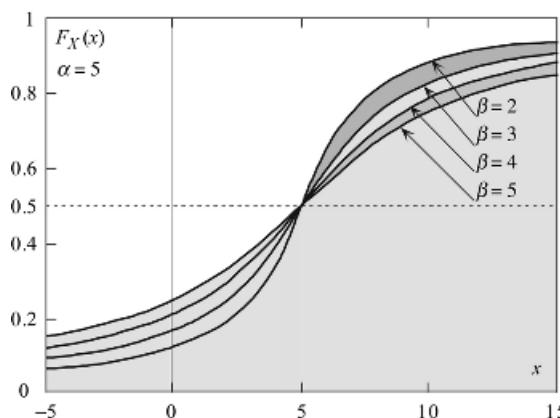


FIGURE 7.13.3

The distribution function  $F_X(x)$  can be determined as follows:

$$F_X(x) = \int_0^x \frac{1}{\sqrt{2\pi}\sigma(\xi - \alpha)} e^{-(\ln(\xi - \alpha) - \mu)^2/2\sigma^2} d\xi \quad (7.13.2)$$

Substituting  $(\xi - \alpha)/\beta = \tan(\theta)$  in Eq. (7.13.2), we have

$$\begin{aligned} F_X(x) &= \int_{-\pi/2}^{\tan^{-1}\{(\xi-\alpha)/\beta\}} \frac{\sec^2(\theta)d\theta}{\pi \sec^2(\theta)} \\ &= \frac{\theta}{\pi} \Big|_{-\pi/2}^{\tan^{-1}\{(\xi-\alpha)/\beta\}} = \frac{1}{\pi} \tan^{-1}\left(\frac{x-\alpha}{\beta}\right) + \frac{1}{2} \end{aligned} \quad (7.13.3)$$

The term  $F_X(x)$  is shown in Fig. 7.13.3 for the same values of  $\alpha$  and  $\beta$  as in Fig. 7.13.1.

## 7.14 PARETO DISTRIBUTION

Many synthetic and naturally occurring phenomena, including city sizes, incomes, word frequencies, and earthquake magnitudes, are distributed according to a power-law distribution. A *power law* implies that small occurrences are extremely common, whereas large instances are extremely rare. These can be modeled as a Pareto distribution. In addition, traffic measurements reveal features such as nonstationarity, long-range dependence, and significant fluctuations in different timescales (e.g., self-similarity). The essential features of the burstiness observed in measurements can be modeled as a heavy-tailed on-time and off-time distributions such as the Pareto distribution.

The Pareto density is given by

$$f_X(x) = \frac{\alpha}{\beta} \left[ \frac{\beta}{x + \beta} \right]^{\alpha+1}, \quad x > 0 \quad (7.14.1)$$

and the distribution function can be obtained from integrating Eq. (7.14.1):

$$F_X(x) = \int_0^x \frac{\alpha}{\beta} \left[ \frac{\beta}{\xi + \beta} \right]^{\alpha+1} d\xi = 1 - \left[ \frac{\beta}{x + \beta} \right]^\alpha \quad (7.14.2)$$

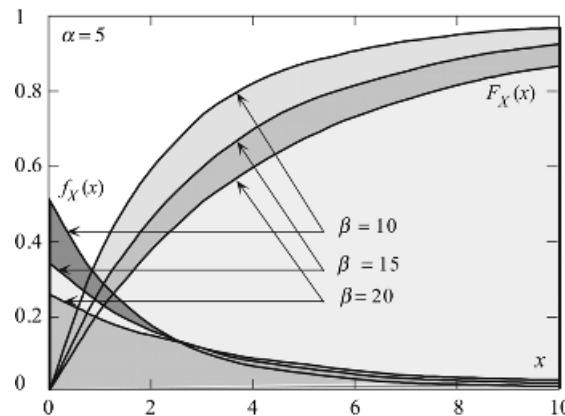


FIGURE 7.14.1

The Pareto density and the distribution functions are plotted in Fig. 7.14.1 for  $\alpha = 5$  and  $\beta = 10, 15, 20$ . Note that the tails are long compared to other distributions. One of the uses of the Pareto distribution is to model the high end of the distribution of incomes.

## 7.15 GIBBS DISTRIBUTION

The Gibbs distribution has been widely used in image processing and pattern recognition problems. An example using this distribution in tomographic imaging is presented in Chapter 23. It is a generalization of the Maxwell–Boltzmann distribution and has origins in statistical mechanics. It is defined on a random field  $N$  as

$$\text{where, } P(\omega) = \frac{1}{Z} e^{-(1/T)U(\omega)} \quad (7.15.1)$$

$U(\omega)$	$\sum_{j,k \in N} w_{jk} V(\omega_j - \omega_k)$	Total energy function
$V(\omega)$		Potential function, where $\omega$ is some vector field
$w_{jk}$		Weighting functions
$j, k$		Pixels in neighboring field $N$
$T$		Absolute temperature that controls sharpness of distribution
$Z$		Normalization constant given by $Z = \sum_{\omega} e^{-(1/T)U(\omega)}$ , also called the <i>partition function</i>

The potential function  $V(\omega)$  depends on the local configuration of the field. Some of the properties of potential functions [38] are given below:

1.  $V(0) = 0$
2.  $V(\omega) \geq 0$ : nonnegative
3.  $V(\omega) = V(-\omega)$ : even function
4.  $\frac{d}{d\omega} V(\omega) > 0$  for  $\omega > 0$
5.  $\lim_{\omega \rightarrow \infty} V(\omega) \rightarrow \infty$
6.  $\sup_{0 \leq \omega < \infty} \frac{d}{d\omega} V(\omega) < \infty$
7.  $\frac{d^2}{d\omega^2} V(\omega) > \infty$

## 7.16 MIXED DISTRIBUTIONS

There are any number of derived and mixed distributions, depending on the application, and new distributions are also constantly being formulated. For example, the double-double-exponential or double-Laplace distribution, given by

$$f(x) = \frac{p}{2\alpha} e^{-\left(\frac{|x|}{\alpha}\right)} + \frac{1-p}{2\beta} e^{-\left(\frac{|x|}{\beta}\right)} \quad (7.16.1)$$

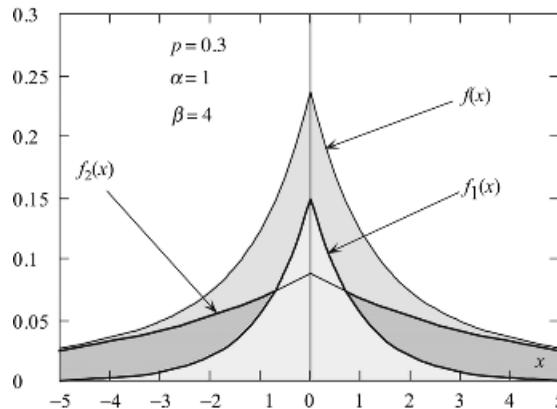


FIGURE 7.16.1

with

$$f_1(x) = \frac{p}{2\alpha} e^{-\left(\frac{|x|}{\alpha}\right)} : f_2(x) = \frac{1-p}{2\beta} e^{-\left(\frac{|x|}{\beta}\right)}$$

is shown in Fig. 7.16.1 with  $p = 0.3$ ,  $\alpha = 1$ ,  $\beta = 4$ .

This distribution has been specially constructed for modeling navigation errors over transatlantic flights consisting of both the pilot and instrument errors [51]. The pilot errors are modeled by  $f_2(x)$  and the instrument errors, by  $f_1(x)$ . The instrument errors have a sharper peak than do pilot errors, which extend over a greater distance. Whereas the instrument errors are easier to eliminate, the pilot errors are more difficult. The composite error curve is  $f(x)$  shown in Fig. 7.16.1.

The distributions given here are by no means exhaustive, and only some of the more important distributions are discussed. More comprehensive sets of distributions can be obtained from Web references W1 and W2 (see References at end of book). A complete list of all the densities and distributions discussed in this chapter is tabulated in Section 7.17.

## 7.17 SUMMARY OF DISTRIBUTIONS OF CONTINUOUS RANDOM VARIABLES

Name	Density Function	Closed-Form Distribution Function
Uniform	$\frac{1}{b-a}$	$\frac{x-a}{b-a}$
Triangular	$\frac{x-a}{(b-a)^2}, \quad a < x \leq b$ $\frac{-(x-2b+a)}{(b-a)^2}$ $b < x \leq 2b-a$	$\frac{1}{2} \left( \frac{x-a}{b-a} \right)^2 \quad a < x \leq b$ $1 - \frac{1}{2} \left( \frac{x+a-2b}{b-a} \right)^2$ $b < x \leq 2b-a$

(continued)

Name	Density Function	Closed-Form Distribution Function
Exponential	$\frac{1}{\lambda} e^{-\lambda x} u(x), \quad \lambda > 0$	$(1 - e^{-\lambda x}) u(x), \quad \lambda > 0$
Laplace	$\frac{\lambda}{2} e^{-\lambda x }, \quad \lambda \geq 0$	$\frac{1}{2} e^{\lambda x}, \quad x \leq 0$
		$1 - \frac{1}{2} e^{-\lambda x}, \quad x > 0$
Erlang	$\frac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!}, \quad n \text{ integer}$ $x > 0$	$1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k e^{-\lambda x}}{k!}$ $n \text{ integer}, \quad x > 0$
Gamma	$\frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad \alpha, \lambda, x > 0$	$\frac{\Gamma_x(\alpha)}{\Gamma(\alpha)}$ Incomplete gamma
Weibull	$\alpha \lambda^\alpha (x - \mu)^{\alpha-1} e^{-[\lambda(x-\mu)]^\alpha}$ $x > \mu$	$1 - e^{-[\lambda(x-\mu)]^\alpha}, \quad x > \mu$
Standard Gaussian	$\frac{1}{\sqrt{2\pi}} e^{-(x^2/2)}$	$\frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)$
Chi-square	$\frac{(\chi^2)^{(n/2)-1} e^{-\chi^2/2\sigma^2}}{(2\sigma^2)^{n/2} \Gamma\left(\frac{n}{2}\right)}$	$\frac{\Gamma_{\chi^2/2\sigma^2}\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \quad x = \chi^2$ Incomplete gamma
Chi	$\frac{1}{2^{(n/2)-1} \Gamma\left(\frac{n}{2}\right)} \frac{x^{n-1}}{\sigma^2} e^{-(1/2)(x/\sigma)^2}$ $x > 0$	$\frac{\Gamma_{x^2/2\sigma^2}\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$ Incomplete gamma
Rayleigh	$\frac{z}{\sigma^2} e^{-(z^2/\sigma^2)}, \quad z > 0$	$1 - e^{-(z^2/2\sigma^2)}, \quad z > 0$
Maxwell	$\sqrt{\frac{2}{\pi}} \frac{x^3}{\sigma^3} e^{-(x^2/2\sigma^2)}, \quad x > 0$	$\sigma\sqrt{2} \operatorname{erf}\left(\frac{x}{2\sigma^2}\right) - \sqrt{\frac{2}{\pi}} e^{-(x^2/2\sigma^2)}$ $x > 0$
Rice	$\frac{x}{\sigma^2} e^{-[(x^2+m^2)/2\sigma^2]} I_0\left(\frac{mx}{\sigma^2}\right), \quad x > 0$	No closed form

(continued)

Name	Density Function	Closed-Form Distribution Function
Nakagami	$\frac{2}{\Gamma(m)} \left(\frac{m}{\Omega}\right)^m x^{2m-1} e^{-mx^2/\Omega}$ $x > 0$	$\frac{\Gamma_{mx^2/\Omega}(m)}{\Gamma(m)}$ Incomplete gamma
Student- $t(v)$	$\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi} \Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{t^2}{v}\right)^{-[(v+1)/2]}$	No closed form
Snedecor $F(v_1, v_2)$	$\frac{\Gamma\left(\frac{v_1+v_2}{2}\right) \left(\frac{v_1}{v_2}\right)^{v_1/2} x^{(v_1-2)/2}}{\Gamma\left(\frac{v_1}{2}\right) \Gamma\left(\frac{v_2}{2}\right) \left(1 + \frac{v_1}{v_2}x\right)^{(v_1+v_2)/2}}$ $x > 0$	No closed form
Lognormal	$\frac{1}{\sqrt{2\pi} \sigma(x-\alpha)} e^{-[\ln(x-\alpha)-\mu]^2/2\sigma^2}$	$\frac{1}{2} \operatorname{erf}\left(\frac{\ln(x-\alpha)-\mu}{\sigma\sqrt{2}}\right)$
Beta	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$ $0 \leq x \leq 1$	No closed form
Cauchy	$\frac{\beta}{\pi[(x-\alpha)^2 + \beta^2]}$	$\frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{x-\alpha}{\beta}\right)$
Pareto	$\frac{\alpha}{\beta} \left(\frac{\beta}{x+\beta}\right)^{\alpha+1}, \quad x > 0$	$1 - \left(\frac{\beta}{x+\beta}\right)^\alpha, \quad x > 0$
Definitions	Zero-order modified Bessel function $I_0(x) = \frac{1}{\pi} \int_0^\pi e^{x \cos(\theta)} d\theta$	Error function $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} d\xi$ Incomplete gamma function $\Gamma_x(\alpha) = \int_0^x \xi^{\alpha-1} e^{-\xi} d\xi$

# Conditional Densities and Distributions

## 8.1 CONDITIONAL DISTRIBUTION AND DENSITY FOR $P\{A\} \neq 0$

In Section 2.2 we defined conditional probability for events  $A$  and  $B$  that are subsets of the sample space  $S$  as

$$\left. \begin{aligned} P\{B | A\} &= \frac{P\{A \cap B\}}{P\{A\}} \\ P\{A \cap B\} &= P\{B | A\}P\{A\} \end{aligned} \right\} \quad \text{if } P\{A\} \neq 0 \quad (8.1.1)$$

We will extend this to distributions  $F_X(x)$  and densities  $f_X(x)$ . The conditional distribution  $F_X(x | A)$ , given that an event  $A$  has occurred, is defined as

$$F_X(x | A) = P\{X \leq x | A\} = \frac{P\{X \leq x, A\}}{P\{A\}} \quad \text{if } P\{A\} \neq 0 \quad (8.1.2)$$

and if the event  $\{X \leq x\}$  is independent of  $A$ , then  $P\{X \leq x | A\} = P\{X \leq x\}$ . The corresponding conditional density  $f_X(x | A)$  is defined by

$$\begin{aligned} f_X(x | A) &= \frac{dF_X(x | A)}{dx} \\ &= \lim_{\Delta x \rightarrow 0} \frac{P\{x < X \leq x + \Delta x | A\}}{\Delta x} \end{aligned} \quad (8.1.3)$$

### Properties of Conditional Distributions

1.  $F_X(\infty | A) = 1; F_X(-\infty | A) = 0$
2.  $F_X(x | A)$  is right-continuous
3.  $F_X(x_2 | A) - F_X(x_1 | A) = P\{x_1 < X \leq x_2 | A\}$
4.  $F_X(x | A) = \int_{-\infty}^x f_X(\xi | A) d\xi$

**Example 8.1.1** A die is tossed and, corresponding to the faces of the die, a random variable is defined by  $X = 10k$ ,  $k = 1, 2, 3, 4, 5, 6$ . The conditioning event  $A$  is given by  $A = \{\text{the toss resulted in an even number}\}$ .

We will first find the unconditional distribution  $F_X(x)$  and then the conditional distribution  $F_{X|A}(x)$ . Using the same mapping techniques discussed in Chapter 2, the unconditional cdf is given by

$$F_X(x) = \frac{1}{6} \sum_{k=1}^6 u(x - 10k)$$

Since the sample space has been reduced to 3 points by the conditioning event  $A$ , the conditional cdf  $F_{X|A}(x)$  is given by

$$\begin{aligned} F_{X|A}(x) &= \frac{P\{X \leq 10k, X \text{ even}\}}{P\{X \text{ even}\}} = \frac{F_X(10k), k = 2, 4, 6}{1/2} \\ &= \frac{1}{3} \sum_{k=1}^3 u(x - 20k) \end{aligned}$$

Both these distributions are shown in Fig. 8.1.1.

*Conditioning Event A Depends on x.* We shall now consider the case when the event  $A$  is dependent on  $x$ . Three cases arise:

1.  $A = \{X \leq x_1\}$ : Eq. (8.1.2) takes the form:

$$F_X(x | A) = \frac{P\{X \leq x, X \leq x_1\}}{P\{X \leq x_1\}} \quad (8.1.4)$$

Two situations arise:

(a)  $x > x_1$ : Since  $x > x_1$ , the joint event  $\{X \leq x, X \leq x_1\} = \{X \leq x_1\}$ , and hence Eq. (8.1.4) can be rewritten as

$$F_X(x | A) = \frac{P\{X \leq x_1\}}{P\{X \leq x_1\}} = 1 \quad (8.1.5)$$

(b)  $x \leq x_1$ : In this case the joint event  $\{X \leq x, X \leq x_1\} = \{X \leq x\}$ , and hence Eq. (8.1.4) becomes

$$F_X(x | A) = \frac{P\{X \leq x\}}{P\{X \leq x_1\}} = \frac{F_X(x)}{F_X(x_1)} \quad (8.1.6)$$

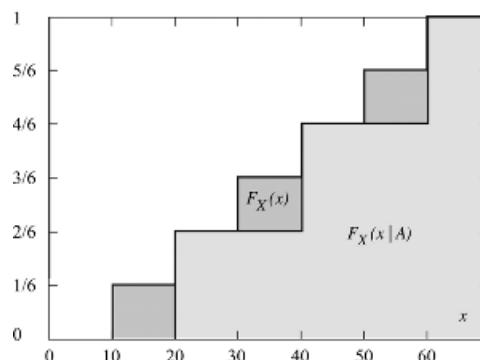


FIGURE 8.1.1

The corresponding density function is given by

$$f_X(x | A) = \begin{cases} \frac{f_X(x)}{F_X(x_1)} & x \leq x_1 \\ 0 & x > x_1 \end{cases} \quad (8.1.7)$$

2.  $A = \{x_1 < X \leq x_2\}$  with  $x_1 < x_2$

Analogous to Eq. (8.1.4), we have the following:

$$F_X(x | A) = \frac{P\{X \leq x, x_1 < X \leq x_2\}}{P\{x_1 < X \leq x_2\}} \quad (8.1.8)$$

Here three situations arise:

(a)  $x > x_2$ :

$$F_X(x | A) = \frac{P\{x_1 < X \leq x_2\}}{P\{x_1 < X \leq x_2\}} = 1 \quad (8.1.9)$$

(b)  $x_1 > x \leq x_2$ :

$$F_X(x | A) = \frac{P\{x_1 < X \leq x\}}{P\{x_1 < X \leq x_2\}} = \frac{F_X(x) - F_X(x_1)}{F_X(x_2) - F_X(x_1)} \quad (8.1.10)$$

(c)  $x \leq x_1$ :

$$F_X(x | A) = \frac{P\{\emptyset\}}{P\{x_1 < X \leq x_2\}} = 0 \quad (8.1.11)$$

The corresponding density function is given by

$$f_X(x | A) = \begin{cases} 0 & x \leq x_1 \\ \frac{f_X(x)}{F_X(x_2) - F_X(x_1)} & x_1 < x \leq x_2 \\ 0 & x > x_1 \end{cases} \quad (8.1.9)$$

**Example 8.1.2** A random variable has a Gaussian distribution, given by

$$f_X(x) = \frac{1}{2\sqrt{2\pi}} \exp\left\{-\left[\frac{1}{2}\left(\frac{x-4}{2}\right)^2\right]\right\}$$

We will find the conditional distribution  $F_X(x | A)$  and the conditional density  $f_X(x | A)$  given that (1)  $A = \{X \leq 4.5\}$  and (2)  $A = \{1.5 < X \leq 4.5\}$ . Using Eq. (8.1.6), we obtain

1.  $A = \{X \leq 4.5\}$ :

$$F_X(x | A) = \frac{F_X(x)}{F_X\left(\frac{9}{2}\right)} = \frac{1}{F_X\left(\frac{9}{2}\right)} \frac{1}{2} \left[ 1 + \operatorname{erf}\left(\frac{x-4}{2\sqrt{2}}\right) \right]$$

and using Eq. (8.1.10), we obtain

2.  $A = \{1.5 < X \leq 4.5\}$ :

$$F_X(x | A) = \frac{F_X(x) - F_X(x_1)}{F_X(x_2) - F_X(x_1)} = \frac{\frac{1}{2} \left[ 1 + \operatorname{erf}\left(\frac{x-4}{2\sqrt{2}}\right) \right] - F_X\left(\frac{3}{2}\right)}{F_X\left(\frac{9}{2}\right) - F_X\left(\frac{3}{2}\right)}$$

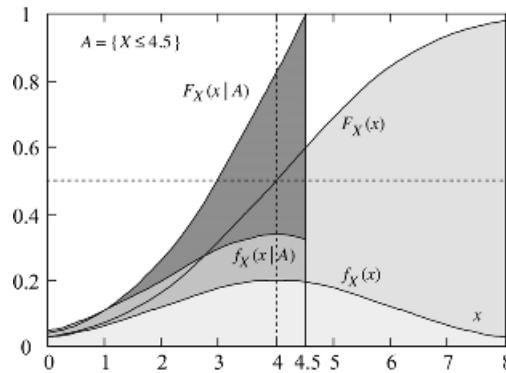


FIGURE 8.1.2

where

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} d\xi$$

The corresponding density functions are as follows for (1)

$$f_X(x | A) = \frac{\frac{1}{2\sqrt{2\pi}} \exp \left\{ -\left[ \frac{1}{2} \left( \frac{x-4}{2} \right)^2 \right] \right\}}{F_X\left(\frac{9}{2}\right)}$$

and for (2)

$$f_X(x | A) = \frac{\frac{1}{2\sqrt{2\pi}} \exp \left\{ -\left[ \frac{1}{2} \left( \frac{x-4}{2} \right)^2 \right] \right\}}{F_X\left(\frac{9}{2}\right) - F_X\left(\frac{3}{2}\right)}$$

The conditional distribution and densities  $F_X(x | A)$  and  $f_X(x | A)$  along with the unconditional distributions and densities  $F_X(x)$  and  $f_X(x)$  are shown in Fig. 8.1.2 for (1)  $A = \{X \leq 4.5\}$  and in Fig. 8.1.3 for (2)  $A = \{1.5 < X \leq 4.5\}$ .

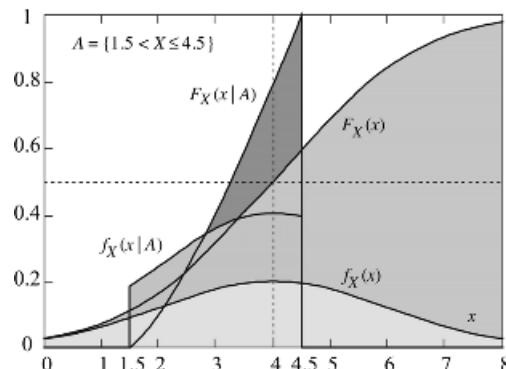


FIGURE 8.1.3

## 8.2 CONDITIONAL DISTRIBUTION AND DENSITY FOR $P\{A\} = 0$

In Section 8.1 we assumed that the probability of the conditioning event  $A$  is not zero. We will now determine the conditional probabilities when  $P\{A\} = 0$ . We need to find  $P\{B | X = x_1\}$ . Since  $P\{X = x_1\} = 0$  for a continuous random variable, we have to use limiting forms to determine this probability. We will write  $P\{B | X = x_1\}$  as

$$P\{B | X = x_1\} = \lim_{\Delta x \rightarrow 0} P\{B | x_1 < X \leq x_1 + \Delta x\} \quad (8.2.1)$$

The conditional probability  $P\{B | x_1 < X \leq x_1 + \Delta x\}$  in Eq. (8.2.1) can be written as follows:

$$\begin{aligned} P\{B | x_1 < X \leq x_1 + \Delta x\} &= \frac{P\{B, x_1 < X \leq x_1 + \Delta x\}}{P\{x_1 < X \leq x_1 + \Delta x\}} \\ &= \frac{P\{x_1 < X \leq x_1 + \Delta x | B\}P\{B\}}{F_X(x_1 + \Delta x) - F_X(x_1)} \\ &= \frac{F_X\{x_1 + \Delta x | B\} - F_X\{x_1 | B\}P\{B\}}{F_X(x_1 + \Delta x) - F_X(x_1)} \end{aligned} \quad (8.2.2)$$

Dividing numerator and denominator of Eq. (8.2.2) by  $\Delta x$  and taking the limit as  $\Delta x \rightarrow \infty$  results in

$$\begin{aligned} P\{B | X = x_1\} &= \lim_{\Delta x \rightarrow 0} \frac{\{F_X\{x_1 + \Delta x | B\} - F_X\{x_1 | B\}P\{B\}\}/\Delta x}{\{F_X(x_1 + \Delta x) - F_X(x_1)\}/\Delta x} \\ &= \frac{f_X(x_1 | B)P\{B\}}{f_X(x_1)} \end{aligned} \quad (8.2.3)$$

### Case of $B: \{Y = y\}$

In an analogous manner we can find an expression for the conditional probability  $P\{B | X = x\}$  when  $B$  is the event  $B: \{Y = y\}$ . In this case we can write

$$\begin{aligned} P\{Y = y | X = x\} &= \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} P\{y < Y \leq y + \Delta y | x < X \leq x + \Delta x\} \\ &= \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{P\{y < Y \leq y + \Delta y, x < X \leq x + \Delta x\}}{P\{x < X \leq x + \Delta x\}} \end{aligned} \quad (8.2.4)$$

Dividing the numerator and denominator of Eq. (8.2.4) by  $\Delta x \Delta y$ , we have,

$$\begin{aligned} &\lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} P\{y < Y \leq y + \Delta y | x < X \leq x + \Delta x\} \\ &= \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{[P\{y < Y \leq y + \Delta y | x < X \leq x + \Delta x\}]/\Delta x \Delta y}{[P\{x < X \leq x + \Delta x\}]/\Delta x \Delta y} \\ &= \frac{f_{XY}(x, y)}{f_X(x)} \Delta y \end{aligned}$$

Hence

$$\begin{aligned} \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{P\{y < Y \leq y + \Delta y \mid x < X \leq x + \Delta x\}}{\Delta y} &= f_{Y|X}(y \mid X = x) \\ &= f_{Y|x}(y \mid x) = \frac{f_{XY}(x, y)}{f_X(x)} \end{aligned} \quad (8.2.5)$$

The conditional distribution  $F_{Y|X}(y \mid X \leq x) = P\{Y \leq y \mid X \leq x\}$  can be obtained from the basic definition of conditional probability as follows:

$$F_{Y|X}(y \mid X \leq x) = P\{Y \leq y \mid X \leq x\} = \frac{P\{Y \leq y, X \leq x\}}{P\{X \leq x\}} = \frac{F_{XY}(x, y)}{F_X(x)} \quad (8.2.6)$$

Even though Eq. (8.2.5) is of exactly the same form as Eq. (8.2.6), it cannot be obtained from the definition of conditional probability but must be obtained from limiting considerations.

### Point of Clarity

Since  $f_{Y|X}(y \mid x)$  has been defined as  $f_{Y|X}(y \mid X = x)$ , it is not the derivative with respect to  $y$  of the distribution  $F_{Y|X}(y \mid x)$ , which has been defined as  $F_{Y|X}(y \mid X \leq x)$ . Bearing this in mind, we can write the following equations for clarity:

$$\begin{aligned} \frac{dF_{Y|X}(y \mid X \leq x)}{dy} &= \frac{dF_{Y|X}(y \mid x)}{dy} = f_{Y|X}(y \mid X \leq x) \\ \frac{dF_{Y|X}(y \mid X = x)}{dy} &= f_{Y|X}(y \mid X = x) = f_{Y|x}(y \mid x) \end{aligned} \quad (8.2.7)$$

**Example 8.2.1** The probability that a machine fails in  $x$  months is given by  $(1 - e^{-x/d})$   $u(x)$ . If the machine failed sometime before  $c$  months, it is desired to find the conditional distribution  $F_X(x \mid X \leq c)$  and the corresponding conditional density. Here we can use the definition of conditional distribution and write for  $F_X(x \mid X \leq c)$ :

$$\begin{aligned} F_X(x \mid X \leq c) &= P\{X \leq x \mid X \leq c\} = \frac{P\{X \leq x, X \leq c\}}{P\{X \leq c\}} \\ &= \begin{cases} \frac{1 - e^{-x/d}}{1 - e^{-c/d}} & 0 < x \leq c \\ 1 & x > c \\ 0 & x \leq 0 \end{cases} \end{aligned}$$

The corresponding density function for the event  $\{X \leq c\}$  is

$$f_X(x \mid X \leq c) = \begin{cases} \frac{de^{-x/d}}{1 - e^{-c/d}} & 0 < x \leq c \\ 0 & \text{otherwise} \end{cases}$$

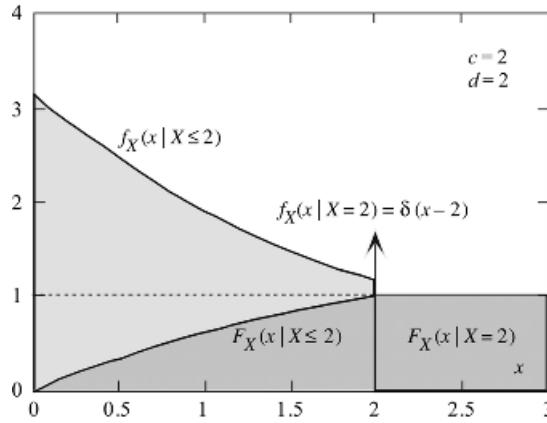


FIGURE 8.2.1

We can now find the conditional distribution  $F_X(x | X = c)$  if it failed at  $x = c$  months. Since  $P\{X = c\} = 0$ , this has to be solved using limiting considerations:

$$\begin{aligned}
 F_X(x | X = c) &= \lim_{\Delta c \rightarrow 0} P\{X \leq x | c < X \leq c + \Delta c\} \\
 &= \lim_{\Delta c \rightarrow 0} \frac{P\{X \leq x, c < X \leq c + \Delta c\}}{P\{c < X \leq c + \Delta c\}} \\
 &= \lim_{\Delta c \rightarrow 0} \frac{P\{c < X \leq c + \Delta c | X \leq x\} P\{X \leq x\} / \Delta c}{P\{c < X \leq c + \Delta c\} / \Delta c} \\
 &= f_X(c | X \leq x) \frac{F_X(x)}{f_X(c)} = \frac{f_X(c, X \leq x)}{F_X(x)} \frac{F_X(x)}{f_X(c)}
 \end{aligned}$$

Hence

$$F_X(x | X = c) = \begin{cases} 1 & x > c \\ 0 & x \leq c \end{cases}$$

and the corresponding density function for the event  $\{X = c\}$  is given by

$$f_X(x | X = c) = \delta(x - c)$$

These distribution and density functions are shown in Fig. 8.2.1 for  $c = 2$ ,  $d = 2$ .

**Example 8.2.2** The joint density function of random variables  $X$  and  $Y$  is

$$f_{XY}(x, y) = \begin{cases} \frac{1}{\pi} & x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The marginal density function  $f_X(x)$  is obtained by integrating  $f_{XY}(x, y)$  over all  $y$ :

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{dy}{\pi} = \frac{2\sqrt{1-x^2}}{\pi}$$

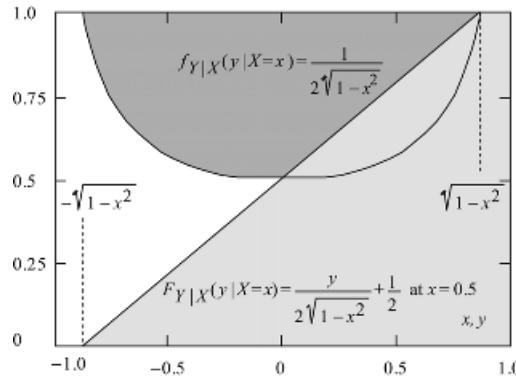


FIGURE 8.2.2

From Eq. (8.2.5), the conditional density  $f_{Y|X}(y | X = x)$  is given by

$$f_{Y|X}(y | X = x) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{1/\pi}{2\sqrt{1-x^2}/\pi} = \begin{cases} \frac{1}{2\sqrt{1-x^2}} & x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Note that for each value of  $x$  the pdf  $f_{Y|X}(y | X = x)$  is a constant and not a function of  $y$ . The corresponding conditional distribution function  $F_{Y|X}(y | X = x)$  is obtained by integrating the pdf from  $-\sqrt{1-x^2}$  to  $y$  as shown below:

$$F_{Y|X}(y | X = x) = \int_{-\sqrt{1-x^2}}^y \frac{1}{2\sqrt{1-x^2}} dy = \frac{y}{2\sqrt{1-x^2}} + \frac{1}{2}$$

In Fig. 8.2.2, the conditional pdf  $f_{Y|X}(y | X = x)$  is shown as a function of  $x$  and the conditional cdf  $F_{Y|X}(y | X = x)$  is shown as a function of  $y$  for  $x = 0.5$ .

**Example 8.2.3** This is an interesting example that again illustrates finding the conditional probability when the probability of the conditioning event is zero. Two friends  $A$  and  $B$  plan to meet between 7:00 and 8:00 p.m. However, they agree that each of them will wait only 15 min for the other. Each of them arrives independently at a random time between 7 and 8 p.m. We have to compute the probability that they will meet given that  $A$  arrives at 7:20 p.m.

Let us first define these random variables.

$$X = \{\text{time of arrival of } A \text{ after 7:00 p.m. in minutes}\}$$

$$Y = \{\text{time of arrival of } B \text{ after 7:00 p.m. in minutes}\}$$

The pdf  $f_X(x)$  and  $f_Y(y)$  are both uniformly distributed in  $(0, 60)$  as shown in Fig. 8.2.3. Since  $X$  and  $Y$  are independent, the joint density  $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$  and is given by  $f_{XY}(x, y) = \frac{1}{60} \times \frac{1}{60} = \frac{1}{3600}$ . The  $60 \times 60$  Cartesian product space is shown in Fig. 8.2.4.

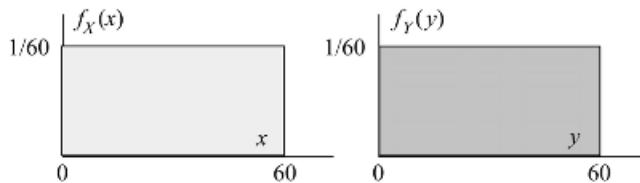


FIGURE 8.2.3

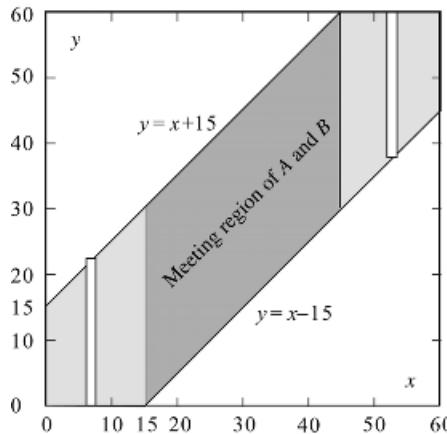


FIGURE 8.2.4

If  $A$  comes at time  $x$  minutes after 7:00 p.m., then for the meeting to take place, we have to find the probability of the event  $|Y - X| \leq 15$  under the condition  $X = x$ . Hence, we can write  $P\{|Y - X| \leq 15 | X = x\} = P\{-15 + X \leq Y \leq 15 + X | X = x\}$ , and we have to find the area under the shaded portion in Fig. 8.2.4. In essence, we have to integrate  $f_Y(y)$  in three different regions: (1)  $0 < y \leq x + 15$ ,  $0 < x = 15$ , (2)  $x - 15 < y \leq x + 15$ ,  $15 < x \leq 45$ , and (3)  $x + 15 < y \leq 60$ ,  $45 < x \leq 60$ . Performing the indicated integration, we obtain

$$P\{A \text{ and } B \text{ meeting} | X = x\} = \begin{cases} \int_0^{15+x} \frac{dx}{60} & 0 < x < 15 \\ \int_{-15+x}^{15+x} \frac{dx}{60} & 15 < x \leq 45 \\ \int_{-15+x}^{60} \frac{dx}{60} & 45 < x \leq 60 \end{cases}$$

$$= \begin{cases} \frac{15+x}{60} & 0 < x \leq 15 \\ \frac{1}{2} & 15 < x \leq 45 \\ \frac{75-x}{60} & 45 < x \leq 60 \end{cases}$$

The probability that  $A$  and  $B$  meet if  $A$  arrives exactly  $x$  minutes after 7:00 p.m. is shown in Fig. 8.2.5.

From the curve, the  $P\{\text{meet} | X = 20\} = 0.5$ . The total probability of  $A$  and  $B$  meeting is given by

$$P\{\text{meet}\} = \int_0^{60} P\{\text{meet} | X = x\} f_X(x) dx$$

$$= \frac{1}{60} \left( \frac{60}{4} + \frac{30}{4} + \frac{15}{4} \right) = \frac{105}{4.60} = \frac{7}{16} = 0.4375$$

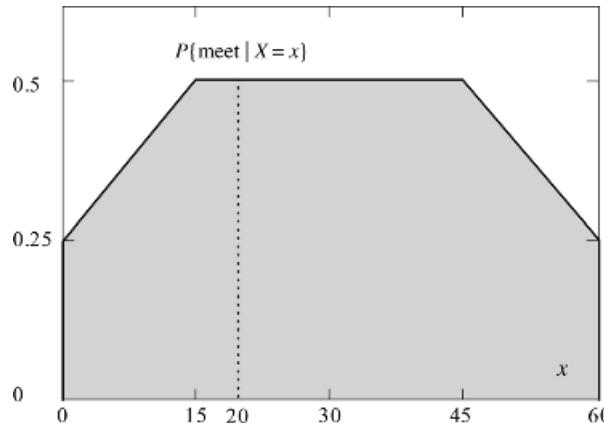


FIGURE 8.2.5

### 8.3 TOTAL PROBABILITY AND BAYES' THEOREM FOR DENSITIES

We will use Eq. (8.2.3), shown in Eq. (8.3.1) as the starting point to derive the total probability law for density functions:

$$P\{B | X = x_1\} = \frac{f_X(x_1 | B)P\{B\}}{f_X(x_1)} \quad (8.3.1)$$

Cross-multiplying and integrating both sides of the resulting equation, we obtain

$$\int_{-\infty}^{\infty} P\{B | X = x_1\}f_X(x_1)dx_1 = \int_{-\infty}^{\infty} f_X(x_1 | B)P\{B\}dx_1 \quad (8.3.2)$$

In Eq. (8.3.2) we obtain the term

$$\int_{-\infty}^{\infty} f_X(x_1 | B)P\{B\}dx_1 = P\{B\} \int_{-\infty}^{\infty} f_X(x_1 | B)dx_1 = P\{B\}$$

and hence Eq. (8.3.2) becomes

$$\int_{-\infty}^{\infty} P\{B | X = x\}f_X(x)dx = P\{B\} \quad (8.3.3)$$

where the subscript on  $x$  has been omitted without loss of generality. Equation (8.3.3) is the total probability law for density functions. We can now formulate Bayes' theorem for density functions. From Eq. (8.3.1) we can write

$$\begin{aligned} f_X(x | B) &= \frac{P\{B | X = x\}f_X(x)}{P\{B\}} \\ &= \frac{P\{B | X = x\}f_X(x)}{\int_{-\infty}^{\infty} P\{B | X = x\}f_X(x)dx} \end{aligned} \quad (8.3.4)$$

where  $f_X(x | B)$  is the a posteriori pdf after observing the event  $B$  given the a priori pdf  $f_X(x)$ . If  $B$  is the event  $B = \{Y = y\}$ , then Eq. (8.3.4) can be written as

$$f_{X|Y}(x | y) = \frac{f_{Y|X}(y | x)f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y | x)f_X(x)dx} \quad (8.3.5)$$

Equation (8.3.5) is Bayes' theorem for connecting the a priori density  $f_X(x)$  with the a posteriori density  $f_{X|Y}(x | y)$  after observing the event  $\{Y = y\}$ .

**Example 8.3.1** Let  $X$  be the random variable that a navigation system fails *before*  $x$  hours. Three suppliers  $A$ ,  $B$ , and  $C$  have supplied the system to airlines with probabilities  $P\{A\} = 0.35$ ,  $P\{B\} = 0.25$ ,  $P\{C\} = 0.4$ . The conditional probabilities of failure for the suppliers  $A$ ,  $B$ ,  $C$  are as follows:

$$F_X(x | A) = (1 - e^{-ax})u(x), \quad a = 1.625 \times 10^{-4}$$

$$F_X(x | B) = (1 - e^{-bx})u(x), \quad b = 5.108 \times 10^{-4}$$

$$F_X(x | C) = (1 - e^{-cx})u(x), \quad c = 2.877 \times 10^{-4}$$

1. From the conditional probabilities of failure, we can find the probability density  $f_X(x)$  as follows:

$$f_X(x | A) = ae^{-ax}u(x), \quad a = 1.625 \times 10^{-4}; \quad P\{A\} = 0.35$$

$$f_X(x | B) = be^{-bx}u(x), \quad b = 5.108 \times 10^{-4}; \quad P\{B\} = 0.25$$

$$f_X(x | C) = ce^{-cx}u(x), \quad c = 2.877 \times 10^{-4}; \quad P\{C\} = 0.40$$

and

$$\begin{aligned} f_X(x) &= f_X(x | A)P\{A\} + f_X(x | B)P\{B\} + f_X(x | C)P\{C\} \\ &= (0.56875e^{-ax} + 1.277e^{-bx} + 1.1508e^{-cx})10^{-4}u(x) \end{aligned}$$

2. If the navigation system fails *after*  $x$  hours, we will find the probabilities  $P\{A | X > x\}$ ,  $P\{B | X > x\}$ ,  $P\{C | X > x\}$ .  $P\{X > x\}$  is given by

$$\begin{aligned} P\{X > x\} &= (1 - F_X(x | A))P\{A\} + (1 - F_X(x | B))P\{B\} + (1 - F_X(x | C))P\{C\} \\ &= e^{-ax}0.35 + e^{-bx}0.25 + e^{-cx}0.40 \end{aligned}$$

The probability  $P\{A | X > x\}$  can be found from the equation

$$P\{A | X > x\} = \frac{P\{X > x | A\}P\{A\}}{P\{X > x\}} = \frac{(1 - F_X(x | A))P\{A\}}{P\{X > x\}}$$

with similar equations for  $P\{B | X > x\}$  and  $P\{C | X > x\}$ . Hence these probabilities can be given by

$$P\{A | X > x\} = \frac{e^{-ax}0.35}{e^{-ax}0.35 + e^{-bx}0.25 + e^{-cx}0.40}$$

$$P\{B | X > x\} = \frac{e^{-bx}0.25}{e^{-ax}0.35 + e^{-bx}0.25 + e^{-cx}0.40}$$

$$P\{C | X > x\} = \frac{e^{-cx}0.40}{e^{-ax}0.35 + e^{-bx}0.25 + e^{-cx}0.40}$$

3. If the navigation system fails *exactly* at time  $X = x$ , we will find  $P\{A | X = x\}$ ,  $P\{B | X = x\}$ , and  $P\{C | X = x\}$ . From Eq. (8.2.3) we have

$$P\{A | X = x\} = \frac{f_X(x | A)P\{A\}}{f_X(x)} = \frac{0.56875 \times 10^{-4}e^{-ax}}{f_X(x)}$$

$$P\{B | X = x\} = \frac{f_X(x | B)P\{B\}}{f_X(x)} = \frac{1.277 \times 10^{-4}e^{-bx}}{f_X(x)}$$

$$P\{C | X = x\} = \frac{f_X(x | C)P\{C\}}{f_X(x)} = \frac{1.1508 \times 10^{-4}e^{-cx}}{f_X(x)}$$

4. We will find the most likely supplier in part (3) if  $x$  ranges from 1000 to 10,000 and the total probability of a wrong decision  $P_e$ .

The most likely supplier if the system failed at exactly  $x$  hours can be obtained by drawing the graphs of the probabilities  $P\{A | X = x\}$ ,  $P\{B | X = x\}$ , and  $P\{C | X = x\}$  as shown in Fig. 8.3.1.

From the graph in Fig. 8.3.1 we can see that if the system fails between  $x = 1000$  and  $x = 2322.22$  h, the most likely supplier is  $C$ , and if it fails between  $x = 2322.22$  and  $x = 5629.17$  h, the most likely supplier is  $A$ . If it fails between  $x = 5629.17$  and  $x = 10,000$  h, the most likely supplier is  $B$ .

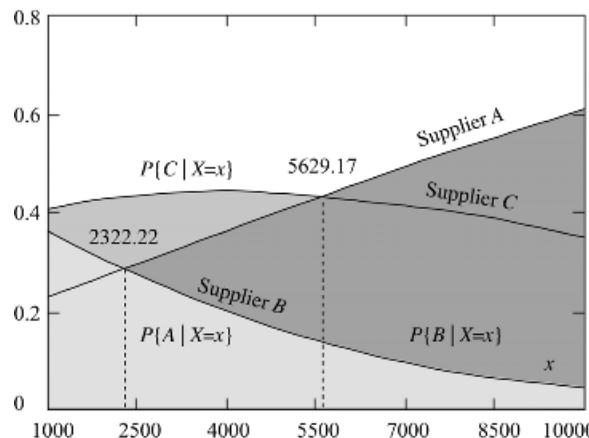


FIGURE 8.3.1

10,000 h, the most likely supplier is A. The conditional probability  $P\{B | X = x\}$  is smaller than the other two conditional probabilities in the two ranges.

The probability of wrong decision  $P_e$  is given by

$$P_e = P\{X \leq 5629.17 | A\}P\{A\} + P\{X > 5629.17 | C\}P\{C\} + P\{B\}$$

Since supplier B does not figure in the wrong decision,  $P\{B\}$  appears without any multiplier. Hence

$$\begin{aligned} P\{\text{wrong decision}\} = P_e &= [1 - e^{-5629.17a}] \cdot 0.35 + [1 - (1 - e^{5629.17c})] \cdot 0.4 + 0.25 \\ &= 0.539 \end{aligned}$$

# Joint Densities and Distributions

## 9.1 JOINT DISCRETE DISTRIBUTION FUNCTIONS

If the random variables  $\{X = X_1, X_2, \dots, X_n\}$ , are discrete, then we can define a joint probability mass function as

$$p_{X_1 \dots X_n}(x_1, \dots, x_n) = P\{X_1 = x_1, \dots, X_n = x_n\} \quad (9.1.1)$$

and the total sum

$$\sum_m^n \dots \sum_j^n \sum_i^n p_{X_1 \dots X_n}(x_i, \dots, x_m) = 1$$

In addition, the marginal distributions are defined by

$$\sum_k^n p_{X_1 \dots X_n}(x_1, \dots, x_k) = p_{X_1 \dots X_{n-1}}(x_1, \dots, x_{n-1}) \quad (9.1.2)$$

**Example 9.1.1 (Discrete)** An urn contains 4 red, 5 green, and 6 blue balls, and 3 balls are taken at random. Given the number of red balls represented by the random variable  $X$  and the number of green balls by the random variable  $Y$ , we want to find the joint probability  $p_{XY}(k,m)$  of  $P\{X = k, Y = m\}$ , where  $k$  ranges from 0 to 4 and  $m$  ranges from 0 to 5. The total number of points in the sample space is

$$C(n,r) = \binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{15!}{3!12!} = 455$$

The number of ways in which 3 balls can be picked from 15 balls is to choose  $k$  balls from 4 red balls,  $m$  balls from 5 green balls, and  $3 - k - m$  balls from 6 blue balls. Hence, the probability  $p_{XY}(k,m)$  is given by

$$p_{XY}(k,m) = \begin{cases} C(6, 3 - k - m)C(4, k)C(5, m) & k + m \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

$k \backslash m$	0	1	2	3	Row Sum $P\{X = k\}$
0	$\frac{20}{455}$	$\frac{75}{455}$	$\frac{60}{455}$	$\frac{10}{455}$	$\frac{165}{455}$
1	$\frac{60}{455}$	$\frac{120}{455}$	$\frac{40}{455}$	$\frac{0}{455}$	$\frac{220}{455}$
2	$\frac{36}{455}$	$\frac{30}{455}$	$\frac{0}{455}$	$\frac{0}{455}$	$\frac{66}{455}$
3	$\frac{4}{455}$	$\frac{0}{455}$	$\frac{0}{455}$	$\frac{0}{455}$	$\frac{4}{455}$
Column Sum $P\{Y = m\}$	$\frac{120}{455}$	$\frac{225}{455}$	$\frac{100}{455}$	$\frac{10}{455}$	$\frac{455}{455}$

FIGURE 9.1.1

These probabilities are shown in Fig. 9.1.1 for all values of  $k$  and  $m$ .

$$P_{XY} = \frac{1}{C(15,3)} \begin{bmatrix} p(0,0) & p(0,1) & p(0,2) & p(0,3) \\ p(1,0) & p(1,1) & p(1,2) & p(1,3) \\ p(2,0) & p(2,1) & p(2,2) & p(2,3) \\ p(3,0) & p(3,1) & p(3,2) & p(3,3) \end{bmatrix}$$

As the probability mass functions  $P\{X = k\}$  and  $P\{Y = m\}$  appear as sums in the margin, they are referred to as *marginal probability* mass functions. Thus, the probability of 2 red balls  $P\{X = 2\} = 66/455$  and the probability of 3 green balls  $P\{Y = 3\} = 10/455$ .

## 9.2 JOINT CONTINUOUS DISTRIBUTION FUNCTIONS

The joint distribution of  $n$  random variables  $\{X_1, X_2, \dots, X_n\}$ , is the probability of the joint event  $\{X_1 \leq x_1, \dots, X_n \leq x_n\}$  and is given by

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = P\{X_1 \leq x_1, \dots, X_n \leq x_n\} \quad (9.2.1)$$

In particular, if  $X$  and  $Y$  are two random variables, the joint distribution function  $F_{XY}(x, y)$  is given by

$$F_{XY}(x, y) = P\{X \leq x, Y \leq y\} \quad (9.2.2)$$

The corresponding density function is given by

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} \quad (9.2.3)$$

Using the definition of the second derivative, Eq. (9.2.3) can also be given in terms of the distribution function as follows:

$$f_{XY}(x, y) = \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{F_{XY}(x + \Delta x, y + \Delta y) + F_{XY}(x, y) - \{F_{XY}(x + \Delta x, y) + F_{XY}(x, y + \Delta y)\}}{\Delta x \Delta y} \quad (9.2.4)$$

Similarly, the distribution function  $F_{XY}(x, y)$  can also be given in terms of the density function by

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(\xi, \eta) d\eta d\xi \quad (9.2.5)$$

The marginal density functions  $f_X(x)$  and  $f_Y(y)$  are obtained by integrating the joint density function  $F_{XY}(x,y)$  over all  $y$  and  $x$ , respectively, giving

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy; \quad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx \quad (9.2.6)$$

### Properties of Joint Densities and Distributions

1.  $F_{XY}(-\infty, -\infty) = 0; F_{XY}(\infty, \infty) = 1$
2.  $F_{XY}(-\infty, y) = 0; F_{XY}(x, -\infty) = 0$
3.  $F_{XY}(x, \infty) = F_X(x); F_{XY}(\infty, y) = F_Y(y)$
4.  $P\{x_1 < X \leq x_2, Y \leq y\} = F_{XY}(x_2, y) - F_{XY}(x_1, y)$   
 $P\{X \leq x, y_1 < Y \leq y_2\} = F_{XY}(x, y_2) - F_{XY}(x, y_1)$  (9.2.7)

5.  $P\{x_1 < X \leq x_2, y_1 < Y \leq y_2\}$   
 $= F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1)$  (9.2.8)

6.  $\frac{\partial F_{XY}(x,y)}{\partial x} = \int_{-\infty}^y f_{XY}(x, \eta)d\eta$   
 $\frac{\partial F_{XY}(x,y)}{\partial y} = \int_{-\infty}^x f_{XY}(\xi, y)d\xi$  (9.2.9)

In many problems we are given  $f_{XY}(x,y)$  in the region  $\{(x_1, x_2], (y_1, y_2]\}$  and the distribution function  $F_{XY}(x,y)$  is desired in the entire  $XY$  plane. In these cases, we can visualize the  $XY$  plane and delineate the various regions as shown in Fig. 9.2.1 for which  $F_{XY}(x,y)$  can be determined.

Given  $F_{XY}(x,y)$  in region I, we find that using the properties enunciated earlier:  $F_{XY}(x,y) = 0$  in region V,  $F_{XY}(x,y) = F_X(x)$  in region II,  $F_{XY}(x,y) = F_Y(y)$  in region III, and  $F_{XY}(x,y) = 1$  in region IV. We shall give three different examples to illustrate how the joint distribution function can be determined.

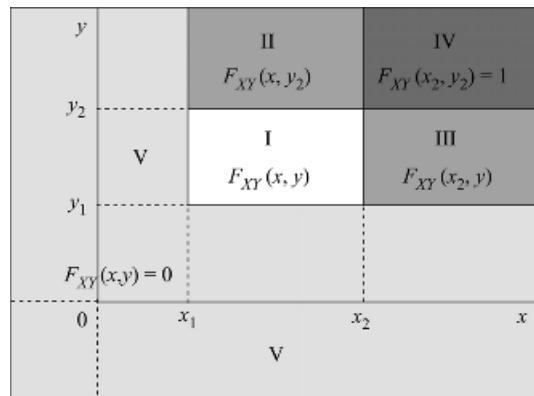


FIGURE 9.2.1

**Example 9.2.1 (Regions Bounded by Constants)** The joint pdf  $f_{XY}(x,y)$  is given by

$$f_{XY}(x,y) = \begin{cases} K(8-x-y) & 1 < x \leq 3, 1 < y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

The constant  $K$  can be found by integrating over the range, yielding

$$\int_1^3 \int_1^2 K(8-x-y) dy dx = 9K = 1 \text{ or } K = \frac{1}{9}$$

We plot the regions in Fig. 9.2.2 as shown in Fig. 9.2.1.

Region V:  $x \leq 1$  or  $y \leq 1$

$$F_{XY}(x,y) = 0$$

Region IV:  $-x > 3, y > 2$

$$F_{XY}(x,y) = 1$$

Region I:  $1 < x \leq 3, 1 < y \leq 2$

$$\begin{aligned} \int_1^x \int_1^y \frac{1}{9}(8-\xi-\eta) d\eta d\xi &= \int_1^x \frac{1}{9} \left( 8\eta - \xi\eta - \frac{\eta^2}{2} \right) \Big|_1^y \\ &= \int_1^x \frac{1}{9} \left( 8y - \xi y - \frac{y^2}{2} + \xi - \frac{15}{2} \right) d\xi \\ &= \frac{1}{9} \left( 8y\xi - \frac{\xi^2 y}{2} - \frac{\xi y^2}{2} + \frac{\xi^2}{2} - \frac{15\xi}{2} \right) \Big|_1^x \\ &= \frac{1}{9} \left( 8y\xi - \frac{\xi^2 y}{2} - \frac{\xi y^2}{2} + \frac{\xi^2}{2} - \frac{15\xi}{2} \right) \\ &= \frac{1}{9} \left( 8xy - \frac{x^2 y}{2} - \frac{xy^2}{2} + \frac{x^2}{2} + \frac{y^2}{2} - \frac{15x}{2} - \frac{15y}{2} + 7 \right) \end{aligned}$$

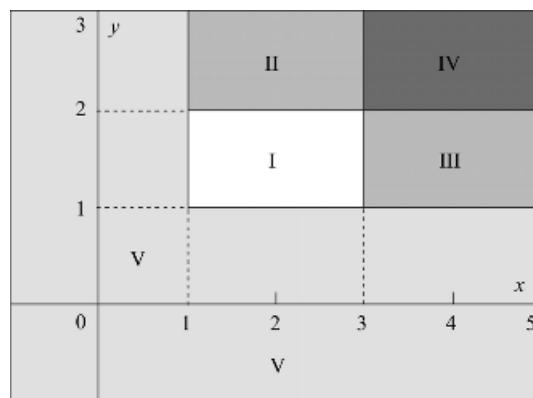


FIGURE 9.2.2

Region II:  $1 < x \leq 3, y > 2$  [marginal distribution  $F_X(x)$ ]

$$F_{XY}(x,2) = F_X(x) = \frac{1}{9} \left( \frac{-x^2}{2} + \frac{13x}{2} - 6 \right)$$

Region III:  $x > 3, 1 < y \leq 2$  [marginal distribution  $F_Y(y)$ ]

$$F_{XY}(3,y) = F_Y(y) = \frac{1}{9} (-y^2 + 12y - 11)$$

By differentiating the distribution function, we can write the density function for the various regions as follows:

$$f_{XY}(x,y) = \begin{cases} 0 & x \leq 1 \text{ or } y \leq 1 \\ \frac{1}{9}(8-x-y) & 1 < x \leq 3, 1 \leq y \leq 2 \\ \frac{1}{9}(12-2y) & x > 3, 1 \leq y \leq 2 \\ \frac{1}{9}\left(\frac{13}{2}-x\right) & 1 < x \leq 3, y > 2 \\ 0 & x > 3, y > 2 \end{cases}$$

We can also conclude that since  $f_{XY}(x,y) \neq f_X(x)f_Y(y)$ , the random variables  $X$  and  $Y$  are not independent. Having determined the joint probability  $F_{XY}(x,y)$ , we can find the probability for other regions such as  $\{1.5 < X \leq 2.5, 1 < Y \leq 1.5\}$  as shown in Fig. 9.2.3.

$$\begin{aligned} P\{1.5 < X \leq 2.5, 1 < Y \leq 1.5\} &= \int_{1.5}^{2.5} \int_1^{1.5} \frac{1}{9}(8-x-y) dy dx \\ &= F_{XY}(2.5, 1.5) - F_{XY}(1.5, 1.5) \\ &\quad - F_{XY}(2.5, 1) + F_{XY}(1.5, 1) \\ &= 0.4167 - 0.1528 - 0 - 0 = 0.2639 \end{aligned}$$

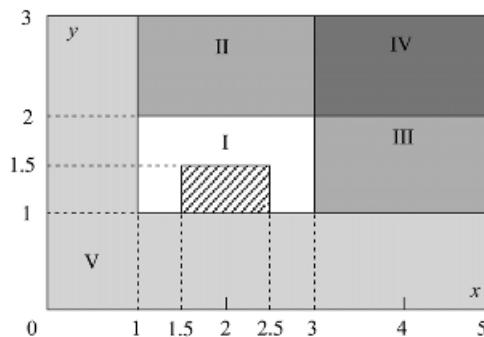


FIGURE 9.2.3

We can also find the conditional probability of  $Y \leq y$  given that  $X \leq x$  as follows

$$\begin{aligned} F_{Y|X}(y|X \leq x) &= \frac{F_{XY}(x,y)}{F_X(x)} \\ &= \frac{\frac{1}{9} \left( 8xy - \frac{x^2y}{2} - \frac{xy^2}{2} + \frac{x^2}{2} + \frac{y^2}{2} - \frac{15x}{2} - \frac{15y}{2} + 7 \right)}{\frac{1}{9} \left( \frac{-x^2}{2} + \frac{13x}{2} - 6 \right)} \\ &= \frac{y^2 + xy - x - 15y + 14}{x - 12} \end{aligned}$$

At the endpoints  $x = 3$  and  $y = 2$ , we have  $F_{Y|X}(y|X \leq 3) = F_Y(y)$  and  $F_{Y|X}(2|X \leq 3) = F_Y(2) = 1$ . The corresponding conditional density function  $f_{Y|X}(y|X \leq x)$  is given by

$$f_{Y|X}(y|X \leq x) = \frac{x + 2y - 15}{x - 12}, \quad 1 < x \leq 3, \quad 1 < y \leq 2$$

We will now find the conditional density  $f_{Y|X}(y|X = x)$  from Eq. (8.2.5)

$$f_{Y|X}(y|X = x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{\frac{1}{9}(8-x-y)}{\frac{1}{9} \left( \frac{13}{2} - x \right)} = \frac{2(8-x-y)}{13-2x}$$

in the region  $1 < x \leq 3$ ,  $1 < y \leq 2$ . The corresponding distribution function  $F_{Y|X}(y|X = x)$  in the region  $1 < x \leq 3$ ,  $1 < y \leq 2$  is obtained by integrating the density function and is given by

$$\begin{aligned} F_{Y|X}(y|X = x) &= \int_1^y \frac{2(8-x-\eta)}{13-2x} d\eta \\ &= \frac{y^2 + 2xy - 2x - 16y + 15}{2x - 13} \end{aligned}$$

We see clearly from this example that  $F_{Y|X}(y|X \leq x)$  is different from  $F_{Y|X}(y|X = x)$  and the corresponding conditional density functions  $f_{Y|X}(y|X \leq x)$  and  $f_{Y|X}(y|X = x)$  are also different. Note that  $f_{Y|X}(y|X = x)$  has been defined as  $f_{Y|X}(y|x)$  in Chapter 8.

**Example 9.2.2 (One Region Bounded by a Function)** In this example one of the bounds on the region is a functional inequality. The joint pdf  $f_{XY}(x,y)$  of two random variables  $X$  and  $Y$  is given by

$$f_{XY}(x,y) = \begin{cases} \frac{8}{7}(1+3xy) & 0 < x \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We have to find the joint distribution function  $F_{XY}(x,y)$  in the entire  $xy$  plane. Integrating first along the  $\eta$  axis and then along the  $\xi$  axis (Fig. 9.2.4), we have

$$\begin{aligned} F_{XY}(x,y) &= \int_0^x \int_\xi^y \frac{8}{7}(1+3\xi\eta) d\eta d\xi \\ &= \frac{1}{7}(-3x^4 + 6x^2y^2 - 4x^2 + 8xy), \quad 0 < x \leq y \leq 1 \end{aligned}$$

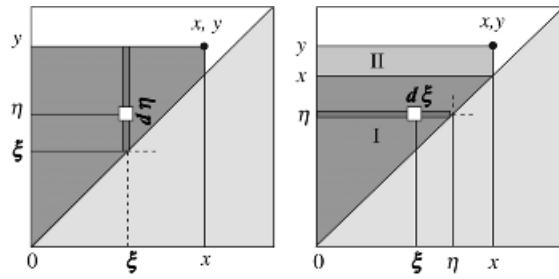


FIGURE 9.2.4

Integrating along the  $\xi$  axis first and then along the  $\eta$  axis (Fig. 9.2.4), we have

$$\begin{aligned}
 F_{XY}(x,y) &= \underbrace{\int_0^x \int_0^{\eta} \frac{8}{7}(1+3\xi\eta) d\xi d\eta}_I \\
 &\quad + \underbrace{\int_x^y \int_0^x \frac{8}{7}(1+3\xi\eta) d\xi d\eta}_II \\
 &= \frac{1}{7}(3x^4 + 4x^2) + \frac{1}{7}(-6x^4 + 6x^2y^2 - 8x^2 + 8xy) \\
 &= \frac{1}{7}(-3x^4 + 6x^2y^2 - 4x^2 + 8xy), \quad 0 < x \leq y \leq 1
 \end{aligned}$$

**Example 9.2.3 (Both Regions Bounded by Functions)** The joint probability density function of two random variables  $X$  and  $Y$  is given by

$$f_{XY}(x,y) = \begin{cases} k(1+2xy) & \begin{cases} 0 < x \leq 1 \\ 0 < x^2 < y \leq x + \frac{1}{2} \leq 1 \end{cases} \\ 0 & \text{otherwise} \end{cases}$$

The two boundaries are shown in Fig. 9.2.5.

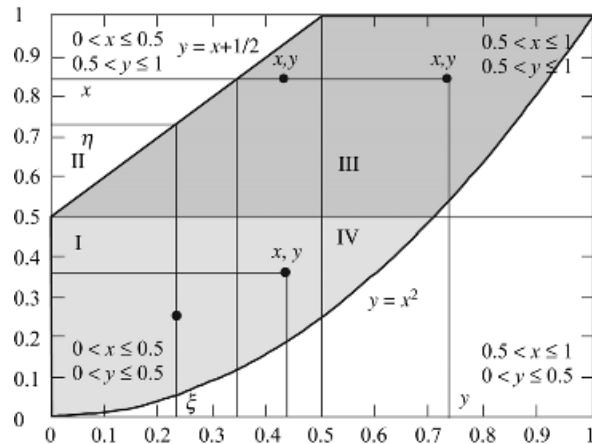


FIGURE 9.2.5

The constant  $k$  is found by integrating  $f_{XY}(x,y)$  over the ranges of definition as shown below:

$$\int_0^{1/2} \int_{x^2}^{x+1/2} k(1+2xy)dy dx + \int_{1/2}^1 \int_{x^2}^1 k(1+2xy)dy dx = 1$$

Solving the above equation we obtain,

$$k \frac{161}{192} = 1 \quad \text{or} \quad k = \frac{192}{161}$$

*Method 1: Marginal Distributions by Integrating  $f_{XY}(x,y)$ .* We will now find the marginal density  $f_X(x)$  by integrating  $f_{XY}(x,y)$  over all  $y$ . In the region  $0 < x \leq \frac{1}{2}$ , we have

$$\begin{aligned} f_X(x) &= \int_{x^2}^{x+1/2} \frac{192}{161} (1+2xy)dy \\ &= \frac{48}{161} (2+5x+4x^3-4x^5), \quad 0 < x \leq \frac{1}{2} \end{aligned}$$

In the region  $\frac{1}{2} < x \leq 1$ , we have

$$\begin{aligned} f_X(x) &= \int_{x^2}^1 \frac{192}{161} (1+2xy)dy \\ &= \frac{192}{161} (1+x-x^2-x^5), \quad \frac{1}{2} < x \leq 1 \end{aligned}$$

By integrating  $f_X(x)$  in these two regions, we can write the distribution function  $F_X(x)$  as

$$F_X(x) = \begin{cases} \frac{8}{161} (12x+15x^2+16x^4-4x^6) & 0 < x \leq \frac{1}{2} \\ \frac{-31}{161} + \frac{32}{161} (6x+3x^2-2x^3-x^6) & \frac{1}{2} < x \leq 1 \\ 1 & x > 1 \\ 0 & \text{otherwise} \end{cases}$$

Similarly, we integrate  $f_{XY}(x,y)$  over all  $x$  to obtain the marginal density  $f_Y(y)$  as shown below:

$$f_Y(y) = \begin{cases} \int_0^{\sqrt{y}} \frac{192}{161} (1+2xy)dx, & 0 < y \leq \frac{1}{2} \\ \int_{y-1/2}^{\sqrt{y}} \frac{192}{161} (1+2xy)dx, & \frac{1}{2} < y \leq 1 \end{cases}$$

or

$$f_Y(y) = \begin{cases} \frac{192}{161} (\sqrt{y}-y^2) & 0 < y \leq \frac{1}{2} \\ \frac{16}{161} (6+12\sqrt{y}-15y+24y^2-12y^3) & \frac{1}{2} < y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Integrating  $f_Y(y)$  in the above regions we obtain the distribution function  $F_Y(y)$  as shown below:

$$F_Y(y) = \begin{cases} \frac{64}{161}(2y^{3/2} + y^3) & 0 < y \leq \frac{1}{2} \\ \frac{-1}{7} + \frac{16}{161}(6y + 8y^{3/2} - 15y^2 + 8y^3 - 3y^4) & \frac{1}{2} < y \leq 1 \\ 1 & y > 1 \\ 0 & \text{otherwise} \end{cases}$$

$f_X(x)$ ,  $F_X(x)$ ,  $f_Y(y)$  and  $F_Y(y)$  are shown in Fig. 9.2.6.

*Method 2: Marginal Distributions from  $F_{XY}(x,y)$ .* We will now find the marginal distributions from the joint distribution  $F_{XY}(x,y)$  obtained by integrating the joint density  $f_{XY}(x,y)$  over the regions of definition.

The joint distribution  $F_{XY}(x,y)$  in the region  $0 < x \leq \frac{1}{2}$ ,  $0 < y \leq \frac{1}{2}$  is obtained from the following integration:

$$\begin{aligned} F_{XY}(x,y) &= \int_0^x \int_{x^2}^y \frac{192}{161}(1 + 2\xi\eta)d\eta d\xi \\ &= \frac{32}{161}(6xy - 2x^3 + 3x^2y^2 - x^6), \quad \begin{cases} 0 < x \leq 1 \\ 0 < y \leq 1/2 \end{cases} \end{aligned}$$

and in the region  $0 < x \leq 1$ ,  $\frac{1}{2} < y \leq 1$ , we have

$$\begin{aligned} F_{XY}(x,y) &= \int_0^{y-(1/2)} \int_{\xi^2}^{\xi+(1/2)} \frac{192}{161}(1 + 2\xi\eta)d\eta d\xi \\ &\quad + \int_{y-(1/2)}^x \int_{\xi^2}^y \frac{192}{161}(1 + 2\xi\eta)d\eta d\xi \end{aligned}$$

or

$$\begin{aligned} F_{XY}(x,y) &= \frac{-1}{7} + \frac{8}{161}(12y - 15y^2 + 8y^3 - 6y^4 \\ &\quad + 24xy + 12x^2y^2 - 8x^3 - 4x^6) \quad \begin{cases} 0 < x < 1 \\ \frac{1}{2} < y \leq 1 \end{cases} \end{aligned}$$

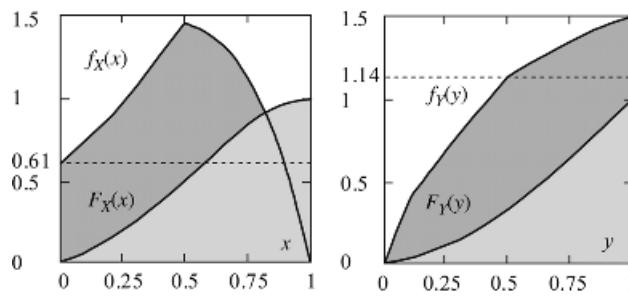


FIGURE 9.2.6

The marginal distributions  $F_X(x)$  and  $F_Y(y)$  can be obtained by substituting suitable boundary values in the joint distribution  $F_{XY}(x,y)$  as shown below:

$$F_X(x) = \begin{cases} F_{XY}\left(x, x + \frac{1}{2}\right), & 0 < x \leq \frac{1}{2}, y > \frac{1}{2} \\ F_{XY}(x, 1), & \frac{1}{2} < x \leq 1, y > \frac{1}{2} \\ F_{XY}(1, 1), & x > 1, y > \frac{1}{2} \\ F_{XY}(0, 0), & x \leq 0, y > \frac{1}{2} \end{cases}$$

Similarly

$$F_Y(y) = \begin{cases} F_{XY}(\sqrt{y}, y), & 0 < y \leq \frac{1}{2}, 0 < x \leq 1 \\ F_{XY}(\sqrt{y}, y), & \frac{1}{2} < y \leq 1, 0 < x \leq 1 \\ F_{XY}(1, 1), & y > 1, x > 1 \\ F_{XY}(0, 0), & y \leq 0, x > 1 \end{cases}$$

and these are exactly the same as derived previously.

### 9.3 BIVARIATE GAUSSIAN DISTRIBUTIONS

If  $X$  and  $Y$  are jointly distributed random variables with mean value  $\mu_X$  and  $\mu_Y$  and variances  $\sigma_X^2$  and  $\sigma_Y^2$ , then the joint density function  $f_{XY}(x,y)$  is given by

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{x-\mu_X}{\sigma_X} \right)^2 - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left( \frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\} \quad (9.3.1)$$

where  $\rho$  is the correlation coefficient that determines the degree of similarity between  $X$  and  $Y$ , given by

$$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} \quad (9.3.2)$$

We have to show that

$$I = \iint_{-\infty}^{\infty} F_{XY}(x,y) dy dx = 1 \quad (9.3.3)$$

Substituting  $\xi = (x - \mu_X)/\sigma_X$  and  $\eta = (y - \mu_Y)/\sigma_Y$  in Eq. (9.3.1), we can write the integral  $I$  as follows:

$$I = \iint_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ \frac{-1}{2(1-\rho^2)} [\xi^2 - 2\rho\xi\eta + \eta^2] \right\} d\eta d\xi \quad (9.3.4)$$

Completing the squares in  $\xi$  in the integral (9.3.4) above, we have

$$I = \iint_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\left[\frac{(\xi-\rho\eta)^2}{1-\rho^2} + \eta\right]\right\} d\eta d\xi \quad (9.3.5)$$

Substituting

$$w = \frac{\xi - \rho\eta}{\sqrt{1-\rho^2}}$$

and

$$dw = \frac{d\xi}{\sqrt{1-\rho^2}}$$

the integral (9.3.5) can be written as follows:

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{w^2}{2}\right\} dw \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{-\frac{\eta^2}{2}\right\} d\eta = 1 \quad (9.3.6)$$

If  $\rho = 0$  in Eq. (9.3.3), then  $X$  and  $Y$  are independent because  $f_{XY}(x,y) = f_X(x)f_Y(y)$ . Thus, if Gaussian-distributed random variables  $X$  and  $Y$  are uncorrelated, they are also independent.

We shall find the conditional pdf  $f_{Y|X}(Y|X=x)$  given  $X=x$ . From Eq. (8.2.5) we have

$$\begin{aligned} f_{Y|X}(y|X=x) &= \frac{f_{XY}(x,y)}{f_X(x)} \\ &= \left\{ \frac{\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{(2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}}{\sqrt{2\pi\sigma_X} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right]} \right\} \\ &= \frac{1}{\sqrt{2\pi\sigma_Y\sqrt{1-\rho^2}}} \exp\left\{-\frac{1}{2\sigma_Y^2(1-\rho^2)} \left[ \left(y - \mu_Y - \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X)\right)^2 \right]\right\} \end{aligned}$$

Thus the conditional density has a conditional mean value of

$$\mu_{Y|X} = \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X)$$

and a conditional variance of  $\sigma_{Y|X}^2 = \sigma_Y^2(1 - \rho^2)$ .

# Moments and Conditional Moments

## 10.1 EXPECTATIONS

The expected value or the mean value of any random variable  $X$  is defined by

$$E[X] = \mu_X = \int_{-\infty}^{\infty} x f_X(x) dx \quad (10.1.1)$$

This definition is valid regardless of whether the random variable is continuous or discrete. Since the pdf of a discrete random variable is always given by a sum of delta functions, we have

$$\begin{aligned} E[X] = \mu_X &= \int_{-\infty}^{\infty} x \left\{ \sum_{i=0}^{\infty} P\{X = x_i\} \delta(x - x_i) \right\} dx \\ &= \sum_{i=0}^{\infty} x_i P\{X = x_i\} \end{aligned} \quad (10.1.2)$$

The expected value is the first moment of the random variable  $X$  about the origin.

**Example 10.1.1** For a random variable uniformly distributed in the interval  $(a, b]$ , we have

$$f_X(x) = \frac{1}{b - a}$$

and hence the mean value is given by

$$E[X] = \frac{1}{b - a} \int_a^b x dx = \frac{1}{b - a} \frac{x^2}{2} \Big|_a^b = \frac{1}{2} \frac{b^2 - a^2}{b - a} = \frac{b + a}{2}$$

**Example 10.1.2** On a toss of a die, a person wins \$10 if {1} or {3} results, loses \$5 if {4} or {6} results, wins \$5 if {2} results, and loses \$10 if {5} results. From Eq. (10.1.2),

the expected value of the winnings is

$$\begin{aligned} E\{X\} &= \mu_X = \sum_{i=1}^6 x_i P\{X = x_i\} \\ &= \frac{1}{6} \{10 + 10 - 5 - 5 + 5 - 10\} = \frac{5}{6} \end{aligned}$$

The expected value of any random variable is a constant. In other words, it smooths the variations of a random variable to a constant. It is a one-way operator, in the sense that once the random variable  $X$  is smoothed, there is no way of getting back the density function. Thus, the expected value of a constant  $a$  is the same constant  $a$ .

### Properties of Expectations

1. It is a linear operator irrespective of the nature of random variables, in the sense that the expected value of the sum equals the sum of expected values

$$E[X + Y] = E[X] + E[Y] \quad (10.1.3)$$

as shown below:

$$\begin{aligned} E[X + Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f_{XY}(xy) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf_{XY}(xy) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf_{XY}(xy) dx dy \\ &= \int_{-\infty}^{\infty} xf_X(x) dx + \int_{-\infty}^{\infty} yf_Y(y) dy = E\{X\} + E\{Y\} \end{aligned}$$

2. On the contrary, the expected value of the product given by

$$E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(xy) dx dy \quad (10.1.4)$$

is not, in general, equal to the product of expected values. However, Equation (10.1.4) can be split only if  $X$  and  $Y$  are independent, in which case we have

$$\begin{aligned} E[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(xy) dx dy \\ &= \int_{-\infty}^{\infty} xf_X(x) dx \int_{-\infty}^{\infty} yf_Y(y) dy \\ &= E[X] E[Y] \end{aligned}$$

and the expected value of the product is equal to the product of expected values. The converse is not true. If  $E[XY] = E[X] E[Y]$ , then  $X$  and  $Y$  are uncorrelated and need not be independent. Analogously, if  $E[XY] = 0$ , then  $X$  and  $Y$  are orthogonal.

We can summarize the following definitions:

*Uncorrelated:*  $E[XY] = E[X] E[Y]$

*Orthogonal:*  $E[XY] = 0$

*Independence:*  $f_{XY}(xy) = f_X(x)f_Y(y)$

**Conditional Expectation**

Given that the random variable  $X$  has occurred, the expected value of  $Y$  is defined by

$$E[Y | X] = \int_{-\infty}^{\infty} y f_{Y|X}(y | x) dy \quad (10.1.5)$$

The conditional expectation operator does not completely smooth the random variable  $Y$  to a constant but smooths it to random variable dependent on  $X$ . Thus, if we take the expectation of the conditional expectation, we get the expected value of  $Y$ :

$$\begin{aligned} E\{E[Y | X]\} &= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} y f_{Y|X}(y | x) dy \right] f_X(x) dx \\ &= \int_{-\infty}^{\infty} y f_{XY}(xy) dy dx \\ &= \int_{-\infty}^{\infty} y f_Y(y) dy = E[Y] \end{aligned} \quad (10.1.6)$$

If  $X$  and  $Y$  are independent, then  $E[Y | X] = E[Y]$  since  $f_{Y|X}(y | x) = f_Y(y)$ .

**Example 10.1.3** A random variable  $Y$  is uniformly distributed in the interval  $(0,5)$ . Hence the distribution function  $F_Y(y)$  is given by  $y/5$  for  $0 < y \leq 5$ . We will find the conditional expectation of  $Y$  given that  $Y > x$  and  $0 < x \leq 5$ . We first find the conditional distribution  $F_Y(y | Y > x)$ :

$$F_Y(y | Y > x) = \frac{P(x < Y \leq y)}{P(Y > x)} = \frac{F_Y(y) - F_Y(x)}{1 - F_Y(x)}$$

Differentiating this equation with respect to  $y$ , the conditional density  $f_Y(y | Y > x)$  is obtained as follows:

$$f_Y(y) = \frac{f_Y(y)}{1 - F_Y(x)} = \frac{1}{5 - x}$$

Hence the conditional expectation  $E[Y | Y > x]$  is given by

$$\begin{aligned} E[Y | Y > x] &= \int_x^5 \frac{y}{5-x} dy = \frac{y^2}{2(5-x)} \Big|_x^5 \\ &= \frac{25 - x^2}{2(5-x)} = \frac{5+x}{2} \end{aligned}$$

**Joint Moment**

We can now define the joint moment between two random variables  $X$  and  $Y$  as follows:

$$E[XY] = \mu_{XY} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x,y) dx dy \quad (10.1.7)$$

The joint moments determine the correlation between  $X$  and  $Y$ .

## 10.2 VARIANCE

The *variance* is the second moment about the mean value also known as the *second central moment*. It is defined by

$$\begin{aligned}\text{var}[X] &= E[X - \mu_X]^2 = \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx \\ &= E[X^2] - 2\mu_X E[X] + \mu_X^2 = E[X^2] - \mu_X^2\end{aligned}\quad (10.2.1)$$

The variance expresses the variability of the random variable about the mean value. If this variability is zero, then the random variable will be no longer random but a constant. This will be discussed further in Chapter 14.

The preceding definition applies for any random variable. However, for a discrete random variable, analogous to Eq. (10.1.2), we have

$$E[X - \mu_X]^2 = \sigma_X^2 = \sum_{i=0}^{\infty} (x_i - \mu_X)^2 P\{X = x_i\} \quad (10.2.2)$$

**Example 10.2.1** For a random variable,  $X$  uniformly distributed between  $(-a, a]$ , the mean value  $\mu_X = 0$ . Hence the variance is given by

$$\sigma_X^2 = E[X^2] = \frac{1}{2a} \int_{-a}^a x^2 dx = \frac{a^2}{3}$$

**Example 10.2.2** From Eq. (10.2.2), the variance of the random variable given in Example 10.1.2 is

$$\begin{aligned}\sigma_X^2 &= \sum_{i=1}^6 \left( x_i - \frac{5}{6} \right)^2 \frac{1}{6} \\ &= \frac{1}{6} \left( \frac{55^2}{6^2} + \frac{55^2}{6^2} + \frac{25^2}{6^2} + \frac{25^2}{6^2} + \frac{25^2}{6^2} + \frac{55^2}{6^2} \right) = \frac{10,950}{216}\end{aligned}$$

### Properties of Variance

$$1. \quad \text{var}[aX] = a^2 \text{ var}[X] = a^2 \sigma_X^2 \quad (10.2.3)$$

$$\begin{aligned}2. \quad \text{var}[X \pm Y] &= E[X \pm Y]^2 - \{E[X] + E[Y]\}^2 \\ &= E[X^2] - \{E[X]\}^2 + E[Y^2] - \{E[Y]\}^2 \\ &\quad \pm 2\{E[XY] - E[X]E[Y]\} \\ &= \sigma_X^2 + \sigma_Y^2 \pm 2\{E[XY] - E[X]E[Y]\}\end{aligned}\quad (10.2.4)$$

3. If  $X$  and  $Y$  are independent, then

$$\text{var}[X \pm Y] = \sigma_X^2 + \sigma_Y^2 \quad (10.2.5)$$

*Conditional Variance.* We can also define a conditional variance of  $Y$  given  $X$  has occurred as

$$\text{var}[Y | X] = E[(Y - \mu_{Y|X})^2 | X] \quad (10.2.6)$$

*Covariance.* The covariance between two random variables  $X$  and  $Y$  is defined by

$$\begin{aligned}\text{cov}[XY] &= \sigma_{XY} = E[X - \mu_X]E[Y - \mu_Y] \\ &= E[XY] - \mu_X\mu_Y\end{aligned}\quad (10.2.7)$$

If  $X$  and  $Y$  are independent, then  $\text{cov}[XY] = 0$ . The converse is not true. If  $\text{cov}[XY] = 0$ , then  $X$  and  $Y$  are uncorrelated but not necessarily independent.

*Correlation Coefficient.* The correlation coefficient is the normalized covariance and is defined by

$$\rho_{XY} = \frac{\text{cov}[XY]}{\sqrt{\text{var}[X]\text{var}[Y]}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}\quad (10.2.8)$$

The correlation coefficient satisfies  $-1 \leq \rho_{XY} \leq 1$  and shows the degree of self-similarity between  $X$  and  $Y$ .

## 10.3 MEANS AND VARIANCES OF SOME DISTRIBUTIONS

### *Discrete Random Variables*

#### *Binomial Distribution*

*Mean:*

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k}\quad (10.3.1)$$

$$\begin{aligned}E[X] &= \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k q^{n-k} \\ &= \sum_{k=0}^n \frac{n(n-1)!}{(k-1)!(n-k)!} p^k p^{k-1} q^{n-k} \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} \\ &= np(p+q)^{n-1} = np\end{aligned}\quad (10.3.2)$$

*Variance:*

$$\text{var}[X] = E[X^2] - \mu_X^2\quad (10.3.3)$$

$$E[X^2] = \sum_{k=0}^n k^2 \frac{n!}{k!(n-k)!} p^k q^{n-k}\quad (10.3.4)$$

Substituting  $k^2 = k(k-1) + k$  in Eq. (10.3.4), we have

$$E[X^2] = \sum_{k=0}^n k(k-1) \frac{n!}{k!(n-k)!} p^k q^{n-k} + \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k q^{n-k}\quad (10.3.5)$$

The second summation in Eq. (10.3.5) is the mean value, already calculated as  $np$  in Eq. (10.3.2). The first summation can be written as follows:

$$\begin{aligned} \sum_{k=0}^n k(k-1) \frac{n!}{k!(n-k)!} p^k q^{n-k} &= n(n-1) p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} q^{n-k} \\ &= n(n-1) p^2 (p+q)^{n-2} = n(n-1) p^2 \end{aligned} \quad (10.3.6)$$

Thus

$$E[X^2] = n(n-1)p^2 + np \quad (10.3.7)$$

and the variance from Eq. (10.3.3) is given by

$$\text{var}[X] = n(n-1)p^2 + np - n^2p^2 = npq \quad (10.3.8)$$

### Poisson Distribution

*Mean:*

$$P(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (10.3.9)$$

$$E[X] = e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda \quad (10.3.10)$$

*Variance.* We use the same technique as in the binomial distribution by substituting  $k^2 = k(k-1) + k$ . Hence

$$\begin{aligned} E[X^2] &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{k(k-1)\lambda^k}{k!} + e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} \\ &= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \\ &= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda = \lambda^2 + \lambda \end{aligned} \quad (10.3.11)$$

Substituting in  $\text{var}[X] = E[X^2] - \mu_X^2$  from Eqs. (10.3.11) and (10.3.10), we have

$$\text{var}[X] = \lambda^2 + \lambda - \lambda = \lambda \quad (10.3.12)$$

For a Poisson distribution, the mean and the variance are the same and equal to  $\lambda$ .

### Continuous Random Variables

#### Exponential Distribution

*Mean:*

$$f_X(x) = \lambda e^{-\lambda x} u(x) \quad (10.3.13)$$

$$E[X] = \lambda \int_0^{\infty} x e^{-\lambda x} dx \quad (10.3.14)$$

Integrating Eq. (10.3.14) by parts, we obtain

$$\begin{aligned} E[X] &= \lambda \int_0^\infty x \frac{de^{-\lambda x}}{-\lambda} = - \int_0^\infty x \, de^{-\lambda x} \\ &= -xe^{-\lambda x} \Big|_0^\infty + \int_0^\infty de^{-\lambda x} = \frac{1}{\lambda} \end{aligned} \quad (10.3.15)$$

*Variance.* We first calculate the second moment:

$$E[X^2] = \lambda \int_0^\infty x^2 e^{-\lambda x} dx \quad (10.3.16)$$

Integrating by parts [Eq. (10.3.16)] twice, we get the result as

$$\lambda \int_0^\infty x^2 e^{-\lambda x} dx = \left[ \frac{e^{-\lambda x}}{\lambda^2} (-\lambda^2 x^2 + 2\lambda x + 2) \right]_0^\infty = \frac{2}{\lambda^2} \quad (10.3.17)$$

and the variance is given by

$$\text{var}[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \quad (10.3.18)$$

### *Gaussian Distribution*

*Mean:*

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \frac{(x - \mu_X)^2}{\sigma_X^2} \right] \quad (10.3.19)$$

$$\begin{aligned} E[X] &= \frac{1}{\sigma_X \sqrt{2\pi}} \int_{-\infty}^\infty x \exp \left[ -\frac{1}{2} \frac{(x - \mu_X)^2}{\sigma_X^2} \right] dx \\ &= \frac{1}{\sigma_X \sqrt{2\pi}} \int_{-\infty}^\infty (x - \mu_X) \exp \left[ -\frac{1}{2} \frac{(x - \mu_X)^2}{\sigma_X^2} \right] dx \\ &\quad + \frac{\mu_X}{\sigma_X \sqrt{2\pi}} \int_{-\infty}^\infty \exp \left[ -\frac{1}{2} \frac{(x - \mu_X)^2}{\sigma_X^2} \right] dx \end{aligned} \quad (10.3.20)$$

In Eq. (10.3.20), the first integral is zero because it is an odd function and the second integral equals  $\mu_X$ . Hence

$$E[X] = \mu_X \quad (10.3.21)$$

*Variance:*

$$\text{var}[X] = \frac{1}{\sigma_X \sqrt{2\pi}} \int_{-\infty}^\infty (x - \mu_X)^2 \exp \left[ -\frac{1}{2} \frac{(x - \mu_X)^2}{\sigma_X^2} \right] dx \quad (10.3.22)$$

Substituting  $y = (x - \mu_X)/\sigma_X$  in Eq. (10.3.22), we have

$$\text{var}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \sigma_X^2 y^2 e^{-(1/2)y^2} dy = \sigma_X^2 \quad (10.3.23)$$

where it can be shown that  $1/\sqrt{2\pi} \int_{-\infty}^\infty y^2 e^{-(1/2)y^2} dy = 1$

## 10.4 HIGHER-ORDER MOMENTS

We can now define higher-order moments. The  $k$ th-order moment is defined by

$$E[X^k] = m_k = \int_{-\infty}^{\infty} x^k f_X(x) dx \quad (10.4.1)$$

Clearly,  $m_0 = 1$  and  $m_1 = \mu_X$ . Similarly, the  $k$ th-order central moments are defined by

$$E[X - \mu_X^k] = v_k = \int_{-\infty}^{\infty} (x - \mu_X^k) f_X(x) dx \quad (10.4.2)$$

Here  $v_0 = 1$ ,  $v_1 = 0$ , and  $v_2 = \sigma_X^2$ .

We can find relationships between the  $k$ th-order moments and the  $k$ th-order central moments. We can write these relationships as follows:

$$\begin{aligned} v_0 &= 1 & m_0 &= 1 \\ v_1 &= 0 & m_1 &= \mu \\ v_2 &= \sigma^2 & m_2 &= v_2 + \mu^2 \\ v_3 &= m_3 - 3m_1 m_2 + 2m_1^3 & m_3 &= v_3 + 3m_1 v_2 + m_1^3 \\ \dots & & \dots & \end{aligned} \quad (10.4.3)$$

The moments of random variables are not arbitrary numbers but must satisfy certain inequalities. For example, since  $E[X - \mu_X]^2$  must be nonnegative, we have

$$\sigma^2 = m_2 - m_1^2 \geq 0 \quad \text{and} \quad m_{2k} - m_k^2 \geq 0 \quad \text{for all } k \quad (10.4.4)$$

A random variable is completely determined if the density function is known. However, knowledge of only two moments gives only a limited information of a random variable. But under certain conditions, knowledge of moments  $m_k$  for all  $k$  will uniquely determine the random variable. However, because only two of the moments of a Gaussian random variable are independent, it can be completely determined with two moments. All other moments are dependent on two moments. We will show that for a Gaussian random variable with density function

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} \quad (10.4.5)$$

satisfies the following moment equations:

$$E[X^k] = \begin{cases} 0 & n = 2k + 1 \text{ for all } k \\ 1 \cdot 3 \cdot 5 \cdots (n-1)\sigma^n & n = 2k \text{ for all } k \end{cases} \quad (10.4.6)$$

Clearly the odd moments are 0 because  $f_X(x)$  is an even function. To show the result for the even moments, we take the equation  $\int_{-\infty}^{\infty} e^{-bx^2} dx = \sqrt{\pi/b}$  and differentiate it  $k$  time with respect to  $b$ , yielding

$$\int_{-\infty}^{\infty} x^{2k} e^{-bx^2} dx = \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{2^k} \sqrt{\frac{\pi}{b^{2k+1}}} \quad (10.4.7)$$

and the result is obtained by substituting  $b = 1/2\sigma^2$  in Eq. (10.4.7). Thus, the Gaussian random variable is completely determined if only two moments are known.

**Higher-Order Cross-Moments**

The concept of joint moments can be extended to higher-order cases. The  $k,j$ th higher-order cross-moment is defined by,

$$E[X^k, Y^j] = m_{kj} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k y^j f_{XY}(x,y) dx dy \quad (10.4.8)$$

and the  $k,j$ th higher order central moment is defined by

$$E[(X - \mu_X)^k, (Y - \mu_Y)^j] = \eta_{kj} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^k (y - \mu_Y)^j f_{XY}(x,y) dx dy \quad (10.4.9)$$

and  $m_{11} = E[XY]$  the joint moment and  $\eta_{11} = \text{cov}[XY] = \sigma_{XY}$  is the covariance.

## 10.5 BIVARIATE GAUSSIAN

Two random variables are jointly Gaussian-distributed if they satisfy the following density function with  $\rho$  as the correlation coefficient:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ \frac{1}{2(1-\rho^2)} \left[ \left( \frac{x - \mu_X}{\sigma_X} \right)^2 - 2\rho \frac{(x - \mu_X)(y - \mu_Y)}{\sigma_X\sigma_Y} + \left( \frac{y - \mu_Y}{\sigma_Y} \right)^2 \right] \right\} \quad (10.5.1)$$

Ordinarily, uncorrelatedness does not imply independence, even though independence implies uncorrelatedness. However, in the case of Gaussian-distributed random variables if  $\rho = 0$ , then Eq. (10.5.1) becomes

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp \left\{ \frac{1}{2} \left[ \left( \frac{x - \mu_X}{\sigma_X} \right)^2 + \left( \frac{y - \mu_Y}{\sigma_Y} \right)^2 \right] \right\} \quad (10.5.2)$$

and clearly

$$\begin{aligned} f_{XY}(x,y) &= f_X(x)f_Y(y) \\ &= \frac{1}{\sqrt{2\pi}\sigma_X} \exp \left\{ \frac{1}{2} \left( \frac{x - \mu_X}{\sigma_X} \right)^2 \right\} \frac{1}{\sqrt{2\pi}\sigma_Y} \exp \left\{ \frac{1}{2} \left( \frac{y - \mu_Y}{\sigma_Y} \right)^2 \right\} \end{aligned} \quad (10.5.3)$$

and hence  $X$  and  $Y$  are independent.

# Characteristic Functions and Generating Functions

## 11.1 CHARACTERISTIC FUNCTIONS

The characteristic function (CF) of any random variable  $X$  (discrete or continuous) is defined by

$$\Phi_X(\omega) = E\{e^{j\omega X}\} = \int_{-\infty}^{\infty} e^{j\omega x} f_X(x) dx \quad (11.1.1)$$

From this definition it is seen that the characteristic function is the Fourier transform of the pdf  $f_X(x)$  where the sign of  $\omega$  is positive instead of negative. However, all the properties of the Fourier transform hold good. From the definition of  $\Phi_X(\omega)$ , we can see that

$$\Phi_X(0) = 1 \quad (11.1.2)$$

This equation is used to check the correctness of the CF.

### Existence of CF

The CF is bounded as shown below. Applying the fact that the absolute value of an integral is less than or equal to the integral of its absolute value, we have

$$\left| \int_{-\infty}^{\infty} e^{j\omega x} f_X(x) dx \right| \leq \int_{-\infty}^{\infty} |e^{j\omega x}| |f_X(x)| dx$$

and since  $|e^{j\omega x}| = 1$  we have

$$\begin{aligned} \left| \int_{-\infty}^{\infty} e^{j\omega x} f_X(x) dx \right| &\leq \int_{-\infty}^{\infty} |e^{j\omega x}| |f_X(x)| dx \\ &= \int_{-\infty}^{\infty} |f_X(x)| dx = 1 \end{aligned}$$

Thus, the CF always exists.

We can also obtain the pdf  $f_X(x)$  from the CF from the inversion formula for the Fourier transform given by

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_X(\omega) e^{-j\omega x} d\omega \quad (11.1.3)$$

[*Comment:* Since  $\Phi_X(\omega)$  is the Fourier transformation of the pdf  $f_X(x)$ , the random variable  $X$  is also completely determined if  $\Phi_X(\omega)$  is known. A Fourier transform of a signal characterizes the frequency content of the signal. No such characterization for the CF can be given; it is just a convenient form of representing the pdf and has other very useful properties.]

**Example 11.1.1** We will find the CF of an exponential distribution given by

$$f_X(x) = \lambda e^{-\lambda x} u(x)$$

Using Eq. (11.1.1), we have

$$\Phi_X(\omega) = \int_0^{\infty} \lambda e^{-\lambda x} e^{j\omega x} dx = \frac{\lambda}{\lambda - j\omega}$$

### **Moment Generating Properties of CF**

In Chapter 10 we found the moments of a random variable using integration techniques. Once the CF is known, the moments can be obtained by differentiation, which is a simpler process than integration. Expanding the CF as a Taylor series about the origin, we can write

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x) \left[ 1 + \frac{j\omega x}{1!} + \frac{(j\omega x)^2}{2!} + \frac{(j\omega x)^3}{3!} + \cdots + \frac{(j\omega x)^k}{k!} + \cdots \right] d\omega \quad (11.1.4)$$

Substituting

$$m_k = \int_{-\infty}^{\infty} x^k f_X(x) dx, \quad k = 0, 1, \dots \quad (11.1.5)$$

in Eq. (11.1.2), we have

$$\Phi_X(\omega) = \left[ 1 + \frac{j\omega m_1}{1!} + \frac{(j\omega)^2 m_2}{2!} + \frac{(j\omega)^3 m_3}{3!} + \cdots + \frac{(j\omega)^k m_k}{k!} + \cdots \right]$$

We now differentiate  $\Phi_X(\omega)$  term by term to yield

$$\frac{d}{d\omega} \Phi_X(\omega) = \left[ \frac{jm_1}{1!} + \frac{2(j)^2 \omega m_2}{2!} + \frac{3(j)^3 \omega^2 m_3}{3!} + \cdots + \frac{k(j)^k \omega^{k-1} m_k}{k!} + \cdots \right] \quad (11.1.6)$$

Substituting  $\omega = 0$  in Eq. (11.1.6), we get

$$\Phi'_X(\omega)|_{\omega=0} = jm_1 \quad \text{and} \quad m_1 = \frac{\Phi'_X(0)}{j} \quad (11.1.7)$$

Differentiating  $\Phi_X(\omega)$  successively  $k$  times and substituting in the resulting equations  $\omega = 0$  we can obtain all the moments of the random variables as shown below:

$$\Phi_X^{(2)}(\omega)|_{\omega=0} = (j)^2 m_2, \dots, \Phi_X^{(k)}(\omega)|_{\omega=0} = (j)^k m_k, \dots, \quad (11.1.8)$$

We will give more examples in a later section.

### Joint Characteristic Functions

We can also define a joint characteristic function for two random variables  $X$  and  $Y$  as

$$\begin{aligned}\Phi_{XY}(\omega_1, \omega_2) &= E\{e^{j(\omega_1 X + \omega_2 Y)}\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) e^{j(\omega_1 x + \omega_2 y)} dx dy\end{aligned}\quad (11.1.9)$$

## 11.2 EXAMPLES OF CHARACTERISTIC FUNCTIONS

We will now give several examples of characteristic functions for both discrete and continuous random variables and determining their means and variances.

### Discrete Random Variables

#### Example 11.2.1 (Binomial Distribution)

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k} \quad (11.2.1)$$

The CF is given by

$$\begin{aligned}\Phi_X(\omega) &= \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} e^{j\omega k} \\ &= \sum_{k=0}^n \binom{n}{k} p e^{j\omega k} q^{n-k} = (p e^{j\omega} + q)^n\end{aligned}\quad (11.2.2)$$

We find that  $\Phi_X(0) = (p + q) = 1$

Mean:

$$\begin{aligned}\Phi'_X(\omega) &= n(p e^{j\omega} + q)^{n-1} j p e^{j\omega} \\ \Phi'_X(0) &= jnp \quad \text{or} \quad E\{X = k\} = m_1 = np\end{aligned}\quad (11.2.3)$$

Variance. We first find  $E\{X^2 = k^2\} = m_2$ :

$$\begin{aligned}\Phi''_X(\omega) &= n(n-1)(p e^{j\omega} + q)^{n-2} (j p e^{j\omega})^2 + n(p e^{j\omega} + q)^{n-1} j^2 p e^{j\omega} \\ \Phi''_X(0) &= -n(n-1)p^2 - np \quad \text{and} \quad m_2 = (np)^2 + npq\end{aligned}$$

Hence

$$\text{var}[X] = m_2 - m_1^2 = npq \quad (11.2.4)$$

**Example 11.2.2 (Poisson Distribution)**

$$p(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (11.2.5)$$

$$\begin{aligned} \Phi_X(\omega) &= \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} e^{j\omega k} \\ &= \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda e^{j\omega k}}{k!} = e^{-\lambda} e^{\lambda e^{j\omega}} = e^{-\lambda(1-e^{j\omega})} \end{aligned} \quad (11.2.6)$$

and  $\Phi_X(0) = 1$ .

*Mean:*

$$\begin{aligned} \Phi'_X(\omega) &= -e^{-\lambda(1-e^{j\omega})} \lambda (-e^{j\omega}) j \\ \Phi'_X(0) &= \lambda j \quad \text{or} \quad E\{X = k\} = m_1 = \lambda \end{aligned} \quad (11.2.7)$$

*Variance.* We first find  $E\{X^2 = k^2\} = m_2$ :

$$\begin{aligned} \Phi''_X(\omega) &= -e^{-\lambda(1-e^{j\omega})} \lambda (-e^{j\omega}) j \cdot \lambda (-e^{j\omega}) j - e^{-\lambda(1-e^{j\omega})} \lambda j (-e^{j\omega}) j \\ \Phi''_X(0) &= -\lambda^2 - \lambda \quad \text{or} \quad m_2 = \lambda^2 + \lambda \end{aligned}$$

Hence

$$\text{var}\{X\} = m_2 - m_1^2 = \lambda \quad (11.2.8)$$

**Example 11.2.3 (Geometric Distribution)**

$$p\{k:p\} = P\{k \text{ failures followed by a success}\} = q^k p \quad (11.2.9)$$

$$\begin{aligned} \Phi_X(\omega) &= \sum_{k=0}^{\infty} pq^k e^{j\omega k} = p \sum_{k=0}^{\infty} (qe^{j\omega})^k \\ &= \frac{p}{1 - qe^{j\omega}} \end{aligned} \quad (11.2.10)$$

and  $\Phi_X(0) = 1$ .

*Mean:*

$$\begin{aligned} \Phi'_X(\omega) &= \frac{-p}{(1 - qe^{j\omega})^2} \cdot (-jq e^{j\omega}) = \frac{jpqe^{j\omega}}{(1 - qe^{j\omega})^2} \\ \Phi'_X(0) &= \frac{jpq}{p^2} \quad \text{or} \quad E\{x = k\} = m_1 = \frac{q}{p} \end{aligned} \quad (11.2.11)$$

*Variance.* We first find  $E\{X^2 = k^2\} = m_2$ .

$$\begin{aligned} \Phi''_X(\omega) &= \frac{-2jpqe^{j\omega}}{(1 - qe^{j\omega})^3} \cdot (-jq e^{j\omega}) + \frac{j^2 p q e^{j\omega}}{(1 - qe^{j\omega})^2} \\ \Phi''_X(0) &= \frac{-2pq^2}{p^3} - \frac{pq}{p^2} \quad \text{or} \quad m_2 = \frac{2q^2}{p^2} + \frac{q}{p} \end{aligned} \quad (11.2.12)$$

Hence variance is

$$m_2 - \mu_X^2 = \frac{2q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2} = \frac{q(p+q)}{p^2} = \frac{q}{p^2} \quad (11.2.13)$$

### Continuous Random Variables

#### Example 11.2.4 (Uniform Distribution)

$$f_X(x) = \frac{1}{b-a} [u(x-b) - u(x-a)] \quad (11.2.14)$$

$$\Phi_X(\omega) = \int_a^b \frac{e^{j\omega x}}{b-a} dx = \frac{e^{j\omega b} - e^{j\omega a}}{j\omega(b-a)} \quad (11.2.15)$$

This is one of the more difficult CFs to obtain the mean and the variance by differentiating Eq. (11.2.15) because of the limiting processes involved. Instead, we expand Eq. (11.2.15) using Taylor series and differentiate the resulting expression. Expansion of Eq. (11.2.15) results in

$$\begin{aligned} \Phi_X(\omega) &= \frac{1}{j\omega(b-a)} \left\{ j\omega(b-a) + \frac{(j\omega)^2}{2!}(b^2 - a^2) + \frac{(j\omega)^3}{3!}(b^3 - a^3) + \dots \right\} \\ &= \left\{ 1 + \frac{j\omega}{2!}(b+a) + \frac{(j\omega)^2}{3!}(b^2 + ab + a^2) + \dots \right\} \end{aligned} \quad (11.2.16)$$

Mean:

$$\begin{aligned} \Phi'_X(\omega) &= \left\{ \frac{j}{2!}(b+a) + \frac{2j^2\omega}{3!}(b^2 + ab + a^2) + \dots \right\} \\ \Phi'_X(0) &= \frac{j(b+a)}{2} \quad \text{or} \quad m_1 = \frac{(b+a)}{2} \end{aligned} \quad (11.2.17)$$

Variance:

$$\begin{aligned} \Phi''_X(\omega) &= \left\{ \frac{2j^2}{3!}(b^2 + ab + a^2) + \dots \right\} \\ \Phi''_X(0) &= \frac{-(b^2 + ab + a^2)}{3} \quad \text{or} \quad m_2 = \frac{b^2 + ab + a^2}{3} \end{aligned}$$

Hence

$$\text{var}[X] = m_2 - m_1^2 = \frac{(b-a)^2}{12} \quad (11.2.18)$$

In most cases it is easier to find the mean and variance using CF. However, in this case it is easier to obtain mean and variance by integration techniques using their definitions.

#### Example 11.2.5 (Exponential Distribution)

$$f_X(x) = \lambda e^{-\lambda x} u(x) \quad (11.2.19)$$

From Example 11.1.1 the CF is

$$\Phi_X(\omega) = \int_0^\infty \lambda e^{-\lambda x} e^{j\omega x} dx = \frac{\lambda}{\lambda - j\omega} \quad (11.2.20)$$

*Mean:*

$$\begin{aligned} \Phi'_X(\omega) &= \frac{j\lambda}{(\lambda - j\omega)^2} \\ \Phi'_X(0) &= \frac{j}{\lambda} \quad \text{or} \quad m_1 = \frac{1}{\lambda} \end{aligned} \quad (11.2.21)$$

*Variance:*

$$\begin{aligned} \Phi''_X(\omega) &= \frac{-2\lambda}{(\lambda - j\omega)^3} \\ \Phi''_X(0) &= \frac{-2}{\lambda^2} \quad \text{or} \quad m_2 = \frac{2}{\lambda^2} \end{aligned} \quad (11.2.22)$$

Hence

$$\text{var}[X] = m_2 - m_1^2 = \frac{1}{\lambda^2} \quad (11.2.23)$$

### Example 11.2.6 (Gaussian Distribution)

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad (11.2.24)$$

$$\Phi_X(\omega) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2 + j\omega x} dx \quad (11.2.25)$$

Completing the squares in the exponent of Eq. (11.2.25), we obtain the resulting expression as follows:

$$\Phi_X(\omega) = e^{j\omega\mu - [(\sigma^2\omega^2)/2]} \left\{ \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[-\frac{(x-\mu-j\omega\sigma^2)^2}{2\sigma^2}\right] dx \right\}$$

The value of the integral within braces is 1 and hence the CF of a Gaussian is

$$\Phi_X(\omega) = e^{j\omega\mu - [(\sigma^2\omega^2)/2]} \quad (11.2.26)$$

*Mean:*

$$\begin{aligned} \Phi'_X(\omega) &= e^{j\omega\mu - [(\sigma^2\omega^2)/2]} (j\mu - \sigma^2\omega) \\ \Phi'_X(0) &= j\mu \quad \text{or} \quad m_1 = \mu \end{aligned} \quad (11.2.27)$$

*Variance:*

$$\begin{aligned} \Phi''_X(\omega) &= e^{j\omega\mu - [(\sigma^2\omega^2)/2]} (j\mu - \sigma^2\omega)^2 - e^{j\omega\mu - [(\sigma^2\omega^2)/2]} \sigma^2 \\ \Phi''_X(0) &= -\mu^2 - \sigma^2 \quad \text{or} \quad m_2 = \mu^2 + \sigma^2 \end{aligned}$$

Hence

$$\text{var}[X] = m_2 - m_1^2 = \sigma^2 \quad (11.2.28)$$

## 11.3 GENERATING FUNCTIONS

Certain simplifications are possible in the definition of the characteristic functions for discrete random variables taking positive integral values. This is called the *generating function* (GF) and is defined for a discrete random variable  $X$  taking positive discrete values as

$$P_X(z) = E\{z^k\} = \sum_{k=0}^{\infty} z^k P\{X = k\} \quad (11.3.1)$$

with  $P_X(1) = 1$ .

The infinite series of Eq. (11.3.1) will converge for values of  $|z| < 1$ . Equation (11.3.1) can be recognized as the  $z$  transform of discrete data with the sign reversal for  $z$ .

**Example 11.3.1** Six dice are tossed. Using the generating function, we want to find the probability that the sum of the faces of the dice add to 24.

The GF for the toss of a single die is

$$P_X(z) = \frac{1}{6}[z + z^2 + z^3 + z^4 + z^5 + z^6]$$

and in closed form this can be written as

$$P_X(z) = \frac{z}{6} \left[ \frac{1 - z^6}{1 - z} \right]$$

Since the tosses are independent, the GF for the toss of 6 dice is  $[P_X(z)]^6$  given by

$$Q_X(z) = [P_X(z)]^6 = \left( \frac{z}{6} \right)^6 \left[ \frac{1 - z^6}{1 - z} \right]^6$$

To find  $P\{\text{faces add up to } 24\}$ , we have to first find the coefficient of  $z^{24}$  in the power series expansion of  $Q(z)$ . Alternately, we have to find the coefficient of  $z^{18}$  in the expansion of

$$R_X(z) = \frac{(1 - z^6)^6}{(1 - z)^6}$$

The expansion of the numerator of  $R_X(z)$  is

$$(1 - z)^6 = 1 - 6z^6 + 15z^{12} - 20z^{18} + 15z^{24} - 6z^{30} + z^{36}$$

and the expansion of the denominator of  $R_X(z)$  is given by the negative exponential

$$(1 - z)^{-6} = \sum_{k=1}^{\infty} \frac{(5 + k - 1)!}{5!(k - 1)!} z^{k-1}$$

We now multiply the two expansions above and find the coefficient  $z^{18}$  in the resulting product. The terms containing  $z^{18}$  are

$$\begin{aligned} z^{18} & \left( -20 + 15 \cdot \frac{(5+7-1)}{5!6!} - 6 \cdot \frac{(5+13-1)!}{5!12!} + 1 \cdot \frac{(5+19-1)!}{5!18!} \right) \\ & = 3451z^{18} \end{aligned}$$

Hence the probability that the sum of the faces add to 24 is

$$P\{\text{sum} = 24\} = \frac{3451}{6^6} = 0.073967$$

a result that would have been very difficult to obtain without generating functions.

### **Moment Generating Properties of GF**

It is easier to use the GF in finding the moments of a discrete random variable taking positive integral values than finding it through CF.

We can differentiate Eq. (11.3.1) and obtain

$$\begin{aligned} P_X(z) &= \sum_{k=0}^{\infty} z^k P\{X = k\} \\ P'_X(z) &= \sum_{k=0}^{\infty} kz^{k-1} P\{X = k\} \\ P'_X(1) &= \sum_{k=0}^{\infty} kP\{X = k\} = m_1 \end{aligned} \tag{11.3.2}$$

Differentiating  $P'_X(z)$  again, we obtain

$$\begin{aligned} P''_X(z) &= \sum_{k=0}^{\infty} k(k-1)P\{X = k\} \\ &= \sum_{k=0}^{\infty} k^2 P\{X = k\} - \sum_{k=0}^{\infty} kP\{X = k\} \\ P''_X(1) &= m_2 - m_1 \end{aligned} \tag{11.3.3}$$

Hence the variance is given by

$$\text{var}[X] = P''_X(1) + P'_X(1) - (P'_X(1))^2 \tag{11.3.4}$$

Unlike the CF, there is no general formula for the  $k$ th moment in the case of GF. In particular

$$P'''_X(1) = m_3 - 3m_2 + 2m_1 \tag{11.3.5}$$

## **11.4 EXAMPLES OF GENERATING FUNCTIONS**

We will now derive generating functions for all discrete random variables discussed in Section 11.2 and determine their means and variances.

**Example 11.4.1 (Binomial Distribution)**

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k} \quad (11.4.1)$$

The GF is given by

$$P_X(z) = \sum_{k=0}^N \binom{n}{k} p z^k q^{n-k} = (pz + q)^N \quad (11.4.2)$$

Clearly  $P_X(1) = 1$ .

*Mean:*

$$\begin{aligned} P'_X(z) &= N(pz + q)^{N-1} p \\ P'_X(1) &= Np \quad \text{or} \quad m_1 = Np \end{aligned} \quad (11.4.3)$$

*Variance:*

$$\begin{aligned} P''_X(z) &= N(N-1)(pz + q)^{N-2} p^2 \\ P''_X(1) &= N(N-1)p \quad \text{or} \quad m_2 = N(N-1)p^2 - Np \end{aligned}$$

Hence

$$\text{var}[X] = m_2 - m_1^2 = Npq \quad (11.4.4)$$

This result has been obtained with comparative ease.

**Example 11.4.2 (Poisson Distribution)**

$$p(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!} \quad (11.4.5)$$

The GF is given by

$$P_X(z) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k z^k}{k!} = e^{\lambda(z-1)} \quad (11.4.6)$$

Note that  $P_X(1) = 1$ .

*Mean:*

$$\begin{aligned} P'_X(z) &= e^{\lambda(z-1)} = e^{\lambda(z-1)} \lambda \\ P'_X(1) &= \lambda \quad \text{or} \quad m_1 = \lambda \end{aligned} \quad (11.4.7)$$

a result that has been trivially obtained compared to Eq. (11.2.7).

*Variance:*

$$\begin{aligned} P''_X(z) &= e^{\lambda(z-1)} \lambda^2 \\ P''_X(1) &= \lambda^2 \quad \text{or} \quad m_2 = \lambda^2 - \lambda \end{aligned} \quad (11.4.8)$$

Hence

$$\text{var}[X] = m_2 - m_1^2 = \lambda \quad (11.4.9)$$

A result that has been obtained much more easily than using Eq. (11.2.8).

**Example 11.4.3 (Geometric Distribution)**

$$p\{k: p\} = q^k p \quad (11.4.10)$$

The GF is

$$P_X(z) = \sum_{k=0}^{\infty} qz^k p = \frac{p}{1 - qz} \quad (11.4.11)$$

Here also  $P_X(1) = 1$ .

*Mean:*

$$\begin{aligned} P'_X(z) &= \frac{pq}{(1 - qz)^2} \\ P'_X(1) &= \frac{q}{p} \end{aligned} \quad (11.4.12)$$

*Variance:*

$$\begin{aligned} P''_X(z) &= 2 \frac{pq^2}{(1 - qz)^3} \\ P''_X(1) &= 2 \frac{pq^2}{(1 - qz)^3} = \frac{2q^2}{p^2} \end{aligned} \quad (11.4.13)$$

or

$$m_2 = \frac{2q^2}{p^2} + \frac{q}{p}$$

Hence

$$\text{var}[X] = m_2 - m_1^2 = \frac{q}{p^2} \quad (11.4.14)$$

## 11.5 MOMENT GENERATING FUNCTIONS

The moment generating function (MGF)  $M_X(t)$  if it exists is defined by

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} f_X(x)e^{tx} dx \quad (11.5.1)$$

where  $t$  is a real variable. This is similar to the characteristic function (CF) [Eq. (11.1.1)], except that  $j\omega$  has been replaced by the real variable  $t$ . Unlike the CF, which always exists, the MGF may or may not exist. Expanding  $E(e^{tX})$  in a power series about the origin and taking expectations, if they exist, we can write

$$\begin{aligned} M_X(t) &= E(e^{tX}) = E\left[1 + tX + \frac{t^2X^2}{2!} + \cdots + \frac{t^nX^n}{n!} + \cdots\right] \\ &= 1 + t\mu + \frac{t^2m_2}{2!} + \cdots + \frac{t^nm_n}{n!} + \cdots \end{aligned} \quad (11.5.2)$$

If  $M_X(t)$  exists, Eq. (11.5.2) can be differentiated term by term to obtain

$$\frac{d}{dx} M_X(t) = \mu + \frac{tm_2}{2!} + \cdots + \frac{t^{n-1}m_n}{n!} + \cdots \quad (11.5.3)$$

Substituting  $t = 0$  in Eq. (11.5.3), we obtain

$$M'_X(t)|_{t=0} = \mu \quad (11.5.4)$$

Differentiating Eq. (11.5.2) successively with respect to  $t$  and setting  $t = 0$  in the resulting equation, we obtain the moments as follows:

$$M_X^{(2)}(t)|_{t=0} = m_2, \dots, m_X^{(n)}(t)|_{t=0} = m_n, \dots \quad (11.5.5)$$

These results are very similar to those in Eq. (11.1.5) except for the absence of  $j = \sqrt{-1}$ .

Since the methodology of finding the moments using MGF is very similar to that of using CF, we will give only three examples.

**Example 11.5.1 (Negative Binomial—Pascal)** If  $N$  is the random variable, then the pmf is

$$nb\{n; k, p\} = P\{n \text{ trials produce } k \text{ successes}\} = \binom{n-1}{k-1} p^k q^{n-k}; n = k, k+1, \dots$$

Or substituting  $m = n - k$ , we can write the pmf with  $M$  as the random variable:

$$\begin{aligned} nb\{m; k, p\} &= P\{(m+k) \text{ trials produce } k \text{ successes}\} \\ &= \binom{k+m-1}{k-1} p^k q^m, \quad m = 0, 1, \dots \end{aligned}$$

Since

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

we can write the MGF for the random variable  $M$ :

$$M_M(t) = \sum_{m=0}^{\infty} \binom{k+m-1}{m} (qe^t)^m p^k$$

The negative binomial expansion of  $(1 - qe^t)^{-k}$  can be written as follows:

$$(1 - qe^t)^{-k} = \sum_{m=0}^{\infty} \binom{k+m-1}{m} (qe^t)^m$$

Substituting this equation in the previous one, we obtain the MGF:

$$M_M(t) = p^k (1 - qe^t)^{-k}$$

Clearly  $M_M(0) = 1$ .

*Mean:*

$$M'_M(t) = p^k qk(1 - qe^t)^{-k-1} e^t$$

$$M'_M(0) = E(M) = \frac{kq}{p}$$

and

$$E(N) = \frac{kq}{p} + k = \frac{k}{p}$$

*Variance.* We first find  $E\{M^2\} = m_2$ :

$$M''_M(t) = p^k qk(1 - qe^t)^{-k-2} (kqe^t + 1)e^t$$

$$M''_M(0) = E(M^2) = \frac{qk(1 + qk)}{p^2}$$

Hence the variance of  $M$  is

$$\text{var}[M] = \frac{qk(1 + qk)}{p^2} - \frac{(qk)^2}{p^2} = \frac{qk}{p^2}$$

The variance of  $N$  can be calculated by using  $N = M + k$  in the equations above.

**Example 11.5.2 (Gaussian Distribution)** The MGF of the Gaussian distribution is given by

$$M_X(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-[(x-\mu)^2/2\sigma^2] + tx} dx$$

Completing the squares, we obtain

$$M_X(t) = e^{t\mu + [(\sigma^2 t^2)/2]} \left[ \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[-\frac{(x - \mu - t\sigma^2)^2}{2\sigma^2}\right] dx \right]$$

The value of the integral within braces is 1, and hence the MGF of a Gaussian is

$$M_X(t) = e^{t\mu + [(\sigma^2 t^2)/2]}$$

*Mean:*

$$M'_X(t) = e^{t\mu - [(\sigma^2 t^2)/2]}(\mu + \sigma^2 t)$$

$$M'_X(0) = \mu$$

*Variance:*

$$M''_X(t) = e^{t\mu + [(\sigma^2 t^2)/2]}(\mu + \sigma^2 t)^2 + e^{t\mu + [(\sigma^2 t^2)/2]} \sigma^2$$

$$M''_X(0) = \mu + \sigma^2 \quad \text{and} \quad \text{var}[X] = \sigma^2$$

**Example 11.5.3 (Gamma Distribution)** We will find the MGF of a gamma density given by

$$f_X(x) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \quad \alpha, \lambda, x > 0$$

From Eq. (11.5.1), we have

$$\begin{aligned} M_X(t) &= \int_0^\infty \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} e^{tx} dx, \quad \alpha, \lambda > 0 \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-t)x} dx \end{aligned}$$

Substituting  $(\lambda - t)x = y$  and using the definition of a gamma function [Eq. (7.5.2)], we can write this equation as

$$\begin{aligned} M_X(t) &= \frac{\lambda^\alpha}{\Gamma(\alpha)(\lambda - t)^\alpha} \int_0^\infty y^{\alpha-1} e^{-y} dy \\ M_X(t) &= \frac{\lambda^\alpha}{\Gamma(\alpha)(\lambda - t)^\alpha} \Gamma(\alpha) = \frac{\lambda^\alpha}{(\lambda - t)^\alpha} \end{aligned}$$

*Mean:*

$$\begin{aligned} M'_X(t) &= \frac{\lambda^\alpha(-\alpha)(-1)}{(\lambda - t)^{\alpha+1}} = \frac{\alpha\lambda^\alpha}{(\lambda - t)^{\alpha+1}} \\ M'_X(0) &= \frac{\alpha}{\lambda} \end{aligned}$$

*Variance.* We first find  $E\{X^2\} = m_2$ :

$$\begin{aligned} M''_X(t) &= \frac{\alpha(\alpha+1)\lambda^\alpha}{(\lambda - t)^{\alpha+2}} \\ M''_X(0) &= m_2 = \frac{\alpha(\alpha+1)\lambda^\alpha}{\lambda^{\alpha+2}} = \frac{\alpha^2 + \alpha}{\lambda^2} \end{aligned}$$

Hence the variance is

$$m_2 - \mu_X^2 = \frac{\alpha^2 + \alpha}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}$$

The means and variances of most of the commonly occurring discrete (Chapter 4) and continuous (Chapters 6 and 7) distributions are shown in Section 11.7.

## 11.6 CUMULANT GENERATING FUNCTIONS

The *cumulant generating function* (CGF)  $K_X(t)$  is defined as the logarithm of the moment generating function  $M_X(t)$ , or  $K_X(t) = \ln[M_X(t)]$ . The *cumulants* also called semiinvariants  $\{\kappa_i\}$  are the coefficients of CGF:

$$K_X(t) = \ln[M_X(t)] = t\kappa_1 + \frac{t^2}{2!}\kappa_2 + \cdots + \frac{t^n}{n!}\kappa_n + \cdots = \sum_{i=0}^{\infty} \frac{t^i}{i!} \kappa_i \quad (11.6.1)$$

We will now find the cumulants in terms of the moments  $m_k$ . From Eq. (11.5.2), we can expand  $\ln[M_X(t)]$  in terms of a Taylor series as follows:

$$\begin{aligned}
 K_X(t) = \ln[M_X(t)] &= \ln \left[ 1 + \left( tm_1 + \frac{t^2 m_2}{2!} + \cdots + \frac{t^n m_n}{n!} + \cdots \right) \right] \\
 &= \left( tm_1 + \frac{t^2 m_2}{2!} + \cdots + \frac{t^n m_n}{n!} + \cdots \right) \\
 &\quad - \frac{1}{2} \left( tm_1 + \frac{t^2 m_2}{2!} + \cdots + \frac{t^n m_n}{n!} + \cdots \right)^2 \\
 &\quad + \frac{1}{3} \left( tm_1 + \frac{t^2 m_2}{2!} + \cdots + \frac{t^n m_n}{n!} + \cdots \right)^3 \pm \cdots \quad (11.6.2)
 \end{aligned}$$

Collecting terms in  $t$ , we have

$$\begin{aligned}
 K_X(t) &= tm_1 + \frac{t^2}{2!} (m_2 - m_1^2) \\
 &\quad + \frac{t^3}{3!} (m_3 - 3m_1 m_2 + 2m_1^3) \\
 &\quad + \frac{t^4}{4!} (m_4 - 3m_2^2 - 4m_1 m_3 + 12m_1^2 m_2 - 6m_1^4) \\
 &\quad + \frac{t^5}{5!} (m_5 - 5m_4 + 30m_1 m_2^2 + 40m_1^2 m_3 - 60m_1^3 m_2 - 24m_1^5) \\
 &\quad + \cdots \quad (11.6.3)
 \end{aligned}$$

Equating coefficients of  $t_i$  in Eqs. (11.6.1) and (11.6.3), the cumulants have been expressed in terms of the moments  $\{m_i\}$  in Eq. (11.6.4), where they are also expressed in terms of the central moments  $v_i = E[X - \mu_X]^i$ :

$$\begin{aligned}
 \kappa_1 &= m_1 = \mu_X \\
 \kappa_2 &= m_2 - m_1^2 = \sigma_X^2 = v_2 \\
 \kappa_3 &= m_3 - 3m_1 m_2 + 2m_1^3 = v_3 \\
 \kappa_4 &= m_4 - 3m_2^2 - 4m_1 m_3 + 12m_1^2 m_2 - 6m_1^4 = v_4 - 3v_2 \\
 \kappa_5 &= m_5 - 5m_4 + 30m_1 m_2^2 + 40m_1^2 m_3 - 60m_1^3 m_2 - 24m_1^5 \\
 &= v_5 - 10v_2 v_3 + \cdots \quad (11.6.4)
 \end{aligned}$$

*Skewness* measures the lack of symmetry of the tails of a density. A density is symmetric if it looks the same on both the left and the right of the centerpoint. Skewness is defined as

$$\gamma_1 = \frac{v_3}{v_2^{3/2}} \quad (11.6.5)$$

*Kurtosis* is defined as a normalized form of the fourth central moment of a distribution. It is a measure of whether the density is peaked or flat relative to a normal distribution. High-kurtosis density functions will have a high peak near the mean, and low-kurtosis

functions will have a flat top. Kurtosis is defined as

$$\gamma_2 = \frac{\nu_4}{\nu_2^2} \quad (11.6.6)$$

For Gaussian distributions the kurtosis is 3.

**Example 11.6.1** We will find the cumulants of the Poisson distribution given by  $p(k; \lambda) = e^{-\lambda}(\lambda^k/k!)$ .

The MGF of the Poisson distribution is  $M_X(t) = e^{-\lambda(1-e^t)}$ , and the CGF is given by

$$K_X(t) = \ln[e^{-\lambda(1-e^t)}] = -\lambda + \lambda e^t$$

Expanding  $K_X(t)$  in Taylor series, we have

$$-\lambda + \lambda e^t = \lambda \left( t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \frac{t^5}{5!} + \dots \right)$$

Thus, all the cumulants are equal to  $\lambda$ , or  $\kappa_i = \lambda$ ,  $i = 1, 2, \dots$ .

**Example 11.6.2** We will find the cumulants of a Gaussian distribution

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right]$$

The MGF is given by  $M_X(t) = e^{t\mu_X + [(t^2\sigma_X^2)/2]}$  and the CGF, by

$$K_X(t) = \ln\left[e^{t\mu_X + [(t^2\sigma_X^2)/2]}\right] = t\mu_X + \frac{t^2\sigma_X^2}{2}$$

and the cumulants are  $\kappa_1 = \mu_X$ ,  $\kappa_2 = \sigma_X^2$  and  $\kappa_i = 0$  for  $i > 2$ . Thus, for the Gaussian distribution, there are only two cumulants.

### Joint Cumulants

By analogy to Eq. (11.6.4), we can define joint cumulants for one to four random variables,  $X_1, X_2, X_3, X_4$ , as follows:

$$\begin{aligned} \kappa_1 &= E[X_1] = \mu_X \\ \kappa_{12} &= E[X_1X_2] - E[X_1]E[X_2] = \sigma_{X_1X_2} \\ \kappa_{123} &= E[X_1X_2X_3] - E[X_1]E[X_2X_3] - E[X_2]E[X_1X_3] \\ &\quad - E[X_3]E[X_1X_2] + 2E[X_1]E[X_2]E[X_3] \\ \kappa_{1234} &= E[X_1X_2X_3X_4] - E[X_1X_2]E[X_3X_4] \\ &\quad - E[X_1X_3]E[X_2X_4] - E[X_1X_4]E[X_2X_3] \\ &\quad - E[X_1]E[X_2X_3X_4] - E[X_2]E[X_1X_3X_4] \\ &\quad - E[X_3]E[X_1X_2X_4] - E[X_4]E[X_1X_2X_3] \\ &\quad + 3E[X_1]E[X_2]E[X_3X_4] + 3E[X_2]E[X_3]E[X_1X_4] \\ &\quad + 3E[X_3]E[X_4]E[X_1X_2] + 3E[X_1]E[X_4]E[X_2X_3] \\ &\quad - 6E[X_1]E[X_2]E[X_3]E[X_4] \end{aligned} \quad (11.6.7)$$

If the mean values of random variables  $Z_1, Z_2, Z_3, Z_4$  are zero, then Eq. (11.6.7) can be considerably simplified as follows:

$$\begin{aligned}\kappa_1 &= E[Z_1] = 0 \\ \kappa_{12} &= E[Z_1 Z_2] \\ \kappa_{123} &= E[Z_1 Z_2 Z_3] \\ \kappa_{1234} &= E[Z_1 Z_2 Z_3 Z_4] - E[Z_1 Z_2] E[Z_3 Z_4] \\ &\quad - E[Z_1 Z_3] E[Z_2 Z_4] - E[Z_1 Z_4] E[Z_2 Z_3]\end{aligned}\tag{11.6.8}$$

If  $X_1, X_2, Y_1, Y_2$  are four random variables, not necessarily independent, we will express the  $\text{cov}[X_1 X_2, Y_1 Y_2]$  in terms of the joint cumulant  $\kappa_{1234}$ :

$$\begin{aligned}\text{cov}[X_1 X_2, Y_1 Y_2] &= E\{[X_1 X_2 - E(X_1 X_2)][Y_1 Y_2 - E(Y_1 Y_2)]\} \\ &= E[X_1 X_2 Y_1 Y_2] - E[X_1 X_2] E[Y_1 Y_2]\end{aligned}\tag{11.6.9}$$

Expressing  $E[X_1 X_2 Y_1 Y_2]$  in terms of the fourth cumulant  $\kappa_{1234}$  given in Eq. (11.6.8), we can rewrite Eq. (11.6.9) as follows:

$$\text{cov}[X_1 X_2, Y_1 Y_2] = \kappa_{1234} + E[X_1 Y_1] E[X_2 Y_2] + E[X_1 Y_2] E[X_2 Y_1]\tag{11.6.10}$$

## 11.7 TABLE OF MEANS AND VARIANCES

$$p + q = 1$$

Random Variable	Mean	Variance
<i>Discrete Distributions</i>		
Bernoulli	$p$	$pq$
Binomial	$Np$	$Npq$ or $Np(1 - p)$
Multinomial	$np_i, \quad i = 1, \dots, m$	$np_i(1 - p_i), \quad i = 1, \dots, m$
Geometric	$\frac{q}{p}$	$\frac{q}{p^2}$
Negative binomial	$E(M) = \frac{kq}{p}$ for $M = N - k$	$\text{var}(M) = \frac{kq}{p^2}$ for $M = N - k$
Hypergeometric	$\frac{Kn}{N}$	$\frac{nK(N - K)(N - n)}{N^2(n - 1)}$
Poisson	$\lambda$	$\lambda$
Logarithmic— Benford	$\log\left[\frac{2^2 \cdot 5^8}{7 \cdot 3^4}\right] \cong 3.44$ $(d = 1, \dots, 9)$	$\log\left[\frac{2^8 \cdot 5^{72}}{7^{13} \cdot 3^{50}}\right] - \log\left[\frac{2^2 \cdot 5^8}{7 \cdot 3^4}\right]^2 \cong 6.06$ $(d = 1, \dots, 9)$
<i>Continuous Distributions</i>		
Uniform	$\frac{b + a}{2}$	$\frac{(b - a)^2}{12}$

(Continued)

Random Variable	Mean	Variance
Exponential	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gaussian	$\mu$	$\sigma^2$
Triangular	$b$	$\frac{(b-a)^2}{6}$
Laplace—double-exponential	0	$\frac{2}{\lambda^2}$
Erlang	$\frac{n}{\lambda}$ , $n$ integer	$\frac{n}{\lambda^2}$ , $n$ integer
Gamma	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Weibull	$\mu + \frac{1}{\lambda} \Gamma\left(\frac{\alpha+1}{\alpha}\right)$	$\frac{1}{\lambda^2} \left[ \Gamma\left(\frac{\alpha+2}{\alpha}\right) - \Gamma^2\left(\frac{\alpha+1}{\alpha}\right) \right]$
Chi-square	$n\sigma^2$	$2n\sigma^4$
Chi	$\frac{\sqrt{2} \cdot \sigma \cdot \Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$	$\sigma^2 \left[ n - \frac{2\Gamma^2\left(\frac{n+1}{2}\right)}{\Gamma^2\left(\frac{n}{2}\right)} \right]$
Rayleigh	$\sigma\sqrt{\frac{\pi}{2}}$	$\sigma^2\left(2 - \frac{\pi}{2}\right)$
Maxwell	$2\sigma\sqrt{\frac{2}{\pi}}$	$\sigma^2\left(3 - \frac{8}{\pi}\right)$
Rice	$\mu_R = \sigma^2 \sqrt{\frac{\pi}{2}} e^{-(m^2/4\sigma^2)} \left[ \left(1 + \frac{m^2}{2\sigma^2}\right) I_0\left(\frac{m^2}{4\sigma^2}\right) + \frac{m^2}{2\sigma^2} I_1\left(\frac{m^2}{4\sigma^2}\right) \right]$ $\sigma_R^2 = m^2 + 2\sigma^2 - \mu_R^2, \text{ } m: \text{noncentrality parameter}$	
Nakagami	$\left(\frac{\Omega}{m}\right)^{1/2} \frac{\Gamma\left(m + \frac{1}{2}\right)}{\Gamma(m)}$	$\Omega - \frac{\Omega}{m} \left[ \frac{\Gamma\left(m + \frac{1}{2}\right)}{\Gamma(m)} \right]^2$
Student- <i>t</i>	0	$\frac{v}{v-2}, \quad v > 2$
Snedecor F	$\frac{v_2}{v_2-2}, \quad v_2 > 2$	$\frac{2v_2^2(v_1 + v_2 - 2)}{v_1(v_2-2)^2(v_2-4)}, \quad v_2 > 4$
Lognormal	$e^{[\mu + (\sigma^2/2)]}$	$e^{(2\mu + \sigma^2)} (e^{\sigma^2} - 1)$
Beta	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}$

(Continued)

Random Variable	Mean	Variance
Cauchy	Not defined	Not defined
Pareto	$\frac{\beta}{\alpha - 1}$ , $\alpha > 1$	$\frac{\alpha\beta^2}{(\alpha - 1)^2(\alpha - 2)}$ , $\alpha > 2$

# Functions of a Single Random Variable

## 12.1 RANDOM VARIABLE $g(X)$

In many probability problems, we have to find the pdf of a function of random variables. We have already defined (in Chapter 5) a random variable  $X$  as the mapping from the sample space  $S$  to space of real numbers  $R_x$ ; that is, it is the set of all outcomes  $\xi$  belonging to the sample space  $S$  such that the unique number  $X(\xi)$  belongs to the real line  $R_x$ . Thus, the random variable  $X$  is a mapping of the probability space  $\{S, \mathcal{F}, P\}$  to the probability space on the real line  $\{R_x, \mathcal{F}_X, P_X\}$ :

$$X: \{S, \mathcal{F}, P\} \rightarrow \{R_x, \mathcal{F}_X, P_X\} \quad (12.1.1)$$

We can now define a random variable  $Y$  as a function of another random variable  $X$ , or  $Y = g(X)$ , as the mapping from the real line  $R_x$  to another real line  $R_y$ , such that it induces a probability space  $\{R_y, \mathcal{F}_Y, P_Y\}$ . In other words, the random variable  $Y$  is a composite function of the random variable  $X$  such that  $Y(\xi) = g(X) = g[X(\xi)]$ . The mapping is given by

$$Y: \{S, \mathcal{F}, P\} \rightarrow \{R_x, \mathcal{F}_X, P_X\} \rightarrow \{R_y, \mathcal{F}_Y, P_Y\} \quad (12.1.2)$$

and shown in Fig. 12.1.1.

### Conditions on $Y = g(X)$ to be a Random Variable

1. The domain of  $g(x)$  must include the range of the random variable (RV)  $X$ .
2. For each  $y \in I_y$ ,  $\{g(x) \leq y\}$  must form a field  $\mathcal{F}_Y$ ; that is, it must consist of the unions and intersections of a countable number of intervals of the real line  $R_y$ .
3. The probability of the event  $\{g(x) = \pm\infty\}$  is 0.

In essence, the transformation  $Y = g(X)$  induces a probability measure  $P_Y$  on  $\mathcal{F}_Y$ , and we have to find this probability measure. From Fig. 12.1.1 we can write

$$P_Y\{Y\} = P_X\{g^{-1}(Y)\} = P_X\{x : g^{-1}(Y) \subset I_x\} \quad (12.1.3)$$

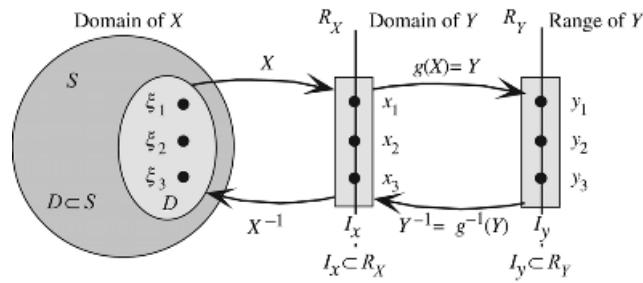


FIGURE 12.1.1

The set  $I_x$  consists of all the countable unions and intersections of the type  $a < x \leq b$ . Without loss in generality, we will assume that  $I_x = \{X \leq y\}$ . Thus, Eq. (12.1.3) becomes

$$P_Y\{Y\} = P\{Y \leq y\} = F_Y(y) = P_X\{x : g^{-1}(Y) \leq y\} = P_X\{x : g(x) \leq y\} \quad (12.1.4)$$

In conclusion, we can say that the probability of  $Y \leq y$  is given by the probability of the set of all values of  $x$  such that  $g(x) \leq y$ .

## 12.2 DISTRIBUTION OF $Y = g(X)$

We will find the distribution  $F_Y(y)$  of the RV  $Y = g(X)$  in terms of the distribution  $F_X(x)$  of the RV  $X$  and the function  $g(x)$ . We shall do this with an arbitrary function  $g(x)$  bounded between  $a$  and  $b$  as shown in Fig. 12.2.1. As  $y$  traverses from  $-\infty$  to  $+\infty$  along the  $g(x)$  axis, we have to find the values of  $x$  along the  $x$  axis that satisfy  $g(x) \leq y$ . In Fig. 12.2.1 there are five distinct regions for  $y$  corresponding to the intersections of  $y$  with  $g(x)$ :

1.  $y \leq a$ , no points of intersection
2.  $a < y \leq y_1$ , one point of intersection  $x_1$
3.  $y_1 < y \leq y_2$ , three points of intersection  $x_2, x'_2, x''_2$
4.  $y_2 < y \leq b$ , one point of intersection  $x_3$
5.  $y \geq b$ , no points of intersection

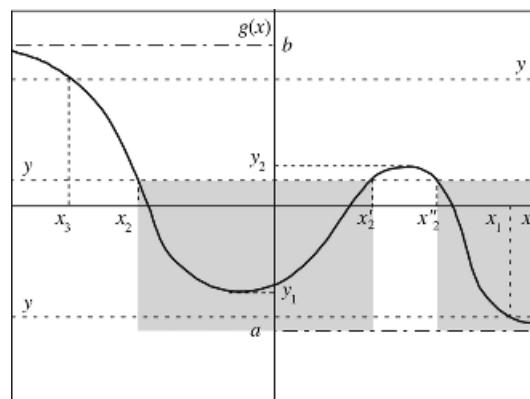


FIGURE 12.2.1

We have to find  $F_Y(y) = \text{probability of } g(x) \leq y$  in each of these five regions.

1.  $y \leq a$ : There are no points along the  $x$  axis for which  $g(x) \leq y$ , or  $I_x = \{\emptyset\}$ ; hence  $F_Y(y) = 0$ .
2.  $a < y \leq y_1$ : The set of all points along the  $x$  axis for which  $g(x) \leq y$ , or  $I_x = (x_1, \infty)$ ; hence  $F_Y(y) = 1 - F_X(x_1)$ .
3.  $y_1 < y \leq y_2$ : The set of all points along the  $x$  axis for which  $g(x) \leq y$ , or  $I_x = (x_2, x'_2] \cup (x''_2, \infty)$ ; hence  $F_Y(y) = F_X(x'_2) - F_X(x_2) + 1 - F_X(x''_2)$ .
4.  $y_2 < y \leq b$ : The set of all points along the  $x$  axis for which  $g(x) \leq y$  or  $I_x = (x_3, \infty)$ ; hence  $F_Y(y) = 1 - F_X(x_3)$ .
5.  $y > b$ : Here the entire  $g(x)$  is below  $y$ ; and hence  $I_x = (-\infty, \infty)$ ; hence  $F_Y(y) = 1$ .

Thus, we have used Eq. (12.1.4) to find  $P\{Y \leq y\}$  along the entire  $g(x)$  axis.

### Special Case

We now consider a situation where  $g(x) = y_1$  for  $x_1 < x \leq x_2$ . In this case we have

$$P\{Y = y_1\} = P\{x_1 < X < x_2\} = F_X(x_2) - F_X(x_1) \quad (12.2.1)$$

Hence  $F_Y(y)$  is discontinuous at  $y = y_1$ , and the discontinuity is equal to  $F_X(x_2) - F_X(x_1)$ .

### Summary of Steps for Finding $F_Y(y) = P\{Y \leq y\}$

1. Graph the function  $g(x)$  along the  $x$  and  $g(x)$  axes. Do not label the  $g(x)$  axis as  $y$ .
2. Draw  $y$  parallel to the  $x$  axis.
3. Find the distinct regions along the  $g(x)$  axis. The regions are determined by the number of intersections of  $y$  with  $g(x)$  as  $y$  goes from  $-\infty$  to  $\infty$  along the  $g(x)$  axis. Let these be  $y_1, y_2, \dots$ .
4. In each region solve for all values of  $x$  for which  $y = g(x)$ . Let these be  $x, x', x'', \dots$ .
5. For each of these regions, find  $I_x$  such that  $g(x) \leq y$ .
6. Find  $F_Y(y) = P\{Y \leq y\} = P\{x \in I_x\} = P\{x : g(x) \leq y\}$ .

We will take a number of examples of increasingly difficult transformations  $g(x)$  and solve for  $P\{Y \leq y\}$  using the steps listed above. We shall assume that in all the following examples  $F_X(x)$  or  $f_X(x)$  is given and we have to find  $F_Y(y)$ .

**Example 12.2.1** Given that  $Y = aX + b$  with  $a > 0$ , we have to find  $F_Y(y)$ . We follow the steps listed above:

1. The function  $g(x) = ax + b$  has been graphed.
2.  $y$  parallel to the  $x$  axis has been drawn in Fig. 12.2.2.
3. In this simple example there is only one region  $-\infty < y < \infty$ .
4. In this region, solving for  $y = ax + b$ , we have  $x = (y - b)/a$ .
5. The region  $I_x$  such that  $g(x) \leq y$  is given by

$$I_x = \left\{ -\infty < x \leq \frac{y - b}{a} \right\}$$

$$6. \quad F_Y(y) = P\{x : g(x) \leq y\} = P\left\{ X \leq \frac{y - b}{a} \right\} = F_X\left(\frac{y - b}{a}\right)$$

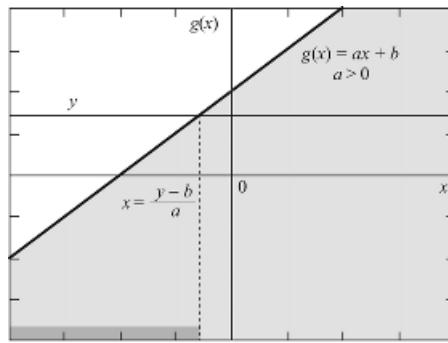


FIGURE 12.2.2

We can now find the density  $f_Y(y)$  by differentiating  $F_Y(y)$  with respect to  $y$ . The result is

$$f_Y(y) = \frac{dF_X((y-b)/a)}{dy} = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

**Example 12.2.2** We repeat Example 12.2.1 with  $a \leq 0$ . We follow the same steps as before.

1. The function  $g(x)$  is graphed in Fig. 12.2.3.
2.  $y$  has been drawn parallel to  $x$ .
3. In this example also, there is only one region  $-\infty < y < \infty$ .
4. In this region, solving for  $y = ax + b$ , we have  $x = (y-b)/a$  or  $x = [(y-b)/|a|]$ .
5. The region  $I_X$  such that  $g(x) \leq y$  is given by

$$\begin{aligned}
 I_X &= \left\{ -\frac{y-b}{|a|} < x \leq \infty \right\} \\
 6. \quad F_Y(y) &= P\{x : g(x) \leq y\} = P\left\{ X > -\frac{y-b}{|a|} \right\} \\
 &= 1 - P\left\{ X > -\frac{y-b}{|a|} \right\} = 1 - F_X\left(-\frac{y-b}{|a|}\right)
 \end{aligned}$$

The corresponding density function  $f_Y(y)$  is given by differentiating  $F_Y(y)$  with respect to  $y$ , and the result is

$$f_Y(y) = \frac{dF_X\left(\frac{y-b}{a}\right)}{dy} = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

The result is very similar to that obtained in Example 12.2.1 except for the absolute value of  $a$ .

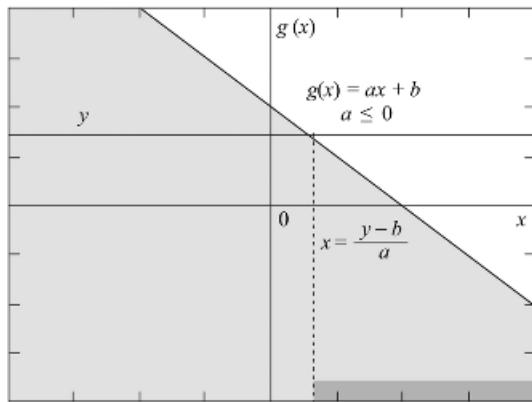


FIGURE 12.2.3

**Example 12.2.3** Let  $g(x) = (x - a)^2$ . This example is different from the previous two examples because there are multiple regions for  $y$ . Following similar steps, we obtain

1. The function  $g(x)$  is graphed in Fig. 12.2.4 and
2.  $y$  has been drawn parallel to  $x$ .
3. Here the two regions of  $y$  are  $\{y \leq 0\}$ , no points of intersection and  $\{y > 0\}$ , two points of intersection.
4. In the region  $\{y \leq 0\}$ ,  $I_x = \emptyset$ . In the region  $\{y > 0\}$ , we solve for  $y = g(x)$  and obtain  $x_1 = a + \sqrt{y}$  and  $x_2 = a - \sqrt{y}$ .
5. Thus,  $I_x = (a - \sqrt{y}, a + \sqrt{y}]$ ,  $y > 0$ .
6. In the first region  $F_Y(y) = 0$ :  $y \leq 0$ . In the second region

$$\begin{aligned} F_Y(y) &= P(a - \sqrt{y} < X \leq a + \sqrt{y}) \\ &= F_X(a + \sqrt{y}) - F_X(a - \sqrt{y}) : y > 0 \end{aligned}$$

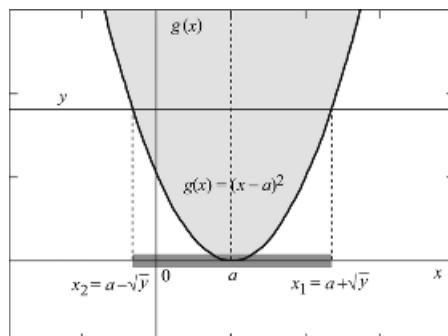


FIGURE 12.2.4

The corresponding density function is

$$\begin{aligned}f_Y(y) &= 0 : y \leq 0 \\f_Y(y) &= \frac{d}{dy}(F_X(a + \sqrt{y}) - F_X(a - \sqrt{y})) \\&= \frac{1}{2\sqrt{y}}(f_X(a + \sqrt{y}) + f_X(a - \sqrt{y})) : y > 0\end{aligned}$$

If  $X$  is a zero mean Gaussian distributed random variable given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x^2/2\sigma^2)}$$

and if  $Y = X^2$ , then  $f_Y(y)$  is given by

$$f_Y(y) = \frac{1}{2\sqrt{y}} \frac{1}{\sigma\sqrt{2\pi}} \left( e^{-(y/2\sigma^2)} + e^{-(y/2\sigma^2)} \right) = \frac{1}{\sigma\sqrt{2\pi y}} e^{-(y/2\sigma^2)}$$

which is a chi-square distribution with one degree of freedom [Eq. (7.7.1)].

**Example 12.2.4** This example is a complement of Example 12.2.3. Here  $g(x) = 1/(x - a)^2$  and we have to find  $F_Y(y)$ . We follow the same procedure laid out earlier.

1. The function  $g(x)$  is graphed in Fig. 12.2.5.
2.  $y$  has been drawn parallel to the  $x$  axis.
3. Here the two regions of  $y$  are  $\{y \leq 0\}$ , no points of intersection and  $\{y > 0\}$ , two points of intersection.
4. In the region  $\{y \leq 0\}$ ,  $I_x = \{\emptyset\}$  and  $F_Y(y) = 0$ . In the region  $\{y > 0\}$ , we solve for  $y = g(x)$  and obtain  $x_1 = a + (1/\sqrt{y})$  and  $x_2 = a - (1/\sqrt{y})$ .
5. Thus

$$I_x = \left( -\infty, a - \frac{1}{\sqrt{y}} \right] \cup \left( a + \frac{1}{\sqrt{y}}, \infty \right]$$

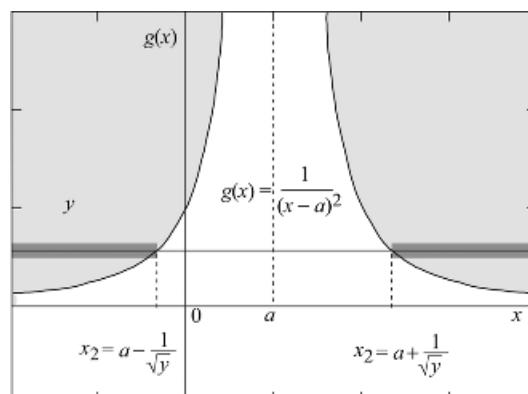


FIGURE 12.2.5

6. In the region indicated in item 5 (above), we obtain

$$\begin{aligned} F_Y(y) &= P\left(X \leq a - \frac{1}{\sqrt{y}}\right) + 1 - P\left(X \leq a + \frac{1}{\sqrt{y}}\right) \\ &= F_X\left(a - \frac{1}{\sqrt{y}}\right) + 1 - F_X\left(a + \frac{1}{\sqrt{y}}\right) : y > 0 \end{aligned}$$

The corresponding density function is

$$f_Y(y) = 0 : y \leq 0$$

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} \left( F_X\left(a - \frac{1}{\sqrt{y}}\right) + 1 - F_X\left(a + \frac{1}{\sqrt{y}}\right) \right) \\ &= \frac{1}{2y^{3/2}} \left[ f_X\left(a - \frac{1}{\sqrt{y}}\right) + f_X\left(a + \frac{1}{\sqrt{y}}\right) \right] : y > 0 \end{aligned}$$

**Example 12.2.5** The previous examples illustrated continuous functions of  $g(x)$ . In this example  $g(x)$  is a discrete function as shown in Fig. 12.2.6.

Note that  $g(x)$  is not a true function because there are many points of  $y$  at  $x = \pm a$ . We can still find  $F_Y(y)$  given  $F_X(x)$ . The procedure is no different from that in the previous examples.

1. The function  $g(x)$  has already been graphed in Fig. 12.2.6.
2.  $y$  has been drawn parallel to the  $x$  axis.
3. There are four region of  $y$ , given by (1)  $y \leq -b$ , (2)  $-b < y \leq 0$ , (3)  $0 < y \leq b$ , (4)  $y > b$ .
4. The intersection points are (1) none, (2)  $x_1 = -a$ , (3)  $x_2 = a$ , (4) none.
5. The corresponding four regions along the  $x$  axis are (1)  $I_x = \{\emptyset\}$ , (2)  $I_x = \{-\infty, -a\}$ , (3)  $I_x = \{-\infty, a\}$ , (4)  $I_x = \{-\infty, \infty\}$ .
6. The distribution function  $F_Y(y)$  in the various regions are (1)  $y \leq -b : F_Y(y) = 0$ , (2)  $-b < y \leq 0 : F_Y(y) = F_X(-a)$ , (3)  $0 < y \leq b : F_Y(y) = F_X(a)$ , (4)  $y > b : F_Y(y) = 1$ . The distribution function  $F_Y(y)$  is a “staircase” function, indicating that  $Y$  is a discrete random variable and the density function  $f_Y(y)$

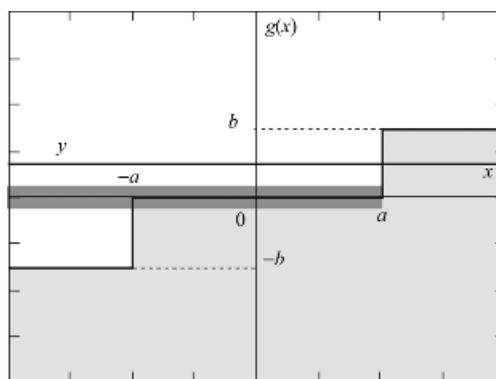


FIGURE 12.2.6

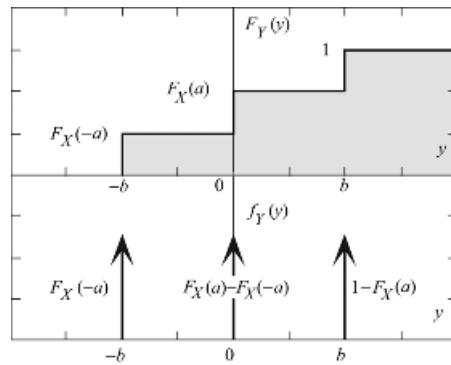


FIGURE 12.2.7

- is (a)  $f_Y(y) = 0$ , (b)  $f_Y(y) = F_X(-a)\delta(x + b)$ , (c)  $f_Y(y) = [F_X(a) - F_X(-a)]\delta(x)$ , (d)  $[1 - F_X(a)]\delta(x - b)$ .

The distribution and density functions are shown in Fig. 12.2.7.

**Example 12.2.6** We shall now take an example of a piecewise-continuous  $g(x)$  with a specified  $f_X(x)$  and  $F_X(x)$ . The functions  $g(x)$ ,  $f_X(x)$  and  $F_X(x)$  are shown in Fig. 12.2.8.

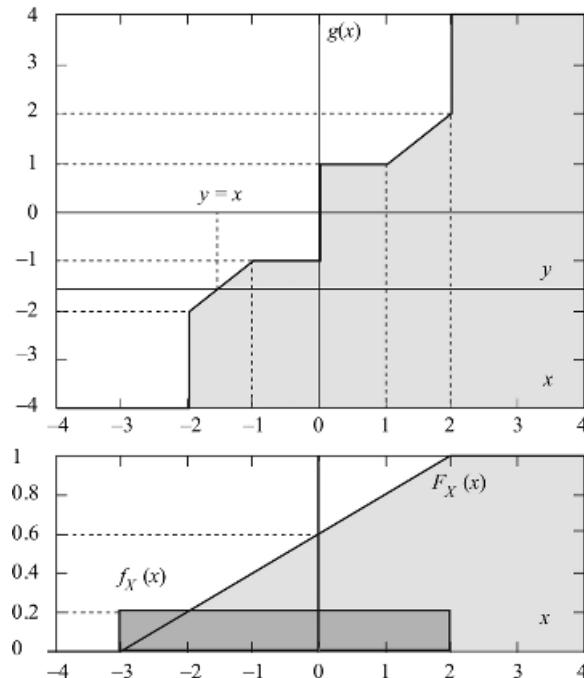


FIGURE 12.2.8

We have to find  $F_Y(y)$  given that  $f_X(x)$  is defined by Fig. 12.2.8 as

$$f_X(x) = \begin{cases} \frac{1}{5} & -3 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Since  $f_X(x)$  is specified,  $F_X(x)$  is determined by integrating  $f_X(x)$ . The result is

$$F_X(x) = \begin{cases} 0 & x \leq -3 \\ \frac{x}{5} + \frac{3}{5} & -3 < x \leq 2 \\ 1 & x > 2 \end{cases}$$

and is shown in Fig. 12.2.8. In this example, the range of  $f_X(x)$ ,  $x \in \{-3, 2\}$  has to be factored in determining  $F_Y(y)$ . Other than that, we will still follow the procedure laid out earlier.

1. The function  $g(x)$  has already been graphed in Fig. 12.2.8.
2.  $y$  will scan from  $-\infty$  to  $\infty$  along the  $g(x)$  axis.
3. The regions of  $y$  are given by (a)  $y \leq -4$ , (b)  $-4 < y \leq -2$ , (c)  $-2 < y \leq -1$ , (d)  $-1 < y \leq 0$ , (e)  $0 < y \leq 1$ , (f)  $1 < y \leq 2$ , (g)  $2 < y \leq 4$ , (h)  $y > 4$ .
4. The points of intersection of  $y$  with  $g(x)$  are (a) none, (b)  $x = -2$ , (c)  $x = y$ , (d)  $x = 0$ , (e)  $x = 1$ , (f)  $x = y$ , (g)  $x = 2$ , (h) none.
5. The corresponding regions  $I_x$  are (a)  $I_x = \{\emptyset\}$ , (b)  $I_x = \{-\infty, -2\}$ , (c)  $I_x = \{-\infty, y\}$ , (d)  $I_x = \{-\infty, 0\}$ , (e)  $I_x = \{-\infty, 1\}$ , (f)  $I_x = \{-\infty, y\}$ , (g)  $I_x = \{-\infty, 2\}$ , (h)  $I_x = \{-\infty, \infty\}$ .
6. The distributions functions in the six regions are
  - (a)  $y \leq -4 : F_Y(y) = 0$
  - (b)  $-4 < y \leq -2 : F_Y(y) = F_X(-2) - F_X(-4) = \frac{1}{5}$
  - (c)  $-2 < y \leq -1 : F_Y(y) = F_X(y) = \frac{y}{5} + \frac{3}{5}$
  - (d)  $-1 < y \leq 0 : F_Y(y) = F_X(0) = \frac{3}{5}$
  - (e)  $0 < y \leq 1 : F_Y(y) = F_X(1) = \frac{4}{5}$
  - (f)  $1 < y \leq 2 : F_Y(y) = F_X(y) = \frac{y}{5} + \frac{3}{5}$
  - (g)  $2 < y \leq 4 : F_Y(y) = F_X(2) = 1$  since  $F_X(x) = 1$  for all values of  $x > 2$
  - (h)  $y > 4 : F_Y(y) = 1$

By differentiating  $F_Y(y)$  in every region, we can express the density function  $f_Y(y)$  as

$$\begin{aligned} f_Y(y) = & \frac{1}{5}\delta(y+3) + \frac{1}{5}[u(y+1) - u(y+2)] + \frac{1}{5}\delta(y+1) + \frac{1}{5}\delta(y) \\ & + \frac{1}{5}[u(y-2) - u(y-1)] \end{aligned}$$

where  $u(y)$  is the usual step function.  $F_Y(y)$  and  $f_Y(y)$  are graphed in Fig. 12.2.9.

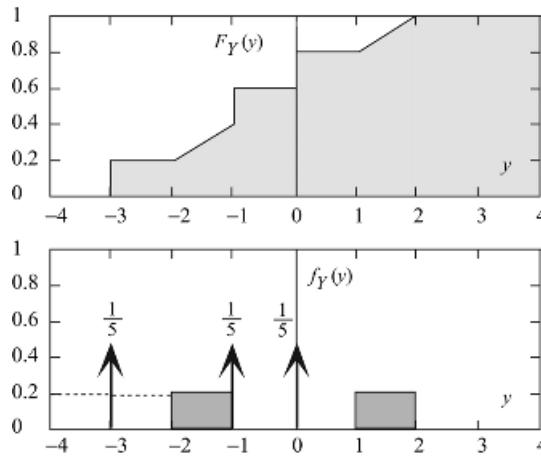


FIGURE 12.2.9

**Example 12.2.7** This example is similar to the previous one with a piecewise-continuous function  $g(x)$ . The functions  $g(x)$ ,  $f_X(x)$  and  $F_X(x)$  are given by

$$g(x) = \begin{cases} 0: & x \leq -2 \\ -1: & -2 < x \leq 0 \\ \frac{2}{3}x: & 0 < x \leq 3 \\ 2: & x > 3 \end{cases}; \quad f_X(x) = \begin{cases} \frac{1}{4}: & -2 < x \leq 2 \\ 0: & \text{otherwise} \end{cases};$$

$$F_X(x) = \begin{cases} 0: & x \leq -2 \\ \frac{x}{4} + \frac{1}{2}: & -2 < x \leq 2 \\ 1: & x > 2 \end{cases}$$

The functions  $g(x)$ ,  $f_X(x)$ , and  $F_X(x)$  are shown in Fig. 12.2.10. We have to find  $F_Y(y)$ .

Following the same procedure as before

1. The function  $g(x)$  has been graphed in Fig. 12.2.10.
2.  $y$  has been drawn, and the intersection of  $y$  with  $g(x)$  is given by  $x = \frac{3}{2}y$ .
3. The regions of  $y$  are given by (a)  $y \leq -1$ , (b)  $-1 < y \leq 0$ , (c)  $0 < y \leq \frac{4}{3}$ , (d)  $\frac{4}{3} < y \leq 2$ , (e)  $y > 2$ . Even though  $g(x)$  extends upto  $x = 3$ , the function  $f_X(x)$  terminates at  $x = 2$  and  $F_X(x) = 1$  beyond  $x = 2$ .
4. The points of intersection of  $y$  with  $g(x)$  are (a) none, (b)  $x = -2, 0$ , (c)  $x = \frac{3}{2}y$ , (d)  $x = \frac{3}{2}y$ , (e) none.
5. The corresponding regions  $I_x$  are (a)  $I_x = \{\emptyset\}$ , (b)  $I_x = (-2, 0]$ , (c)  $I_x = (-\infty, \frac{2}{3}x]$ , (d)  $I_x = (-\infty, \frac{2}{3}x]$ , (e)  $I_x = (-\infty, \infty]$ .
6. The distribution functions in the four regions are (a)  $y \leq -1: F_Y(y) = 0$ , (b)  $-1 < y \leq 0: F_Y(y) = F_X(0) - F_X(-2) = \frac{1}{2}$ , (c)  $0 < y \leq \frac{4}{3}: F_Y(y) = F_X(\frac{3}{2}y) = \frac{3}{8}y + \frac{1}{2}$ , (d)  $\frac{4}{3} < y \leq 2: F_Y(y) = 1$ , (e)  $y \geq 2: F_Y(y) = 1$ .

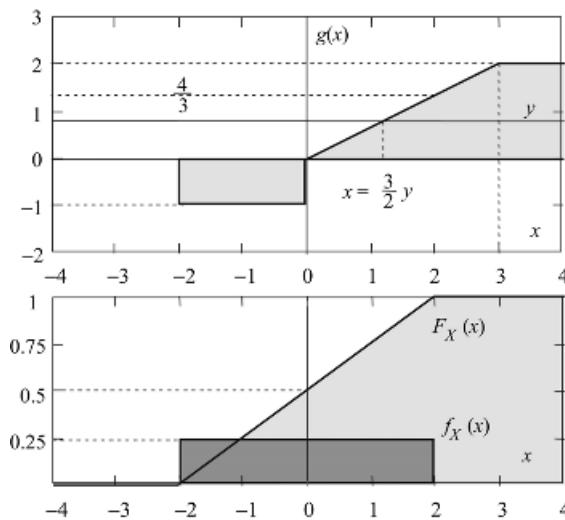


FIGURE 12.2.10

The corresponding density function  $f_Y(y)$  is given by

$$f_Y(y) = \frac{1}{2} \delta(y + 1) + \frac{3}{8} \left[ u\left(y - \frac{4}{3}\right) - u(y) \right]$$

The functions  $F_Y(y)$  and  $f_Y(y)$  are shown in Fig. 12.2.11.

**Example 12.2.8** We shall now take a more complicated example and find  $F_Y(y)$ . In this example  $g(x)$  is defined by

$$g(x) = \frac{1}{x^2 - 1} \quad \text{for all } x$$

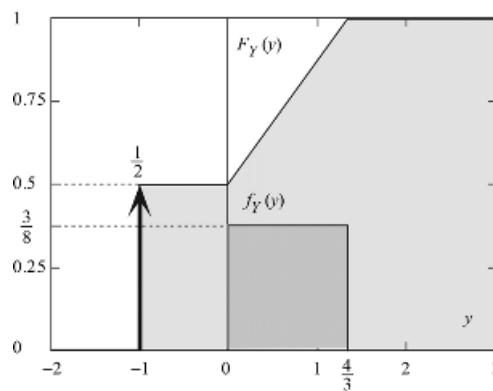


FIGURE 12.2.11

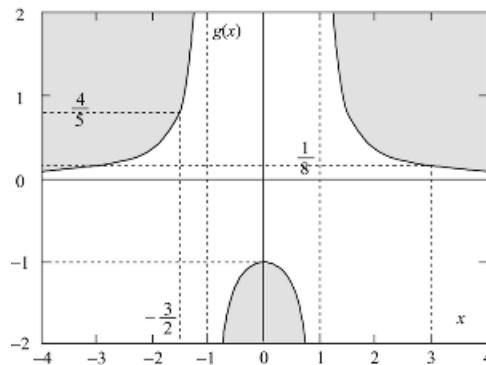


FIGURE 12.2.12

The probability density function  $f_X(x)$  and the distribution function  $F_X(x)$  are given by

$$f_X(x) = \begin{cases} \frac{2}{9} & \text{if } -1.5 < x \leq 3 \\ 0 & \text{otherwise} \end{cases}; \quad F_X(x) = \begin{cases} 0 & \text{if } x \leq -1.5 \\ \frac{2}{9}x + \frac{1}{3} & \text{if } -1.5 \leq x \leq 3 \\ 1 & \text{if } x > 3 \end{cases}$$

The function  $g(x)$  is shown in Fig. 12.2.12 and  $f_X(x)$  and  $F_X(x)$  are shown in Fig. 12.2.13.

1.  $g(x)$  has been graphed in Fig. 12.2.12.
2.  $y$  scans from  $-\infty$  to  $+\infty$  along the  $g(x)$  axis.
3. The regions of  $y$  are given by (a)  $y \leq -1$ , two points of intersection in the range of  $f_X(x)$ ,  $(-1.5, 3]$ ; (b)  $-1 < y \leq 0$ , no points of intersection; (c)  $0 < y \leq \frac{1}{8}$ , two points of intersection beyond the range of  $f_X(x)$ ,  $(-1.5, 3]$ ; (d)  $\frac{1}{8} < y \leq \frac{4}{5}$ , one point of intersection in the range of  $f_X(x)$ ,  $(-1.5, 3]$ , (e)  $y > \frac{4}{5}$ , two points of intersection in the same range.
4. The points of intersection of  $y$  with  $g(x)$  are (a)  $x_1 = -\sqrt{1 + (1/y)}$  and  $x_2 = \sqrt{1 + (1/y)}$ , (b) none, (c)  $x_1 = -\sqrt{1 + (1/y)}$  and  $x_2 = \sqrt{1 + (1/y)}$  beyond the range of  $(-1.5, 3]$ , (d) one point of intersection  $x_2 = \sqrt{1 + (1/y)}$  within the

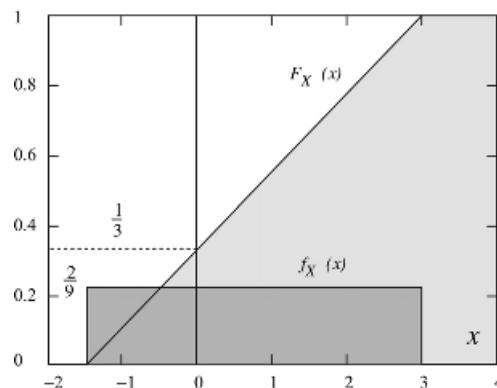


FIGURE 12.2.13

range and another  $x_1 = -\sqrt{1 + (1/y)}$  beyond the range, (e)  $x_1 = -\sqrt{1 + (1/y)}$  and  $x_2 = \sqrt{1 + (1/y)}$ .

5. The corresponding regions  $I_x$  are

$$(a) \quad I_x = \left\{ -1, -\sqrt{1 + \frac{1}{y}} \right\} \cup \left\{ \sqrt{1 + \frac{1}{y}}, 1 \right\}$$

$$(b) \quad I_x = (-1, 1]$$

$$(c) \quad I_x = \left\{ -\infty, -\sqrt{1 + \frac{1}{y}} \right\} \cup \{-1, 1\} \left\{ \sqrt{1 + \frac{1}{y}}, \infty \right\}$$

$$(d) \quad I_x = \left\{ -\infty, -\sqrt{1 + \frac{1}{y}} \right\} \cup \{-1, 1\} \left\{ \sqrt{1 + \frac{1}{y}}, \infty \right\}$$

$$(e) \quad I_x = \left\{ -\infty, -\sqrt{1 + \frac{1}{y}} \right\} \cup \{-1, 1\} \left\{ \sqrt{1 + \frac{1}{y}}, \infty \right\}$$

6. The distribution functions in the various regions are

$$\begin{aligned} (a) \quad y \leq -1 : F_Y(y) &= F_X\left(-\sqrt{1 + \frac{1}{y}}\right) - F_X(-1) + F_X(1) - F_X\left(\sqrt{1 + \frac{1}{y}}\right) \\ &= -\frac{2}{9}\sqrt{1 + \frac{1}{y}} + \frac{1}{3} - \left(-\frac{2}{9} + \frac{1}{3}\right) + \left(\frac{2}{9} + \frac{1}{3}\right) \\ &\quad - \left(\frac{2}{9}\sqrt{1 + \frac{1}{y}} + \frac{1}{3}\right) \\ &= \frac{4}{9} - \frac{4}{9}\sqrt{1 + \frac{1}{y}} \end{aligned}$$

$$(b) \quad -1 < y \leq 0 : F_Y(y) = F_X(1) + F_X(-1) = \left(\frac{2}{9} + \frac{1}{3}\right) - \left(-\frac{2}{9} + \frac{1}{3}\right) = \frac{4}{9}$$

$$\begin{aligned} (c) \quad 0 < y \leq \frac{1}{8} : F_Y(y) &= F_X\left(-\sqrt{1 + \frac{1}{y}}\right) - F_X(-\infty) + F_X(1) - F_X(-1) \\ &\quad + F_X(\infty) - F_X\left(\sqrt{1 + \frac{1}{y}}\right) \\ &= 0 - 0 + \frac{4}{9} + 1 - 1 = \frac{4}{9} \end{aligned}$$

$$\begin{aligned}
\text{(d)} \quad \frac{1}{8} < y \leq \frac{4}{5} : F_Y(y) &= F_X\left(-\sqrt{1 + \frac{1}{y}}\right) - F_X(-\infty) + F_X(1) - F_X(-1) \\
&\quad + F_X(\infty) - F_X\left(\sqrt{1 + \frac{1}{y}}\right) \\
&= 0 - 0 + \frac{4}{9} + 1 - \left(\frac{2}{9}\sqrt{1 + \frac{1}{y}} + \frac{1}{3}\right) = \frac{10}{9} - \frac{2}{9}\sqrt{1 + \frac{1}{y}} \\
\text{(e)} \quad y > \frac{4}{5} : F_Y(y) &= F_X\left(-\sqrt{1 + \frac{1}{y}}\right) - F_X(-\infty) + F_X(1) - F_X(-1) \\
&\quad + F_X(\infty) - F_X\left(\sqrt{1 + \frac{1}{y}}\right) \\
&= \left(-\frac{2}{9}\sqrt{1 + \frac{1}{y}} + \frac{1}{3}\right) - 0 + \frac{4}{9} + 1 - \left(\frac{2}{9}\sqrt{1 + \frac{1}{y}} + \frac{1}{3}\right) \\
&= \frac{13}{9} - \frac{4}{9}\sqrt{1 + \frac{1}{y}}
\end{aligned}$$

The density function  $f_Y(y)$  can be obtained by differentiating  $F_Y(y)$ . Functions  $F_Y(y)$  and  $f_Y(y)$  are shown in the following equation.  $F_Y(y)$  is graphed in Fig. 12.2.14a, and  $f_Y(y)$  is graphed in Fig. 12.2.14b.

$$\begin{aligned}
F_Y(y) &= \begin{cases} \frac{4}{9} - \frac{4}{9}\sqrt{1 + \frac{1}{y}} & \text{if } y \leq -1 \\ \frac{4}{9} & \text{if } -1 < y \leq \frac{1}{8} \\ \frac{10}{9} - \frac{2}{9}\sqrt{1 + \frac{1}{y}} & \text{if } \frac{1}{8} < y \leq \frac{4}{5} \\ \frac{13}{9} - \frac{4}{9}\sqrt{1 + \frac{1}{y}} & \text{if } y > \frac{4}{5} \end{cases} ; \\
f_Y(y) &= \begin{cases} \frac{2}{9} \frac{1}{y^2 \sqrt{1 + \frac{1}{y}}} & \text{if } y \leq -1 \\ 0 & \text{if } -1 < y \leq \frac{1}{8} \\ \frac{1}{9} \frac{1}{y^2 \sqrt{1 + \frac{1}{y}}} & \text{if } \frac{1}{8} < y \leq \frac{4}{5} \\ \frac{2}{9} \frac{1}{y^2 \sqrt{1 + \frac{1}{y}}} & \text{if } y > \frac{4}{5} \end{cases}
\end{aligned}$$

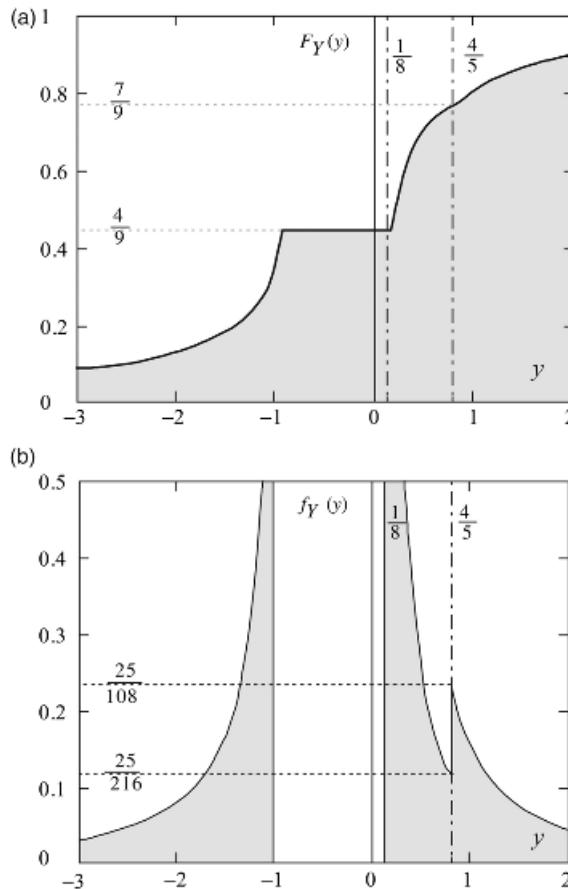


FIGURE 12.2.14

**Example 12.2.9** In Example 12.2.8 we will find the conditional probability of  $Y$  conditioned on the event  $A = |X| > 1$ . Since  $|X| > 1$ , the part of the function  $g(x)$  that lies between  $-1$  and  $1$  has no effect on the conditional distribution. Thus,  $g(x)$  is given by

$$g(x) = \frac{1}{x^2 - 1} \quad \text{for } |x| < 1$$

The conditional density function  $f_X(x | A)$  with  $A = |X| > 1$  is given by

$$\begin{aligned} f_X(x | A) &= \frac{f_X(x)}{p(|X| > 1)} = \frac{f_X(x)}{F_X(-1) + 1 - F_X(1)} \\ &= \frac{f_X(x)}{\left(-\frac{2}{9} + \frac{1}{3}\right) + 1 - \left(\frac{2}{9} + \frac{1}{3}\right)} = \frac{9}{5} f_X(x) \quad \text{for } |x| > 1 \end{aligned}$$

Hence, the conditional density  $f_X(x|A)$  and the conditional distribution  $F_X(x|A)$  can be written as

$$f_X(x|A) = \begin{cases} \frac{2}{5} & \text{if } -1.5 < x \leq -1 \\ \frac{2}{5} & \text{if } 1 < x \leq 3 \\ 0 & \text{otherwise} \end{cases};$$

$$F_X(x|A) = \begin{cases} 0 & \text{if } x < -1.5 \\ \frac{2x+3}{5} & \text{if } -1.5 < x \leq -1 \\ \frac{1}{5} & \text{if } -1 < x \leq 1 \\ \frac{2x-1}{5} & \text{if } 1 < x \leq 3 \\ 1 & \text{if } x > 3 \end{cases}$$

The function  $g(x|A)$  is shown in Fig. 12.2.15.  $f(x|A)$  and  $F_X(x|A)$  are shown in Fig. 12.2.16.

Using exactly the same procedure as in Example 12.3.9, we can find the conditional distribution  $F_Y(y|A)$  and the conditional density  $f_Y(y|A)$  for the regions (a)  $y \leq \frac{1}{8}$ , (b)  $\frac{1}{8} < y \leq \frac{4}{5}$ , and (c)  $y > \frac{4}{5}$ . The results are shown below:

$$F_Y(y|A) = \begin{cases} 0 & \text{if } y \leq \frac{1}{8} \\ \frac{2}{5} \left( 3 - \sqrt{1 + \frac{1}{y}} \right) & \text{if } \frac{1}{8} < y \leq \frac{4}{5} \\ \frac{1}{5} \left( 9 - 4 \sqrt{1 + \frac{1}{y}} \right) & \text{if } y > \frac{4}{5} \end{cases}$$

$$f_Y(y|A) = \begin{cases} 0 & \text{if } y \leq \frac{1}{8} \\ \frac{1}{5} \left( \frac{1}{y^2 \sqrt{1 + \frac{1}{y}}} \right) & \text{if } \frac{1}{8} < y \leq \frac{4}{5} \\ \frac{2}{5} \left( \frac{1}{y^2 \sqrt{1 + \frac{1}{y}}} \right) & \text{if } y > \frac{4}{5} \end{cases}$$

$F_Y(y|A)$  and  $f_Y(y|A)$  are also graphed in Fig. 12.2.17.

**Example 12.2.10** In the previous examples  $g(x)$  was explicitly given, and we had to find  $F_Y(y)$  given  $f_X(x)$ . In this example  $g(x)$  is not explicitly given, but we will have to determine it from the problem. Points are uniformly distributed along the circumference of a unit circle. From each point a tangent is drawn that intersects the  $x$  axis (Fig. 12.2.18). We have to find the distribution of the points of intersection along the  $x$  axis. Since the

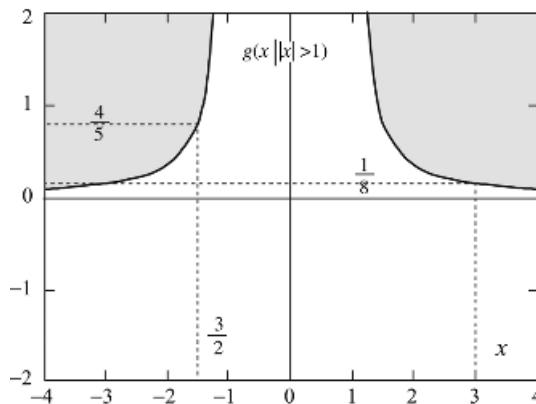


FIGURE 12.2.15

points are uniformly distributed, the angle  $\Theta$  that the point subtends at the center of the circle is a random variable that is uniformly distributed as  $U(0, 2\pi)$ .

The transformation mapping  $X = g(\Theta)$  between  $\Theta$  and  $X$  is found as follows. Since the radius of the circle is 1, the cosine of the angle  $\theta$  subtended by any point on the unit circle at the center of the circle is given by  $\cos(\theta) = (1/x)$ . Hence, the transformation between the random variable  $\Theta$  and the random variable  $X$  is given by

$$X = g(\Theta) = \sec(\Theta)$$

The function  $g(\theta)$  is graphed in Fig. 12.2.19.

The density function  $f_\Theta(\theta)$  and the distribution function  $F_\Theta(\theta)$  of the random variable  $\Theta$  are given by

$$f_\Theta(\theta) = \begin{cases} \frac{1}{2\pi} & \text{for } 0 < \theta \leq 2\pi \\ 0 & \text{otherwise} \end{cases}; \quad F_\Theta(\theta) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{\theta}{2\pi} & \text{if } 0 < \theta \leq 2\pi \\ 1 & \text{if } x > 2\pi \end{cases}$$

They are shown in Fig. 12.2.20.

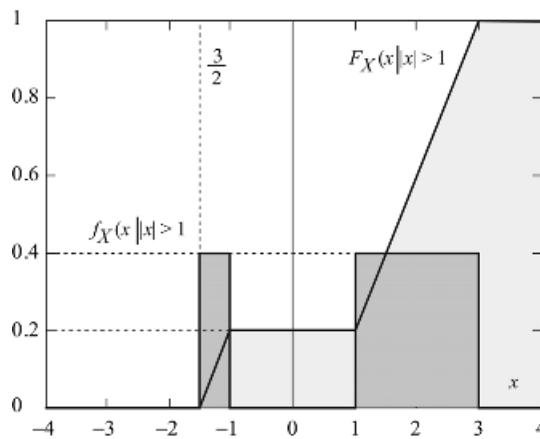


FIGURE 12.2.16

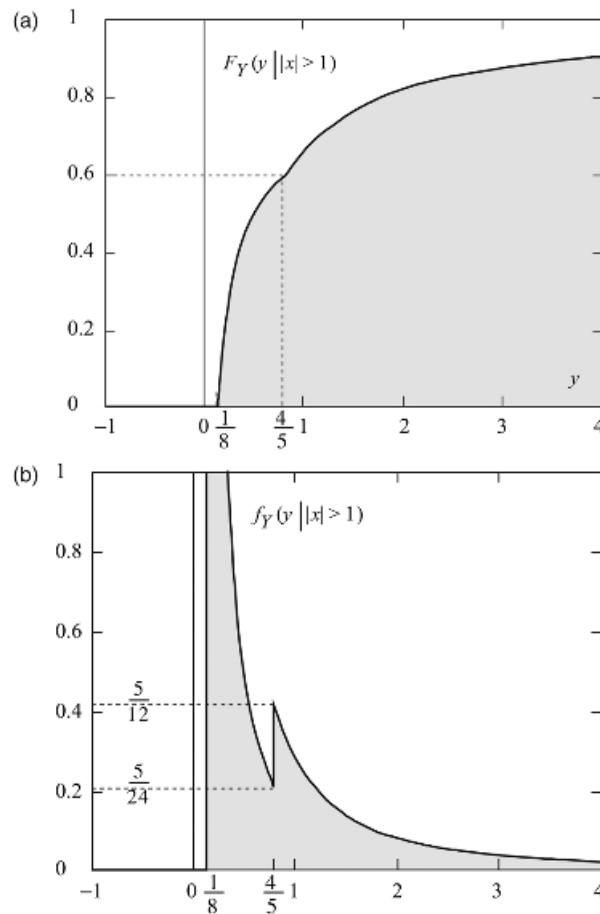


FIGURE 12.2.17

The usual procedure can now be followed to determine the distribution function  $F_X(x)$ .

1. The function  $g(\theta)$  has been graphed.
2. The line  $x$  has been drawn in Fig. 12.2.19.

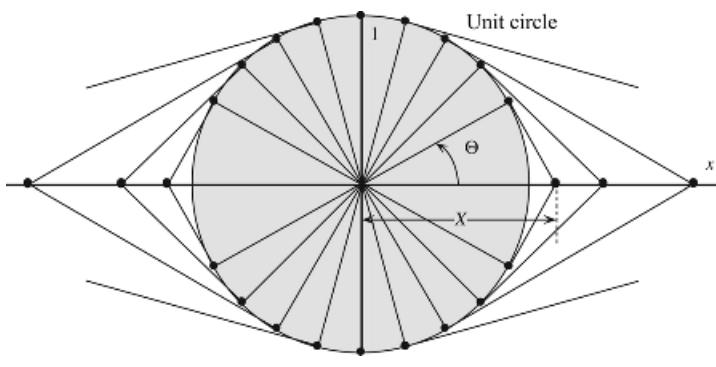


FIGURE 12.2.18

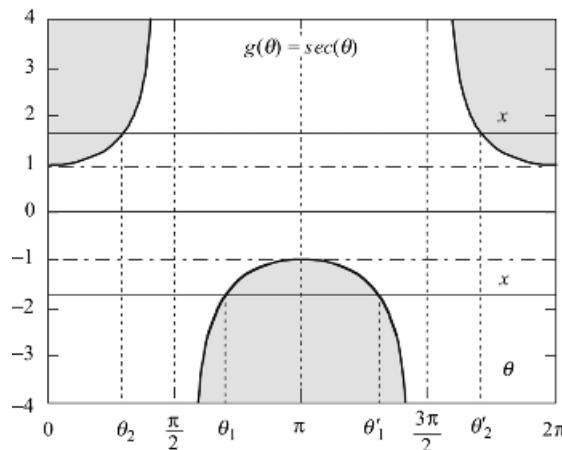


FIGURE 12.2.19

3. The regions along the  $x$  axis are given by (a)  $x \leq -1$ , two points of intersection; (b)  $-1 < x \leq 1$ , no points of intersection; (c)  $x > 1$ , two points of intersection.
4. Intersection points of  $x$  with  $g(\theta)$  are (a)  $\theta_1 = \cos^{-1}(1/x)$  and  $\theta'_1 = 2\pi - \cos^{-1}(1/x)$ , (b) no points of intersection, (c)  $\theta_2 = \cos^{-1}(1/x)$  and  $\theta'_2 = 2\pi - \cos^{-1}(1/x)$ .
5. The corresponding regions  $I_\theta$  are

- (a)  $I_\theta = \left\{ \frac{\pi}{2}, \cos^{-1}\left(\frac{1}{x}\right) \right\} \cup \left\{ 2\pi - \cos^{-1}\left(\frac{1}{x}\right), \frac{3\pi}{2} \right\}$
- (b)  $I_\theta = \left\{ \frac{\pi}{2}, \frac{3\pi}{2} \right\}$
- (c)  $I_\theta = \left\{ 0, \cos^{-1}\left(\frac{1}{x}\right) \right\} \cup \left\{ \frac{\pi}{2}, \frac{3\pi}{2} \right\} \cup \left\{ 2\pi - \cos^{-1}\left(\frac{1}{x}\right), 2\pi \right\}$

6. The distribution functions in the various regions are

- (a)  $F_X(x) = \frac{1}{2\pi} \cos^{-1} \frac{1}{x} - \frac{1}{4} + \frac{3}{4} - \left( 1 - \frac{1}{2\pi} \cos^{-1} \frac{1}{x} \right) = \frac{1}{\pi} \cos^{-1} \frac{1}{x} - \frac{1}{2}$
- (b)  $F_X(x) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$

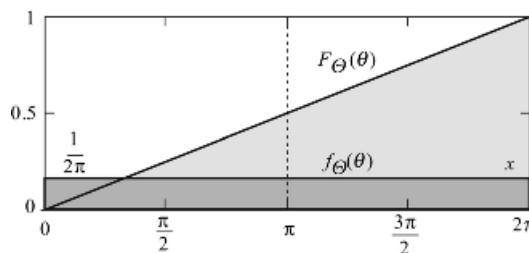


FIGURE 12.2.20

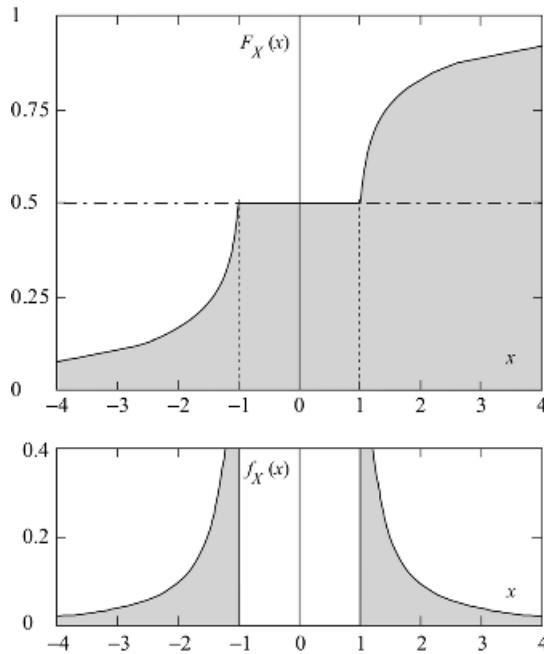


FIGURE 12.2.21

$$(c) F_X(x) = \frac{1}{2\pi} \cos^{-1} \frac{1}{x} + \frac{1}{2} + 1 - \left(1 - \frac{1}{2\pi} \cos^{-1} \frac{1}{x}\right) = \frac{1}{\pi} \cos^{-1} \frac{1}{x} + \frac{1}{2}$$

This distribution is called the *arcsine law*. The density function  $f_X(x)$  is obtained by differentiating the distribution function  $F_X(x)$ . They are both shown below, and the graphs are shown in Fig. 12.2.21:

$$F_X(x) = \begin{cases} \frac{1}{\pi} \cos^{-1} \frac{1}{x} - \frac{1}{2} & \text{if } x \leq -1 \\ \frac{1}{2} & \text{if } -1 < x \leq 1 \\ \frac{1}{\pi} \cos^{-1} \frac{1}{x} + \frac{1}{2} & \text{if } x > 1 \end{cases};$$

$$f_X(x) = \begin{cases} \frac{1}{\pi x \sqrt{x^2 - 1}} & \text{if } x \leq -1 \\ 0 & \text{if } -1 < x \leq 1 \\ \frac{1}{\pi x \sqrt{x^2 - 1}} & \text{if } x > 1 \end{cases}$$

**Example 12.2.11** As a final example, we will take the case when  $g(\theta) = \cos(\theta)$  in the region  $-\pi < \theta \leq \pi$ . In the previous example  $g(x) = \sec(\theta)$ , and here  $g(x) = \cos(\theta)$ . We will also assume as in the previous example that  $f_\theta(\theta)$  is uniformly distributed in

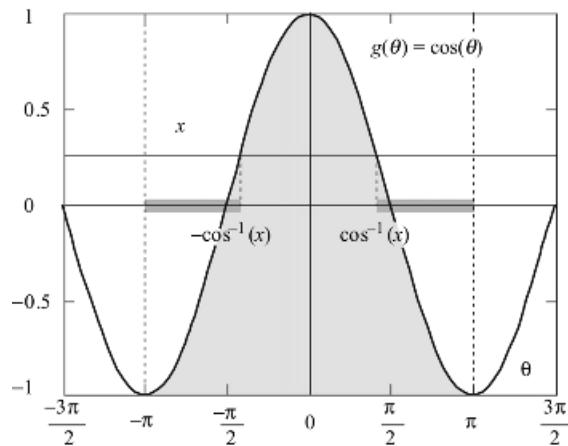


FIGURE 12.2.22

$(-\pi, \pi]$ . Thus,  $f_\Theta(\theta)$  and  $F_\Theta(\theta)$  are given by

$$f_\Theta(\theta) = \begin{cases} \frac{1}{2\pi} & \text{for } -\pi < \theta \leq \pi \\ 0 & \text{otherwise} \end{cases}; \quad F_\Theta(\theta) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{\theta}{2\pi} + \frac{1}{2} & \text{if } -\pi < \theta \leq \pi \\ 1 & \text{if } x > \pi \end{cases}$$

The function  $g(\theta) = \cos(\theta)$  is shown in Fig. 12.2.22. The density and distribution functions are graphed in Fig. 12.2.23.

Following the usual procedure

1.  $g(\theta)$  has been graphed.
2. The line  $x$  has been drawn (Fig. 12.2.22).
3. The regions along the  $x$  axis are given by (a)  $x \leq -1$ , no points of intersection; (b)  $-1 < x \leq 1$ , two points of intersection; (c)  $x > 1$ , no points of intersection.
4. Intersection points of  $x$  with  $g(\theta)$  are (a) no points of intersection, (b)  $\theta_1 = -\cos^{-1}(x)$  and  $\theta_2 = \cos^{-1}(x)$ , (c) no points of intersection.
5. The corresponding regions  $I_\theta$  are (a)  $I_\theta = \{\emptyset\}$ , (b)  $I_\theta = \{-\pi, -\cos^{-1}(x)\} \cup \{\cos^{-1}(x), \pi\}$ , (c)  $I_\theta = \emptyset$ .

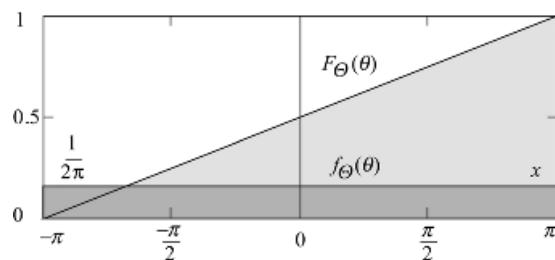


FIGURE 12.2.23

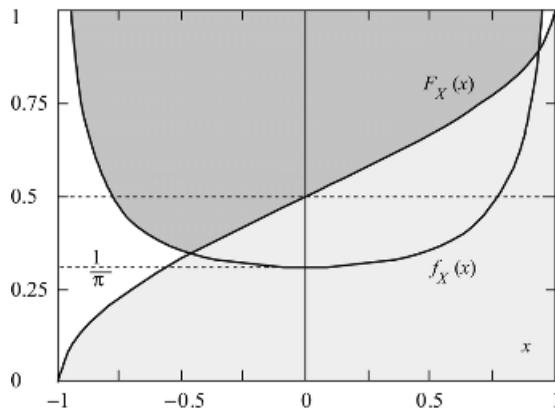


FIGURE 12.2.24

6. The distribution functions in the various regions are

$$(a) F_X(x) = 0$$

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

$$(c) F_X(x) = 1$$

Here also the distribution is the arcsine law.

The density function  $f_X(x)$  is obtained by differentiating the distribution function  $F_X(x)$ . The distribution and density functions are given below and graphed in Fig. 12.2.24.

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq -1 \\ 1 - \frac{1}{\pi} \cos(x) & \text{if } -1 < x \leq 1 \\ 1 & \text{if } x > 1 \end{cases};$$

$$f_X(x) = \begin{cases} 0 & \text{if } x \leq -1 \\ \frac{1}{\pi \sqrt{1-x^2}} & \text{if } -1 < x \leq 1 \\ 0 & \text{if } x > 1 \end{cases}$$

## 12.3 DIRECT DETERMINATION OF DENSITY $f_Y(y)$ FROM $f_X(x)$

In the previous section, for any given  $Y = g(X)$ , we determined  $f_Y(y)$  from  $f_X(x)$  through a circuitous route by converting  $f_X(x)$  to  $F_X(x)$ , using the transformation  $g(X)$  to find  $F_Y(y)$ , and then differentiating  $F_Y(y)$  to obtain  $f_Y(y)$  as shown in the schematic representation in Fig. 12.3.1.

Consider the function  $g(x)$  given in Fig. 12.3.2. The probability of the random variable  $Y$  lying between  $y$  and  $y + dy$  is given by

$$f_Y(y)dy = P\{y < Y \leq y + dy\} \quad (12.3.1)$$

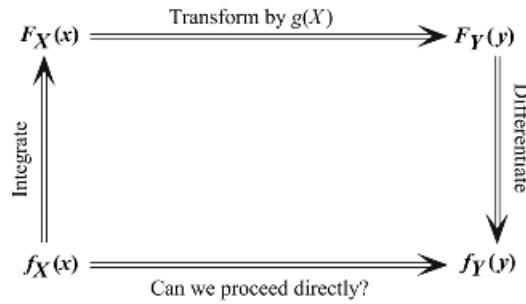


FIGURE 12.3.1

In Fig. 12.3.2  $y = g(x)$  has three real solutions:  $x_1, x_2, x_3$ . Hence the set of values for which  $\{y < g(x) \leq y + dy\}$  is given by the union of the set of values corresponding to the three points of intersection,  $\{x_1 < X \leq x_1 + dx_1\} \cup \{x_2 + dx_2 < X \leq x_2\} \cup \{x_3 < X \leq x_3 + dx_3\}$ , where  $dx_1 > 0$ ,  $dx_3 > 0$ , and  $dx_2 < 0$ . By the conservation of probability measure, we have

$$\begin{aligned}
 f_Y(y)dy &= P\{y < Y \leq y + dy\} \\
 &= P\{x_1 < X \leq x_1 + dx_1\} + P\{x_2 + dx_2 < X \leq x_2\} + P\{x_3 < X \leq x_3 + dx_3\} \\
 &= f_X(x_1)dx_1 + f_X(x_2)|dx_2| + f_X(x_3)dx_3
 \end{aligned} \tag{12.3.2}$$

We can now write

$$f_Y(y) = f_X(x_1) \frac{dx_1}{dy} + f_X(x_2) \frac{|dx_2|}{dy} + f_X(x_3) \frac{dx_3}{dy} \tag{12.3.3}$$

But

$$\frac{dx_1}{dy} = \frac{1}{|g'(x_1)|}, \quad \frac{|dx_2|}{dy} = \frac{1}{|g'(x_2)|}, \quad \frac{dx_3}{dy} = \frac{1}{|g'(x_3)|} \tag{12.3.4}$$

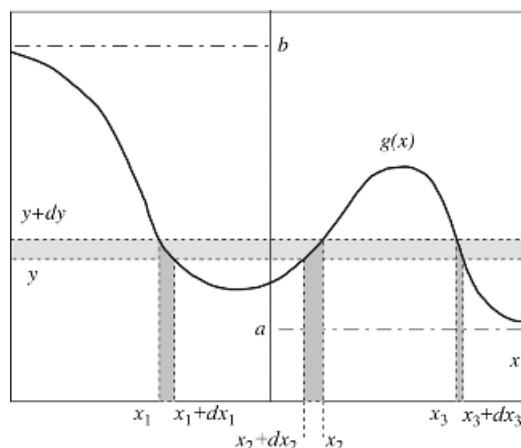


FIGURE 12.3.2

Substituting Eq. (12.3.4) into Eq. (12.3.3), we obtain the result for  $f_Y(y)$  directly from  $f_X(x)$ :

$$f_Y(y) = \frac{f_X(x_1)}{|g'(x_1)|} + \frac{f_X(x_2)}{|g'(x_2)|} + \frac{f_X(x_3)}{|g'(x_3)|} \quad (12.3.5)$$

Equation (12.3.5) can be generalized to the case when there are  $n$  real solutions to the equation  $y = g(x)$ , given by  $\{x_1, x_2, \dots, x_n\}$ :

$$f_Y(y) = \sum_{i=1}^n \frac{f_X(x_i)}{|g'(x_i)|} \quad (12.3.6)$$

From the derivation it is clear that certain restrictions have to be satisfied if we have to determine  $f_Y(y)$  directly from  $f_X(X)$ .

### Conditions for Direct Determination of $f_X(X)$

1. The derivative of  $g(x)$  must exist.
2. There can be no straight-line solutions to  $g(x) = y$ .

We now formulate the following steps for the direct determination of  $f_Y(y)$  from  $f_X(x)$ :

### Steps for Direct Determination of $f_Y(y)$

1. Check whether  $g(x)$  satisfies restrictions 1 and 2.
2. Solve for  $y = g(x)$ . Let the  $n$  real roots be  $\{x_1, x_2, \dots, x_n\}$ .
3. Determine

$$\left| \frac{dg(x)}{dx} \right|_{x=x_i}, \quad i = 1, 2, \dots, n$$

4. Calculate

$$f_Y(y) = \sum_{i=1}^n \frac{f_X(x_i)}{|g'(x_i)|}$$

Thus, if the restrictions on  $g(x)$  are satisfied, it is easier to calculate  $f_Y(y)$  directly from  $f_X(x)$ . We will take some of the examples in the previous section and solve for  $f_X(x)$  directly.

**Example 12.3.1** We will find  $f_Y(y)$  for  $Y = aX + b$  with  $a > 0$  or,  $a \leq 0$ .

1.  $g(x)$  satisfies the restrictions.
2. We solve for  $y = ax + b$ . The solution yields  $x = (y - b)/a$ .

$$\begin{aligned} 3. \quad & \left| \frac{dg(x)}{dx} \right| = \left| \frac{d(ax + b)}{dx} \right| = |a| \\ 4. \quad & f_Y(y) = \frac{1}{|a|} f_X \left( \frac{y - b}{a} \right) \end{aligned} \quad (12.3.7)$$

**Example 12.3.2**  $Y = g(X) = (X - a)^2$ :

1.  $g(x)$  satisfies the restrictions.
2. Solving for  $x$  in  $y = (x - a)^2$ , we obtain two roots:  $x_1 = a - \sqrt{y}$ ,  $x_2 = a + \sqrt{y}$

3.  $\left| \frac{dg(x)}{dx} \right| = \left| \frac{d(x-a)^2}{dx} \right| = |2(x-a)| = 2\sqrt{y}$
4.  $f_Y(y) = \frac{1}{2\sqrt{y}} [f_X(a - \sqrt{y}) + f_X(a + \sqrt{y})]; \quad y > 0 \quad (12.3.8)$

**Example 12.3.3**  $Y = g(X) = 1/(X - a)^2$ .

1.  $g(x)$  satisfies the restrictions.
2. Solving for  $x$  in  $y = 1/(x - a)^2$ , we obtain two roots  $x_1 = a - (1/\sqrt{y})$ ,  $x_2 = a + (1/\sqrt{y})$ .
3.  $\left| \frac{dg(x)}{dx} \right| = \left| \frac{d}{dx} \frac{1}{(x-a)^2} \right| = \left| -2 \frac{1}{(x-a)^3} \right| = |2y^{3/2}|$
4.  $f_Y(y) = \frac{1}{2y^{3/2}} \left[ f_X\left(a - \frac{1}{\sqrt{y}}\right) + f_X\left(a + \frac{1}{\sqrt{y}}\right) \right]; \quad y > 0 \quad (12.3.9)$

**Example 12.3.4 (See Example 12.2.8)** The probability density functions  $f_X(x)$  and  $g(x)$  and are given by

$$f_X(x) = \begin{cases} \frac{2}{9} & \text{if } -1.5 < x \leq 3 \\ 0 & \text{otherwise} \end{cases} \quad g(x) = \frac{1}{x^2 - 1} \quad \text{for all } x$$

Even though  $g(x)$  is defined from  $-\infty$  to  $\infty$ , the range of  $f_X(x)$  plays a critical role in determining  $f_Y(y)$ . Figure 12.2.12 is redrawn in Fig. 12.3.3 to indicate the ranges.

1.  $g(x)$  satisfies the restrictions.
2. Solving for  $x$  in  $y = 1/(x^2 - 1)$ , we obtain two roots:  $x_1 = -\sqrt{1 + (1/y)}$ ,  $x_2 = \sqrt{1 + (1/y)}$ .
3.  $\left| \frac{dg(x)}{dx} \right| = \left| \frac{d}{dx} \frac{1}{(x^2 - 1)} \right| = \left| -2 \frac{x}{(x^2 - 1)^2} \right| = \left| 2y^2 \sqrt{1 + \frac{1}{y}} \right|$
4.  $f_Y(y)$  have to be calculated in the four ranges of  $y$  as shown in Fig. 12.3.3.

The density function  $f_Y(y)$  is calculated as follows:

Region I:  $y \leq -1$ , two points of intersection:

$$f_Y(y) = \frac{1}{2y^2 \sqrt{1 + \frac{1}{y}}} \left( \frac{2}{9} + \frac{2}{9} \right) = \frac{2}{9} \frac{1}{y^2 \sqrt{1 + \frac{1}{y}}}$$

Region II:  $-1 < y \leq \frac{1}{8}$ , no points of intersection:

$$f_Y(y) = 0$$

Region III:  $\frac{1}{8} < y \leq \frac{4}{5}$ , one point of intersection:

$$f_Y(y) = \frac{1}{2y^2 \sqrt{1 + \frac{1}{y}}} \left( \frac{2}{9} \right) = \frac{1}{9} \frac{1}{y^2 \sqrt{1 + \frac{1}{y}}}$$

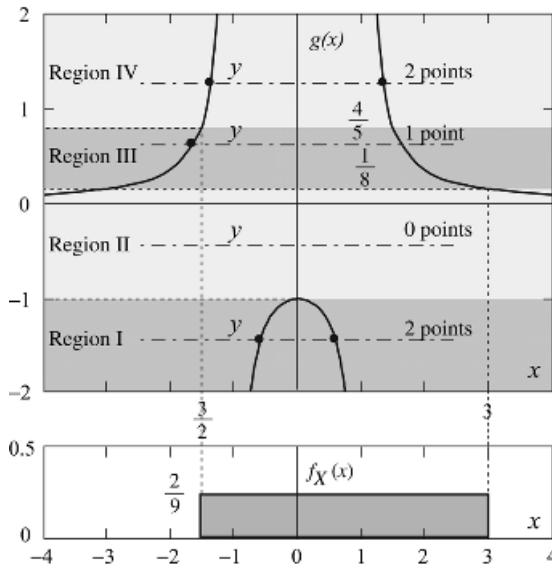


FIGURE 12.3.3

Region IV:  $y > \frac{4}{5}$ , two points of intersection:

$$f_Y(y) = \frac{1}{2y^2 \sqrt{1 + \frac{1}{y}}} \left( \frac{2}{9} + \frac{2}{9} \right) = \frac{2}{9} \frac{1}{y^2 \sqrt{1 + \frac{1}{y}}}$$

Here we have found the result far more easily than in Example 12.2.8.

**Example 12.3.5 (See Example 12.2.10)** In Example 12.2.10, the function  $g(\theta)$  is given by  $g(\theta) = \sec(\theta)$  and  $f_{\Theta}(\theta)$  is given by

$$f_{\Theta}(\theta) = \begin{cases} \frac{1}{2\pi} & \text{for } 0 < \theta \leq 2\pi \\ 0 & \text{otherwise} \end{cases}$$

We will now find  $f_X(x)$  directly from  $f_{\Theta}(\theta)$ .

1.  $g(\theta)$  satisfies the restrictions.
2. Solving for  $\theta$  in  $x = \sec(\theta)$ , we have  $\theta_1 = \sec^{-1}(x)$  and  $\theta'_1 = 2\pi - \sec^{-1}(x)$ .
3.  $\left| \frac{dg(\theta)}{d\theta} \right| = \left| \frac{d}{d\theta} \sec(\theta) \right| = |\sec(\theta) \tan(\theta)| = |x\sqrt{x^2 - 1}|$
4. From Fig. 12.2.19, the regions of  $x$  are (a)  $x \leq -1$ , (b)  $-1 < x \leq 1$ , (c)  $x > 1$ .

- (a)  $x \leq -1$ , two points of intersection:

$$f_X(x) = \frac{1}{x\sqrt{x^2 - 1}} \left( \frac{1}{2\pi} + \frac{1}{2\pi} \right) = \frac{1}{\pi x \sqrt{x^2 - 1}}$$

- (b)  $-1 < x \leq 1$ , no points of intersection,  $f_X(x) = 0$ .

(c)  $x > 1$ , two points of intersection:

$$f_X(x) = \frac{1}{x\sqrt{x^2 - 1}} \left( \frac{1}{2\pi} + \frac{1}{2\pi} \right) = \frac{1}{\pi x \sqrt{x^2 - 1}}$$

These are the same results obtained in Example 12.2.10.

**Example 12.3.6 (See Example 12.2.11)** Here  $g(\theta) = \cos(\theta)$  and  $f_\Theta(\theta)$  is given by

$$f_\Theta(\theta) = \begin{cases} \frac{1}{2\pi} & \text{for } -\pi < \theta \leq \pi \\ 0 & \text{otherwise} \end{cases}$$

The three regions of  $x$  are (a)  $x \leq -1$ , (b)  $-1 < x \leq 1$ , (c)  $x > 1$ .

1.  $g(\theta)$  satisfies the restrictions.
2. The solutions for  $x = \cos(\theta)$  are  $\theta_1 = -\cos^{-1}(x)$  and  $\theta'_1 = \cos^{-1}(x)$ .
3.  $\left| \frac{dg(\theta)}{d\theta} \right| = \left| \frac{d}{d\theta} \cos(\theta) \right| = | -\sin(\theta) | = \left| \sqrt{1 - x^2} \right|$
4.  $f_X(x)$  can now be found in the three regions.
  - (a)  $x \leq -1$ , no points of intersection:  $f_X(x) = 0$ .
  - (b)  $-1 < x \leq 1$ , two points of intersection:

$$f_X(x) = \frac{1}{\sqrt{1 - x^2}} \left( \frac{1}{2\pi} + \frac{1}{2\pi} \right) = \frac{1}{\pi \sqrt{1 - x^2}} \quad (12.3.10)$$

(c)  $x > 1$ , no points of intersection:  $f_X(x) = 0$ .

These are the same results obtained in Example 12.2.11.

**Example 12.3.7** This example will have relevance in the next section. Here  $g(x)$  and  $f_X(x)$  are given by

$$g(x) = \begin{cases} 0 & \text{if } x \leq 1 \\ 2 + \ln\left(\frac{4}{5-x}\right) & \text{if } 1 < x < 5 \\ 0 & \text{if } x \geq 5 \end{cases} \quad f_X(x) = \begin{cases} \frac{1}{4} & \text{if } 1 < x \leq 5 \\ 0 & \text{otherwise} \end{cases}$$

The functions  $g(x)$  and  $f_X(x)$  are shown in Fig. 12.3.4.

1.  $g(x)$  satisfies the restrictions.
2. There are no points of intersection in the region  $y \leq 2$ . Solving for  $y = g(x)$  in the region  $y > 2$ , we obtain

$$y = 2 + \ln\left(\frac{4}{5-x}\right); y - 2 = \ln\left(\frac{4}{5-x}\right) \quad \text{or} \quad 5 - x = 4e^{-(y-2)}$$

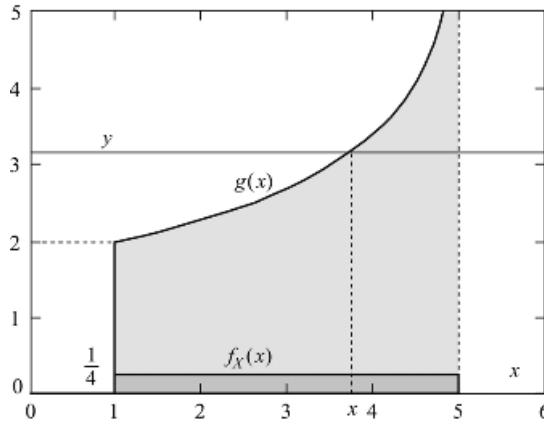


FIGURE 12.3.4

or

$$x = 5 - 4e^{-(y-2)} \quad \text{for } y > 2$$

$$3. \quad \left| \frac{dg(x)}{dx} \right| = \left| \frac{d}{dx} \left[ 2 + \ln \left( \frac{4}{5-x} \right) \right] \right| = \frac{1}{5-x} = 4e^{(y-2)}; \quad y > 2$$

4. Since there are no points of intersection of  $y$  with  $g(x)$  in the region  $y \leq 2$ , it follows that  $f_Y(y) = 0$ . In the region  $y > 2$ ,

$$f_Y(y) = \frac{1}{4} 4e^{-(y-2)} = e^{-(y-2)} \quad \text{for } y > 2$$

## 12.4 INVERSE PROBLEM: FINDING $g(x)$ GIVEN $f_X(x)$ AND $f_Y(y)$

In the previous sections we studied finding  $f_Y(y)$  given  $f_X(x)$  and  $g(x)$ . In this section we will find  $g(x)$  given  $f_X(x)$  and  $f_Y(y)$ . We will use the conservation of probability measure given by an equation similar to Eq. (12.3.3):

$$f_Y(\eta) \Delta \eta = f_X(\xi_1) \Delta \xi_1 + f_X(\xi_2) \Delta \xi_2 + \cdots + f_X(\xi_n) \Delta \xi_n \quad (12.4.1)$$

Taking the limit as  $\Delta \eta, \Delta \xi_1, \Delta \xi_2, \dots, \Delta \xi_n \rightarrow 0$ , all the  $\xi_k$  coalesce to  $\xi$ , and integrating  $\eta$  between  $-\infty$  to  $y$  and  $\xi$  between  $-\infty$  to  $x$ , we obtain

$$\int_{-\infty}^y f_Y(\eta) d\eta = \int_{-\infty}^x f_X(\xi) d\xi \quad (12.4.2)$$

The lefthand side of Eq. (12.4.2) is a function of  $y$ , the righthand side is a function of  $x$ , and the equation can be solved for  $y$ , which yields  $g(x)$  as the transforming function.

**Example 12.4.1** Given  $f_X(x) = 2e^{-2(x-2)}u(x-2)$  and  $f_Y(y) = \frac{1}{4}[u(y-2) - u(y-6)]$ , we have to find the transforming function  $g(x)$ . The functions  $f_X(x)$  and  $f_Y(y)$  are shown in Fig. 12.4.1.

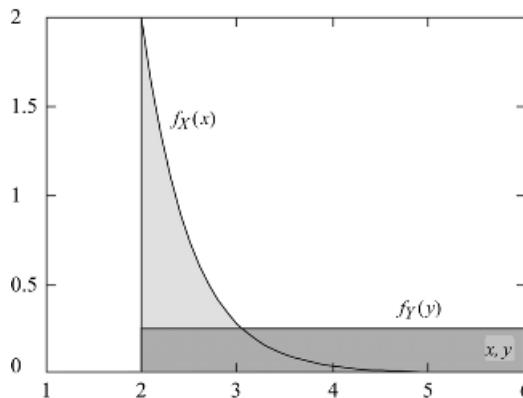


FIGURE 12.4.1

Integrating both  $f_X(x)$  and  $f_Y(y)$ , we obtain the following:

$$\begin{aligned} \int_2^y f_Y(\eta) d\eta &= \int_2^y \frac{1}{4} d\eta = \frac{y-2}{4} && \text{for } 2 < y \leq 6 \\ \int_2^x f_X(\xi) d\xi &= \int_2^x 2e^{-2(\xi-2)} d\xi = 1 - e^{-(x-2)} && \text{for } x > 2 \end{aligned}$$

Thus

$$\frac{y-2}{4} = 1 - e^{-(x-2)} \quad \text{for } 2 < y \leq 6 \quad \text{and } x > 2$$

or

$$y = g(x) = 6 - 4e^{-(x-2)} \quad \text{for } x > 2$$

The function  $g(x)$  is graphed in Fig. 12.4.2. Using the same techniques as in the previous section, we can show that the transformation function  $g(x)$  indeed gives a uniform distribution in  $(2,6]$ , given that  $f_X(x)$  is an exponentially distributed function starting at  $y = 2$ .

**Example 12.4.2 (See Example 12.3.7)** We will revisit Example 12.3.7, where  $f_X(x)$  is uniformly distributed in the interval  $(1,5]$  and  $f_Y(y)$  is exponentially distributed as

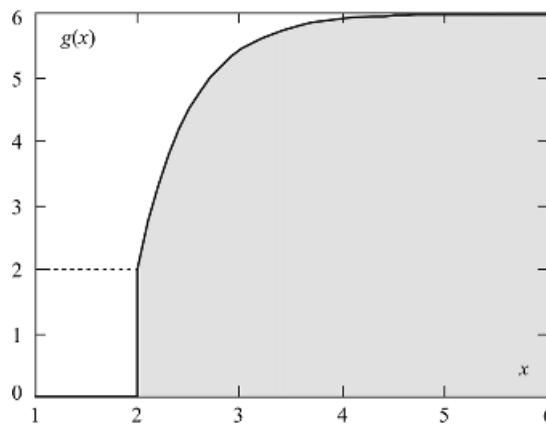


FIGURE 12.4.2

$e^{-(y-2)}$  starting at  $y = 2$ . We have to find the transforming functions. Proceeding as in the previous example, we have

$$\int_2^y e^{-(\eta-2)} d\eta = \int_1^x \frac{1}{4} d\xi \quad \text{or} \quad 1 - e^{-(y-2)} = \frac{x-1}{4}, \quad y > 2, x > 1$$

or

$$e^{-(y-2)} = \frac{5-x}{4}, \quad y > 2, x > 1$$

Taking logarithm on both sides, we obtain

$$-(y-2) = \ln\left(\frac{5-x}{4}\right) \quad \text{or} \quad y = g(x) = 2 + \ln\left(\frac{4}{5-x}\right), \quad x > 1$$

$g(x)$  is exactly the same function that we used in Example 12.3.7 and is shown in Fig. 12.3.4.

## 12.5 MOMENTS OF A FUNCTION OF A RANDOM VARIABLE

### Expectation

We will now find the expectation of the random variable  $Y = g(X)$ . We have already found the density function  $f_Y(y)$ . We can use the usual definition of expectations and write

$$\mu_Y = E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy \quad (12.5.1)$$

However, if only the expected value of  $Y$ , and not the density function  $f_Y(y)$  is required, we can find the result from the knowledge of  $f_X(x)$  and  $g(x)$ . Using the law of conservation of probability measure [Eq. (12.3.2)], we have

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_{-\infty}^{\infty} y f_X(x) dx \quad (12.5.2)$$

Substituting  $y = g(x)$  in Eq. (12.5.2), we obtain the equation for  $E[Y]$  in terms of the density function  $f_X(x)$ :

$$E[Y] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad (12.5.3)$$

If the random variable is of the discrete type, then

$$E[Y] = \sum_i y_i P(y = y_i) = \sum_i g(x_i) P(x = x_i) \quad (12.5.4)$$

### Variance

In a similar manner, the variance of  $Y = g(X)$  can be defined as

$$\begin{aligned} \sigma_Y^2 &= E[Y - \mu_Y]^2 = \int_{-\infty}^{\infty} (y - \mu_Y)^2 f_Y(y) dy \\ &= \int_{-\infty}^{\infty} (y - \mu_Y)^2 f_X(x) dx = \int_{-\infty}^{\infty} [g(x) - \mu_Y]^2 f_X(x) dx \end{aligned} \quad (12.5.5)$$

### Higher-Order Moments

The  $r$ th moment of  $Y = g(X)$  can be defined as follows:

$$\begin{aligned} m_r &= E[Y]^r = \int_{-\infty}^{\infty} (y)^r f_Y(y) dy \\ &= \int_{-\infty}^{\infty} (y)^r f_X(x) dx = \int_{-\infty}^{\infty} [g(x)]^r f_X(x) dx \end{aligned} \quad (12.5.6)$$

**Example 12.5.1** We will take Example 12.2.7 and calculate the expected values by using the density functions  $f_Y(y)$  and  $f_X(x)$ . Function  $f_X(x)$  is uniformly distributed in  $(-2, 2)$  as shown in Fig. 12.2.10. The density function  $f_Y(y)$  for that example is given by

$$f_Y(y) = \frac{1}{2}\delta(y+1) + \frac{3}{8}\left[u\left(y - \frac{4}{3}\right) - u(y)\right]$$

and  $g(x)$ ,  $f_X(x)$ , and  $f_Y(y)$  are shown in Fig. 12.5.1.

Using Eq. (12.5.1) to obtain  $E[Y]$ , we have

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} y \left\{ \frac{1}{2}\delta(y+1) + \frac{3}{8}\left[u\left(y - \frac{4}{3}\right) - u(y)\right] \right\} dy \\ &= -\frac{1}{2} + \int_0^{4/3} \frac{3}{8}y \, dy = -\frac{1}{2} + \frac{3}{8} \cdot \frac{16}{9.2} = -\frac{1}{6} \end{aligned}$$

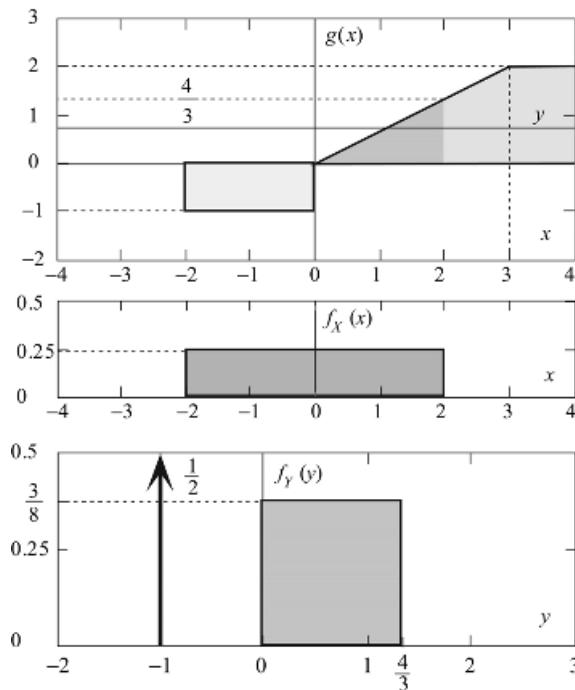


FIGURE 12.5.1

Using Eq. (12.5.3) to obtain  $E[Y]$ , we have

$$\begin{aligned} E[Y] &= \frac{1}{4} \left[ \int_{-2}^0 -1 \, dx + \int_0^2 \frac{2}{3} x \, dx \right] \\ &= -\frac{2}{4} + \frac{1}{4} \cdot \frac{2}{3} \cdot \frac{4}{2} = -\frac{1}{2} + \frac{1}{3} = -\frac{1}{6} \end{aligned}$$

This result could just as well have been obtained by inspection of  $g(x)$  by adding the areas of the rectangle  $(-2)$  and the triangle  $\frac{4}{3}$  and dividing them by  $\frac{1}{4}$ , yielding  $-\frac{1}{6}$ . All three of these methods yield exactly the same result.

We will now compute the variance of  $Y$  from  $f_Y(y)$  using Eq. (12.5.5):

$$\begin{aligned} \text{var}[Y] &= \int_{-\infty}^{\infty} \left( y + \frac{1}{6} \right)^2 \left\{ \frac{1}{2} \delta(y+1) + \frac{3}{8} \left[ u\left(y - \frac{4}{3}\right) - u(y) \right] \right\} \, dy \\ &= \frac{1}{2} \left( -\frac{5}{6} \right)^2 + \int_0^{4/3} \frac{3}{8} \left( y + \frac{1}{6} \right)^2 \, dy = +\frac{25}{72} + \frac{91}{216} = \frac{83}{108} \end{aligned}$$

Computing  $\text{var}[Y]$  from  $f_X(x)$  and using  $\text{var}[Y] = E[Y]^2 - \mu_Y^2$ , we have

$$\begin{aligned} \text{var}[Y] &= \frac{1}{4} \int_{-2}^0 (-1)^2 \, dx + \frac{1}{4} \int_0^2 \left( \frac{2}{3} x \right)^2 \, dx - \frac{1}{36} \\ &= \frac{2}{4} + \frac{1}{4} \cdot \frac{32}{27} - \frac{1}{36} = \frac{83}{108} \end{aligned}$$

and the two results are the same.

**Example 12.5.2** We are given  $f_X(x)$  and  $f_Y(y)$ , and we have to find  $g(x)$  and the expectation and variance of  $Y$  using  $f_Y(y)$  and  $f_X(x)$ :

$$f_X(x) = \begin{cases} \frac{1}{4} & 2 < x \leq 6 \\ 0 & \text{otherwise} \end{cases}; \quad f_Y(y) = \begin{cases} \frac{3y^2}{8} & 0 < y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Integrating  $f_X(\xi)$  from 2 to  $x$  and  $f_Y(\eta)$  from 0 to  $y$ , we have

$$\begin{aligned} \int_2^x \frac{1}{4} \, d\xi &= \int_0^y \frac{3\eta^2}{8} \, d\eta \\ \frac{1}{4}(x-2) &= \frac{y^3}{8}; \quad 2 < x \leq 6, \quad 0 < y \leq 2 \end{aligned}$$

Hence,  $y = g(x) = [2(x-2)]^{1/3}$ . We also note the following:

Expectation with  $f_Y(y)$ :

$$E[y] = \int_0^2 y \cdot \frac{3y^2}{8} \, dy = \frac{3}{2}$$

Expectation with  $f_X(x)$ :

$$E[y] = \int_2^6 \frac{1}{4} [2(x-2)]^{1/3} \, dx = \frac{2^{1/3}}{4} \cdot 3 \cdot 4^{1/3} = \frac{3}{2}$$

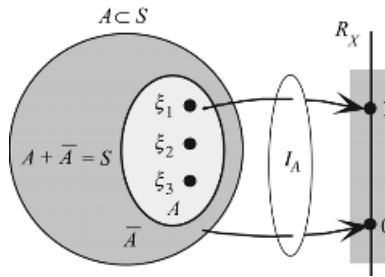


FIGURE 12.5.2

Variance with  $f_Y(y)$ :

$$\text{var}[y] = \int_0^2 \left(y - \frac{3}{2}\right)^2 \cdot \frac{3y^2}{8} dy = \frac{3}{20}$$

Variance with  $f_X(x)$ , using  $\text{var}[Y] = E[Y]^2 - \mu_Y^2$ :

$$\text{var}[y] = \int_2^6 [2(x-2)]^{2/3} \cdot \frac{1}{4} dx - \frac{9}{4} = \frac{3}{5} \cdot 8^{2/3} - \frac{9}{4} = \frac{3}{20}$$

Both methods give the same results.

### Indicator Functions

An indicator function  $I_A(\xi)$  is a real-valued function defined on the sample space  $S$ , having a value 0 or 1 depending on whether the  $\xi$  point is in the event  $A \subset S$ . Or

$$I_A(\xi) = \begin{cases} 1 & \xi \in A \\ 0 & \xi \notin A \end{cases} \quad (12.5.7)$$

The indicator function is a unit step for sets. It is shown in Fig. 12.5.2.

The indicator function is a *zero-one random variable*, and if

$$\begin{aligned} P\{\xi \in A\} &= p \\ P\{\xi \notin A\} &= 1 - p = q \end{aligned} \quad (12.5.8)$$

then the expected value and the variance of the indicator function  $I_A$  are

$$\begin{aligned} E[I_A] &= 1 \cdot p + 0 \cdot q = p \\ \text{var}[I_A] &= 1^2 \cdot p - p^2 = pq \end{aligned} \quad (12.5.9)$$

### Properties of Indicator Functions

1.  $A \subset B \iff I_A \leq I_B$
2.  $I_A = I_A^2 = \dots = I_A^n$
3.  $I_s = 1$ : sample space
4.  $I_{A \cup B} = \max(I_A, I_B)$
5.  $I_{A \cap B} = \min(I_A, I_B)$
6.  $I_{\bar{A}} = 1 - I_A$ :  $\bar{A}$  = complement of  $A$

(12.5.10)

# Functions of Multiple Random Variables

## 13.1 FUNCTION OF TWO RANDOM VARIABLES, $Z = g(X, Y)$

In the previous chapter we discussed a single function of a single random variable. Here we will find the probability  $f_Z(z)$  of a single function  $Z$  of two random variables  $X$  and  $Y$  defined by the joint probability density function  $f_{XY}(x, y)$ . The analysis is very similar to that in the previous chapter.

Following similar analysis given in Chapter 12, we have  $Z = g(X, Y)$  as a function that maps the domain space of  $Z$  given by  $\{C_{xy}, \mathcal{F}_{XY}, P_{XY}\}$  into the range space  $\{R_z, \mathcal{F}_Z, P_Z\}$ , which is a real line, as shown in Fig. 13.1.1

$$\{Z = g(X, Y)\} : \{C_{xy}, \mathcal{F}_{XY}, P_{XY}\} \rightarrow \{R_z, \mathcal{F}_Z, P_Z\} \quad (13.1.1)$$

and the mapping of  $\{X, Y\}$  is from the probability space  $\{S, \mathcal{F}, P\}$  (where,  $S$  is the product space  $S = S_X \times S_Y$ ) into the domain space of  $Z$  given by  $\{C_{xy}, \mathcal{F}_{XY}, P_{XY}\}$

$$X, Y: \{S, \mathcal{F}, P\} \rightarrow \{C_{xy}, \mathcal{F}_{XY}, P_{XY}\} \quad (13.1.2)$$

Alternately, if  $D_{xy} \subset C_{xy}$  and  $I_z \subset R_z$ , then  $I_z$  is the *image* of  $D_{xy}$  under the transformation  $g$  and  $D_{xy}$  is the *inverse image* of  $I_z$  under the transformation  $g^{-1}$ . Hence  $I_z$  is the set of all points  $(x, y)$  belonging to  $D_{xy}$  such that  $g(x, y)$  belongs to  $I_z$ . This is represented by

$$D_{xy} = \{x, y : g(x, y) \in I_z\} \quad (13.1.3)$$

We also have conditions similar to those in the last chapter for  $Z$  to be a random variable.

### Conditions on $Z = g(X, Y)$ to Be a Random Variable

1. The domain of  $g(x, y)$  must include the range of the RV  $Z$ .
2. For each  $z \in I_z$ ,  $\{g(x, y) \leq z\}$  must form a field  $\mathcal{F}_Z$ ; that is, it must consist of the unions and intersections of a countable number of intervals.
3. The probability of the event  $\{g(x, y) = \pm\infty\}$  is 0.

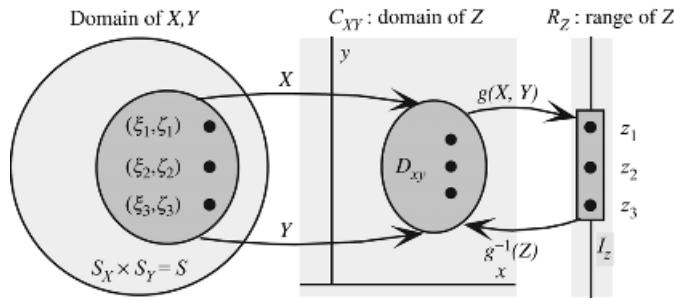


FIGURE 13.1.1

In essence, the transformation  $Z = g(X, Y)$  induces a probability measure  $P_Z$  on  $\mathcal{F}_Z$ , and we have to find this probability measure. From Fig. 13.1.1, we can write

$$P_Z(Z) = P(Z \in I_z) = P_{XY}(g^{-1}(Z) \in D_{xy}) = P_{XY}(X, Y \in D_{xy}) \quad (13.1.4)$$

The set  $I_z$  consists of all the countable unions and intersections of the type  $a < Z \leq b$ . Without loss in generality, we will assume that  $I_z = \{Z \leq z\}$ , and substituting for  $D_{xy}$  from Eq. (13.1.3) in Eq. (13.1.4), we have

$$\begin{aligned} P_Z(Z) &= P_{XY}[(X, Y) \in D_{xy}] = P_{XY}[x, y: g(x, y) \in I_z] \\ &= P_{XY}[x, y: g(x, y) \leq z] = F_{XY}(z) = P(Z \leq z) = F_Z(z) \end{aligned} \quad (13.1.5)$$

Equation (13.1.5) can also be written as

$$P_Z(Z) = P_{XY}[(X, Y) \in D_{xy}] = F_Z(z) = \int_{D_{xy}} \int f_{XY}(\xi, \eta) d\xi d\eta \quad (13.1.6)$$

where  $f_{XY}(x, y)$  is the joint density of the random variables  $X$  and  $Y$ .

In conclusion, we can say that the probability of  $Z \leq z$  is given by the probability of the set of all values of  $x$  and  $y$  such that  $g(x, y) \leq z$ , or by integrating the joint density  $f_{XY}(x, y)$  over the domain  $D_{xy}$ . We will use both these interpretations in solving problems.

### Summary of Steps for Finding $F_Z(z) = P(Z \leq z)$

1. Graph the function  $g(x, y)$  in the  $x$ - $y$  plane.
2. Equate  $g(x, y) = z$ .
3. Find  $D_{xy}$ , the distinct regions in the  $x$ - $y$  plane such that  $g(x, y) \leq z$ .
4. Find  $F_Z(z)$  from Eq. (13.1.5) or (13.1.6).

**Example 13.1.1** Given  $Z = g(X, Y) = \min(X, Y)$ , we will find  $F_Z(z)$  given the joint distribution  $F_{XY}(x, y)$ . We follow the scheme outlined above.

1.  $g(x, y)$  is shown in Fig. 13.1.2.
2.  $g(x, y) = z$  is marked on the diagram.
3. The region  $D_{xy}$  is the set of all points in the  $x$ - $y$  plane such that either  $x \leq z$  or  $y \leq z$ . It is shown by the region ABHCD in Fig. 13.1.2.

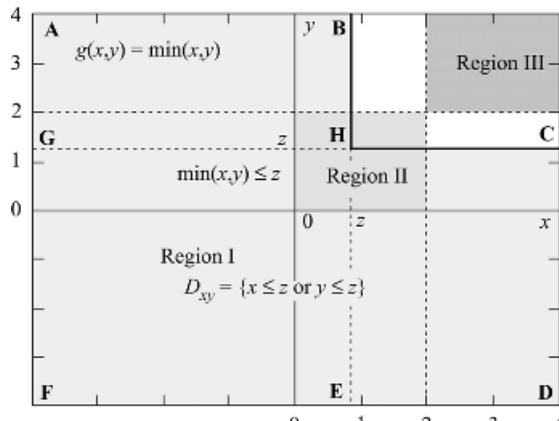


FIGURE 13.1.2

4. The probability of  $D_{xy}$  is given by  $P\{(X \leq z) \text{ or } (Y \leq z)\}$ , which can be written as  $P\{\text{rectangle } ABEF\} + P\{\text{rectangle } GCDF\} - P\{\text{rectangle } GHEF\}$ . We subtract  $P\{\text{rectangle } GHEF\}$  since it occurs twice in the addition of the first two terms. Hence

$$F_Z(z) = F_Y(z) + F_X(z) - F_{XY}(z,z) \quad (13.1.7)$$

If  $X$  and  $Y$  are independent, then Eq. (13.1.7) becomes

$$F_Z(z) = F_Y(z) + F_X(z) - F_X(z)F_Y(z) \quad (13.1.8)$$

and the density function  $f_Z(z)$  is given by

$$f_Z(z) = f_Y(z) + f_X(z) - f_X(z)F_Y(z) - F_X(z)f_Y(z) \quad (13.1.9)$$

We will define the joint density function  $f_{XY}(x,y)$  in the region  $\{0 < x \leq 4\}, \{0 < y \leq 2\}$  as shown in Fig. 13.1.3 and find the distribution  $F_Z(z)$ .

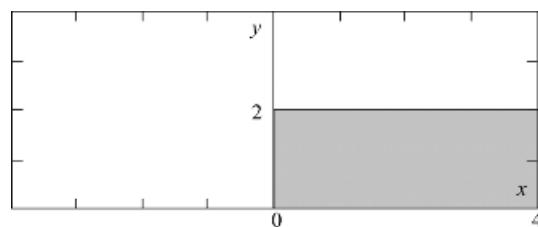


FIGURE 13.1.3

The joint density  $f_{XY}(x,y)$  and the distribution  $F_{XY}(x,y)$  are shown below:

$$f_{XY}(x,y) = \begin{cases} 0 & \text{for } x \leq 0, y \leq 0 \\ \frac{1}{8} & \text{for } 0 < x \leq 4, 0 < y \leq 2 \\ 0 & \text{for } x > 0, y > 0 \end{cases}$$

$$F_{XY}(x,y) = \begin{cases} 0 & \text{for } x \leq 0, y \leq 0 \\ \frac{xy}{8} & \text{for } 0 < x \leq 4, 0 < y \leq 2 \\ \frac{x}{4} & \text{for } 0 < x \leq 4, y > 2 \\ \frac{y}{2} & \text{for } x > 4, 0 < y \leq 2 \\ 1 & \text{for } x > 0, y > 0 \end{cases}$$

1. In region I, for  $\{z \leq 0\}$ ,  $F_Z(z) = 0$ .
2. In region II,  $\{0 < z \leq 2\}$  or  $\{0 < x \leq 2\}$ ,  $\{0 < y \leq 2\}$  we have the following formula from the equation derived earlier:

$$F_Z(z) = \frac{z}{4} + \frac{z}{2} - \frac{z^2}{8} = \frac{z}{8}(6 - z) \quad \text{for } 0 < z \leq 2$$

3. In region III,  $\{2 < z \leq 4\}$  or  $\{2 < x \leq 4\}$ ,  $\{2 < y \leq 4\}$

$$F_Z(z) = F_X(z) + F_Y(2) - F_{XY}(z,2) = \frac{z}{4} + 1 - \frac{z}{4} \cdot 1 = 1 \quad \text{for } 2 < z \leq 4$$

The same result can also be obtained from the fact that the random variable  $Y$  is always less than 2. Since the region  $D_{xy}$  is either  $X \leq z$  or  $Y \leq z$ , and  $z > 2$ , we have  $F_Z(z) = 1$ .

4. In region IV,  $\{z > 4\}$  or  $\{x > 4\}$ ,  $\{y > 4\}$ ,  $F_Z(z) = 1$ .

The density function  $f_Z(z)$  obtained by differentiating  $F_Z(z)$  is shown below.

$$f_Z(z) = \begin{cases} 0 & \text{for } z \leq 0 \\ \frac{3}{4} - \frac{z}{4} & \text{for } 0 < z \leq 2 \\ 0 & \text{for } 2 < z < 4 \\ 0 & \text{for } z > 4 \end{cases}$$

**Example 13.1.2** Given  $Z = g(X, Y) = \max(X, Y)$ , we will find  $F_Z(z)$  given the joint distribution  $F_{XY}(x,y)$ .

We follow the scheme outlined above.

1.  $g(x,y)$  is shown in Fig. 13.1.4.
2.  $g(x,y) = \max(x,y) = z$  is marked on the diagram.
3. The region  $D_{xy}$  is the set of all points in the  $x-y$  plane such that  $x \leq z$  and  $y \leq z$ . This region is shown in Fig. 13.1.4.
4. The probability of  $D_{xy}$  is given by  $P\{(X, Y \leq z)\}$ , which can be written as

$$F_Z(z) = F_{XY}(z,z) \quad (13.1.10)$$

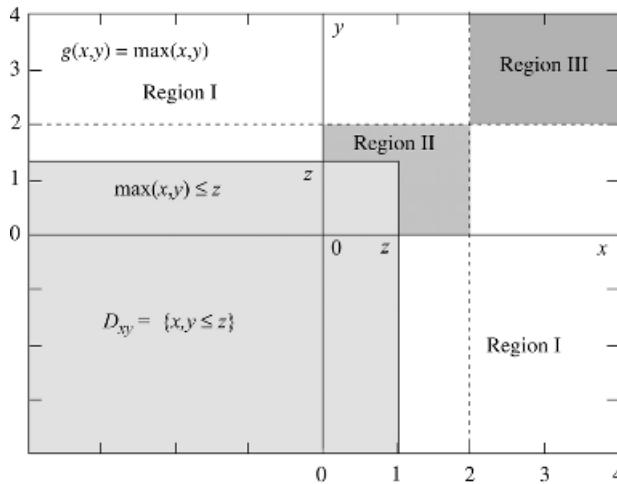


FIGURE 13.1.4

If  $X$  and  $Y$  are independent, then Eq. (13.1.10) becomes

$$F_Z(z) = F_X(z)F_Y(z) \quad (13.1.11)$$

and the density function  $f_Z(z)$  is given by

$$f_Z(z) = f_X(z)F_Y(z) + F_X(z)f_Y(z) \quad (13.1.12)$$

We will use the same joint density function  $f_{XY}(x,y)$  in the region  $\{0 < x \leq 4\}$ ,  $\{0 < y \leq 2\}$  as shown in Fig. 13.1.3 and find the distribution  $F_Z(z)$ .

1. In region I, for  $\{z \leq 0\}$ ,  $F_Z(z) = 0$ .
2. In region II,  $\{0 < z \leq 2\}$  or  $\{0 < x \leq 2\}$ ,  $\{0 < y \leq 2\}$ ,  $F_Z(z)$  is given by

$$F_Z(z) = \frac{z^2}{8} \quad \text{for } 0 < x \leq 2, 0 < y \leq 2$$

3. In the region III,  $\{2 < z \leq 4\}$  or  $\{2 < x \leq 4\}$ ,  $\{2 < y \leq 4\}$  we have

$$F_Z(z) = F_{XY}(z, 2) = \frac{z}{4} \cdot \frac{2}{2} = \frac{z}{4} \quad \text{for } 2 < x \leq 4, 2 < y \leq 4$$

We can also integrate the joint density function  $f_{XY}(x,y)$  in the ranges  $\{0 < x \leq z\}$  and  $\{0 < y \leq 2\}$ :

$$F_Z(z) = \int_0^z \int_0^2 \frac{1}{8} dy dx = \int_0^z \frac{2}{8} dx = \frac{z}{4} \quad \text{for } 2 < x \leq 4, 2 < y \leq 4$$

4. In region IV,  $\{z > 4\}$  or  $\{x > 4\}$ ,  $\{y > 4\}$ ,  $F_Z(z) = 1$ .

The density function  $f_Z(z)$  obtained by differentiating  $F_Z(z)$  is shown below.

$$f_Z(z) = \begin{cases} 0 & \text{for } z \leq 0 \\ \frac{z}{4} & \text{for } 0 < z \leq 2 \\ \frac{1}{4} & \text{for } 2 < z < 4 \\ 0 & \text{for } z \geq 4 \end{cases}$$

**Example 13.1.3** Given that  $Z = g(x, y) = X + Y$ , we have to find  $F_Z(z)$  from the knowledge of  $f_{XY}(x, y)$ .

1. The function  $g(x, y)$  is shown in Fig. 13.1.5.
2. The equation  $x + y = z$  is shown in Fig. 13.1.5.
3.  $D_{xy}$  is shown as the shaded region in the figure.
4. The distribution function  $F_Z(z)$  is

$$f_Z(z) = f_X(z)F_Y(z) + F_X(z)f_Y(z) \quad (13.1.13)$$

The density function  $f_Z(z)$  is obtained by differentiating  $F_Z(z)$ , and the result is

$$\begin{aligned} f_Z(z) &= \frac{d}{dz} \left( \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{XY}(x, y) dy dx \right) = \int_{-\infty}^{\infty} \frac{d}{dz} \left( \int_{-\infty}^{z-x} f_{XY}(x, y) dy \right) dx \\ &= \int_{-\infty}^{\infty} f_{XY}(x, z-x) dx \end{aligned} \quad (13.1.14)$$

If  $X$  and  $Y$  are independent in Eq. (13.1.14), then

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x) dx \quad (13.1.15)$$

Equation (13.1.15) is a convolution integral, and we can use characteristic functions to write

$$\Phi_Z(\omega) = \Phi_X(\omega)\Phi_Y(\omega) \quad (13.1.16)$$

$f_Z(z)$  can be obtained from  $\Phi_Z(\omega)$ .

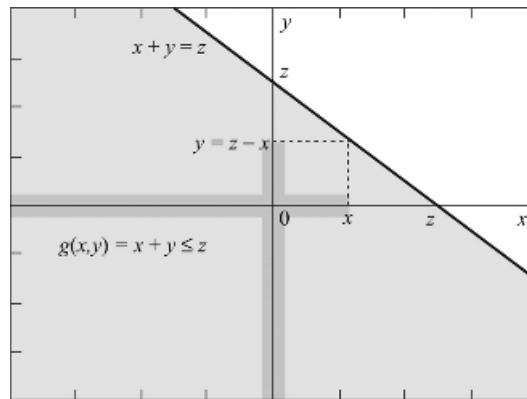


FIGURE 13.1.5

**Example 13.1.4** We are given  $X$  and  $Y$  as independent random variables with  $Z = g(X, Y) = 2X - \frac{1}{2}Y$ . The density functions  $f_X(x)$  and  $f_Y(y)$  are given by

$$f_X(x) = 2e^{-2x}u(x); \quad f_Y(y) = \frac{1}{2}e^{-y/2}u(y)$$

We have to find  $f_Z(z)$ .

Substituting  $U = 2X$  and  $V = \frac{1}{2}Y$ , we can obtain the corresponding densities of  $U$  and  $V$  from Eq. (12.3.7):

$$f_U(u) = \frac{1}{2}f_X\left(\frac{u}{2}\right) = e^{-u}u(u); \quad f_V(v) = 2f_Y(2v) = e^{-v}u(v)$$

With this substitution,  $Z$  can be given by  $Z = U + (-V)$ . Since  $U$  and  $V$  are also independent, we can use Eq. (13.1.16) to find  $f_Z(z)$ . The characteristic functions of  $U$  and  $(-V)$  are

$$\Phi_U(\omega) = \frac{1}{1-j\omega}; \quad \Phi_{-V}(\omega) = \frac{1}{(1+j\omega)}$$

Using the independence property, we obtain

$$\Phi_Z(\omega) = \Phi_U(\omega)\Phi_{-V}(\omega) = \frac{1}{(1-j\omega)(1+j\omega)} = \frac{1}{2}\left(\frac{1}{1-j\omega} + \frac{1}{1+j\omega}\right)$$

Taking the inverse transform, we obtain

$$f_Z(z) = \frac{1}{2}[e^{-z}u(z) + e^z u(-z)] = \frac{1}{2}e^{|z|}$$

The functions  $f_X(x)$ ,  $f_Y(y)$ , and  $f_Z(z)$  are shown in Fig. 13.1.6.

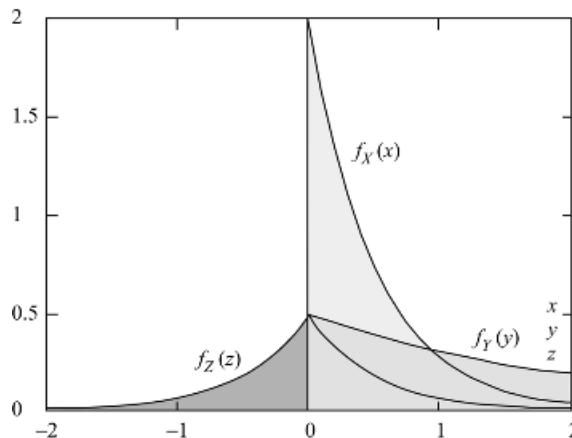


FIGURE 13.1.6

**Example 13.1.5** Two independent random variables  $X$  and  $Y$  are distributed normally as

$$f_X(x) = \frac{1}{\sqrt{2\pi \cdot 3}} e^{-(1/2)[(x-2)/3]^2}; \quad f_Y(y) = \frac{1}{\sqrt{2\pi \cdot 4}} e^{-(1/2)[(y-3)/4]^2}$$

Another random variable  $Z$  is given by  $Z = X - Y$ . We have to find  $f_Z(z)$ .

The characteristic functions of  $X$  and  $(-Y)$  are given by

$$\Phi_X(\omega) = e^{j\omega^2 - [(\omega^2 9)/2]}; \quad \Phi_Y(\omega) = e^{-j\omega 3 - [(\omega^2 16)/2]}$$

Using the independence of  $X$  and  $Y$ , we have

$$\Phi_Z(\omega) = \left( e^{j\omega^2 - [(\omega^2 9)/2]} \right) \left( e^{j\omega^3 - [(\omega^2 16)/2]} \right) = e^{-j\omega - [(\omega^2 25)/2]}$$

and  $f_Z(z)$  is also normally distributed with mean  $-1$  and variance  $25$ , and is given by

$$f_Z(z) = \frac{1}{\sqrt{2\pi \cdot 5}} e^{-(1/2)[(x-1)/5]^2}$$

Hence, the probability of a sum of independent Gaussian random variables is also a Gaussian random variable with mean equaling the sum of means, and variance equaling the sum of variances.

**Example 13.1.6** This example illustrates when the two random variables  $X$  and  $Y$  are not independent but are given by a joint Gaussian density function:

$$f_{XY}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sqrt{1-\rho^2}}[x^2 - 2\rho xy + y^2]\right\} \quad (13.1.17)$$

where the correlation coefficient is  $\rho$  and the random variable  $Z = X + Y$ . Here we can use Eq. (13.1.14) only. Substituting  $f_{XY}(x, y)$  into Eq. (13.1.14), we have

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{XY}(x, z-x) dx \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}[x^2 - 2\rho x(z-x) + (z-x)^2]\right\} dx \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}[2x^2 - 2\rho xz + 2\rho x^2 + z^2 - 2zx]\right\} dx \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{\frac{-1}{2(1-\rho^2)}[2x^2(1+\rho) - 2xz(1+\rho) + z^2]\right\} dx \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-z^2/2(1-\rho^2)} \int_{-\infty}^{\infty} e^{[-(1+\rho)/(1-\rho^2)][x^2 - xz]} dx \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-z^2/2(1-\rho^2)} \int_{-\infty}^{\infty} e^{-[1/(1-\rho)][x^2 - xz]} dx \end{aligned} \quad (13.1.18)$$

Completing squares under the integral sign in Eq. (13.1.18), we have

$$\begin{aligned}
f_Z(z) &= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-z^2/2(1-\rho^2)} \int_{-\infty}^{\infty} \exp\left\{\frac{-1}{1-\rho}\left[x - \frac{z}{2}\right]^2 - \frac{z^2}{4}\right\} dx \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{[-z^2/2(1-\rho^2)] + [z^2/4(1-\rho)]} \int_{-\infty}^{\infty} \exp\left\{\frac{-1}{1-\rho}\left[x - \frac{z}{2}\right]^2\right\} dx \\
&= \frac{1}{\sqrt{2\pi}} e^{-z^2/4(1+\rho)} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \int_{-\infty}^{\infty} \exp\left\{\frac{-1}{2}\left[\frac{x - (z/2)}{(1-\rho)/2}\right]^2\right\} dx \\
&= \frac{1}{\sqrt{2\pi}} e^{-z^2/4(1+\rho)} \frac{1}{\sqrt{2(1+\rho)}} \cdot \underbrace{\frac{1}{\sqrt{2\pi\left(\frac{1-\rho}{2}\right)}} \int_{-\infty}^{\infty} \exp\left\{\frac{-1}{2}\left[\frac{x - (z/2)}{(1-\rho)/2}\right]^2\right\} dx}_{=1} \\
&= \frac{1}{\sqrt{2(1+\rho)}} \cdot \frac{1}{\sqrt{2\pi}} e^{-z^2/4(1+\rho)}
\end{aligned} \tag{13.1.19}$$

Thus, we see that the sum of jointly Gaussian-distributed zero mean, unit variance random variables with correlation coefficient  $\rho$  is also a zero mean Gaussian with variance  $2(1+\rho)$ . If  $\rho=0$ , we have independent Gaussian random variables and variance equals 2, which is the sum of unit variances. If  $\rho=1$ , we have essentially  $Z=2X$  and the variance is 4.

**Example 13.1.7** Two independent Gaussian random variables  $X$  and  $Y$  are with zero mean and variance  $\sigma^2$ . The probability density functions of  $X$  and  $Y$  are

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x/2\sigma^2)}; f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y/2\sigma^2)} \tag{13.1.20}$$

Another random variable  $Z$  is given by  $Z = g(X, Y) = \sqrt{X^2 + Y^2}$ . We have to find the pdf  $f_Z(z)$ . The region  $D_{xy} = \sqrt{x^2 + y^2} \leq z$  is shown in Fig. 13.1.7.

Using Eq. (13.1.6), we have

$$F_Z(z) = \iint_{\sqrt{x^2+y^2} \leq z} \frac{1}{\sigma^2 2\pi} e^{-(\xi^2 + \eta^2)/2\sigma^2} d\xi d\eta$$

We can solve this integral by using polar coordinates using the following substitution:

$$\xi = r \cos \theta; \quad \eta = r \sin \theta; \quad d\xi d\eta = r dr d\theta; \quad \theta = \tan^{-1} \frac{\eta}{\xi}$$

Thus

$$\begin{aligned}
F_Z(z) &= \int_0^{2\pi} \int_0^z \frac{1}{\sigma^2 2\pi} e^{-(r^2/2\sigma^2)} r dr d\theta = \int_0^z \frac{r}{\sigma^2} e^{-(r^2/2\sigma^2)} dr \\
&= \left[1 - e^{(-z^2/2\sigma^2)}\right] u(z)
\end{aligned} \tag{13.1.21}$$

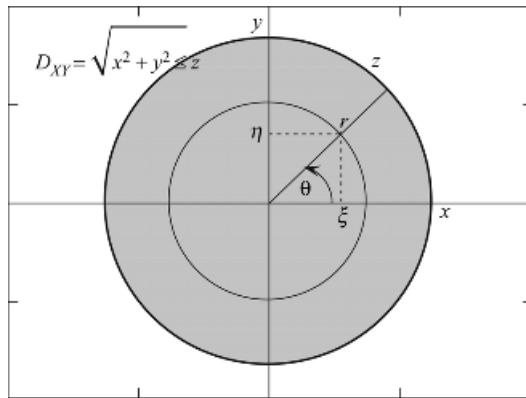


FIGURE 13.1.7

and

$$f_Z(z) = \frac{z}{\sigma^2} e^{-(z^2/2\sigma^2)} u(z) \quad (13.1.22)$$

where  $u(z)$  is the usual unit step function. This is a Rayleigh distribution as given in Eq. (7.8.5) or a chi-square distribution with 2 degrees of freedom.

**Example 13.1.8** This example is similar to the previous example except that  $Z = g(X, Y) = X^2 + Y^2$ .

The region  $D_{xy} = x^2 + y^2 \leq z$ . We can now find the pdf  $f_Z(z)$  using polar coordinates as in the previous example. Thus

$$\begin{aligned} F_Z(z) &= \int_0^{2\pi} \int_0^{\sqrt{z}} \frac{1}{\sigma^2 2\pi} e^{-(r^2/2\sigma^2)} r dr d\theta = \int_0^{\sqrt{z}} \frac{r}{\sigma^2} e^{-(r^2/2\sigma^2)} dr \\ &= \left[ 1 - e^{(-z/2\sigma^2)} \right] u(z) \end{aligned} \quad (13.1.23)$$

and the density function  $f_Z(z)$  is given by

$$f_Z(z) = \left( \frac{1}{2\sigma^2} e^{-z/2\sigma^2} \right) u(z) \quad (13.1.24)$$

This is an exponential distribution or a chi-square distribution given by Eq. (7.7.3) with 2 degrees of freedom.

The density functions for  $Z = \sqrt{X^2 + Y^2}$  and  $Z = X^2 + Y^2$  given by Eqs. (13.1.22) and (13.1.24) are shown in Fig. 13.1.8 for comparison.

**Example 13.1.9** In Example 13.1.7 we assumed that the independent random variables  $X$  and  $Y$  are of zero mean. In this example, we will assume that  $X$  and  $Y$  have the same variance  $\sigma^2$  and that their mean values are  $\mu_X$  and  $\mu_Y$  respectively. The density functions are

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu_X)/2\sigma^2]}; \quad f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-[(y-\mu_Y)/2\sigma^2]} \quad (13.1.25)$$

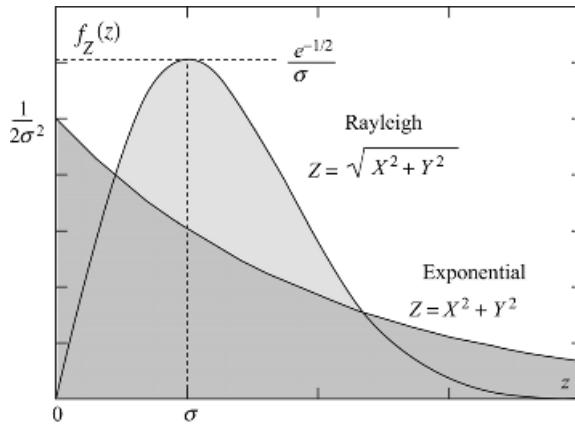


FIGURE 13.1.8

We will now find  $f_Z(z)$  for  $Z = \sqrt{X^2 + Y^2}$ . Function  $F_Z(z)$  is now given by

$$F_Z(z) = \iint_{\sqrt{x^2+y^2} \leq z} \frac{1}{\sigma^2 2\pi} \exp \left[ -\frac{(\xi - \mu_X)^2 + (\eta - \mu_Y)^2}{2\sigma^2} \right] d\xi d\eta; \quad z > 0 \quad (13.1.26)$$

Expanding Eq. (13.1.26) we have,

$$F_Z(z) = \iint_{\sqrt{x^2+y^2} \leq z} \frac{1}{\sigma^2 2\pi} \exp \left[ -\frac{\xi^2 + \eta^2 - 2(\mu_X \xi + \mu_Y \eta) + \mu_X^2 + \mu_Y^2}{2\sigma^2} \right] d\xi d\eta; \quad z > 0 \quad (13.1.27)$$

Substituting polar coordinates in Eq. (13.1.27), we obtain the following:

$$\begin{aligned} F_Z(z) &= \int_0^{2\pi} \int_0^z \frac{1}{\sigma^2 2\pi} \exp \left[ -\frac{r^2 - 2r(\mu_X \cos \theta + \mu_Y \sin \theta) + \mu_X^2 + \mu_Y^2}{2\sigma^2} \right] r dr d\theta \\ &= \frac{e^{-(\mu_X^2 + \mu_Y^2)/\sigma^2}}{\sigma^2} \int_0^z e^{-(1/2)(r/\sigma)^2} \\ &\quad \times \frac{1}{2\pi} \left[ \int_0^{2\pi} \exp \left[ \frac{r(\mu_X \cos \theta + \mu_Y \sin \theta)}{\sigma^2} \right] d\theta \right] r dr; \quad z > 0 \end{aligned} \quad (13.1.28)$$

We now define

$$\begin{aligned} \mu_X &= \left( \sqrt{\mu_X^2 + \mu_Y^2} \right) \cos(\psi); \quad \mu_Y = \left( \sqrt{\mu_X^2 + \mu_Y^2} \right) \sin(\psi) \\ \psi &= \tan^{-1} \left( \frac{\mu_Y}{\mu_X} \right) \end{aligned} \quad (13.1.29)$$

and a noncentrality parameter  $m = \sqrt{\mu_X^2 + \mu_Y^2}$ . Substituting  $m$  in Eqs. (13.1.28) and (13.1.29), we have

$$\begin{aligned} F_Z(z) &= \frac{e^{-m^2/\sigma^2}}{\sigma^2} \int_0^z e^{-(1/2)(r/\sigma)^2} \frac{1}{2\pi} \left[ \int_0^{2\pi} \exp \left[ \frac{rm(\cos \theta \cos \psi + \mu_Y \sin \theta \sin \psi)}{\sigma^2} \right] d\theta \right] r dr \\ &= \frac{e^{-m^2/\sigma^2}}{\sigma^2} \int_0^z e^{-(1/2)(r/\sigma)^2} \frac{1}{2\pi} \left[ \int_0^{2\pi} \exp \left[ \frac{rm(\cos \theta - \psi)}{\sigma^2} \right] d\theta \right] r dr; \quad z > 0 \quad (13.1.30) \end{aligned}$$

The integral under  $\theta$  is the modified Bessel function of zero order with argument  $rm/\sigma^2$  given by

$$I_0\left(\frac{rm}{\sigma^2}\right) = \frac{1}{2\pi} \left[ \int_0^{2\pi} e^{-rm(\cos \theta - \psi)/\sigma^2} d\theta \right]$$

Substituting this value in Eq. (13.1.30) and differentiating with respect to  $z$ , we obtain

$$f_Z(z) = \frac{z}{\sigma^2} e^{-(z^2+m^2)/\sigma^2} I_0\left(\frac{zm}{\sigma^2}\right); \quad z > 0 \quad (13.1.31)$$

This is Rice's distribution as given in Eq. (7.8.11). With  $m = 0$ , we get the Rayleigh distribution.

**Example 13.1.10** In this example we have  $Z = g(X, Y) = XY$ . The function  $g(x, y) = xy = z$  is shown in Fig. 13.1.9 for  $z > 0$ . We have to find  $f_Z(z)$  given the joint density  $f_{XY}(x, y)$ .

The distribution function  $F_Z(z)$  is obtained by integrating  $f_{XY}(x, y)$  in the shaded region given by  $g(x, y) \leq z$ . Since the region  $D_{XY} = \{xy \leq z\}$  is bounded by two disjoint curves, we have to integrate  $f_{XY}(x, y)$  along the two regions  $\{-\infty < x \leq 0\}$  and  $\{0 < x \leq \infty\}$ . Hence

$$\begin{aligned} F_Z(z) &= \iint_{xy \leq z} f_{XY}(x, y) dy dx \\ &= \int_{-\infty}^0 \int_{z/x}^{\infty} f_{XY}(x, y) dy dx + \int_0^{\infty} \int_{-\infty}^{z/x} f_{XY}(x, y) dy dx \quad (13.1.32) \end{aligned}$$

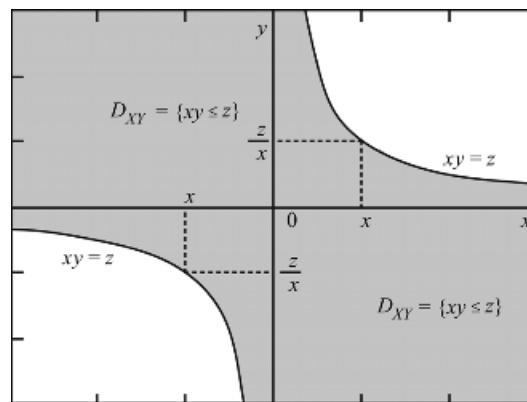


FIGURE 13.1.9

We differentiate Eq. (13.1.32) using the chain rule of differentiation and obtain  $f_Z(z)$ :

$$\begin{aligned} f_Z(z) &= - \int_{-\infty}^0 f_{XY}\left(x, \frac{z}{x}\right) \cdot \frac{1}{x} dx + \int_0^\infty f_{XY}\left(x, \frac{z}{x}\right) \cdot \frac{1}{x} dx \\ &= \int_{-\infty}^\infty \frac{1}{|x|} f_{XY}\left(x, \frac{z}{x}\right) dx; \quad z > 0 \end{aligned} \quad (13.1.33)$$

It can be shown by similar techniques that Eq. (13.1.33) is valid for  $z \leq 0$ . If  $X$  and  $Y$  are independent, then

$$f_Z(z) = \int_{-\infty}^\infty \frac{1}{|x|} f_X(x) f_Y\left(\frac{z}{x}\right) dx \quad \text{for all } z \quad (13.1.34)$$

We note that Eq. (13.1.34) is not a convolution integral even if the product of independent random variables is involved.

The problem in Example 13.1.10 can also be solved using the techniques developed in Chapter 12. We will first find the conditional density  $f_Z(z | X = x)$ , given the conditional density  $f_Y(y | X = x)$  with  $Z = xY$ . Since  $x$  is a scaling factor for the random variable  $Y$ , we can use Eq. (12.3.7) and obtain

$$f_Z(z | X = x) = \frac{1}{|x|} f_Y\left(\frac{z}{x} | X = x\right)$$

Hence, the unconditional density  $f_Z(z)$  is given by

$$f_Z(z) = \int_{-\infty}^\infty \frac{1}{|x|} f_Y\left(\frac{z}{x} | X = x\right) f_X(x) dx = \int_{-\infty}^\infty \frac{1}{|x|} f_{XY}\left(x, \frac{z}{x}\right) dx$$

which is the same as before.

**Example 13.1.11** We will now assume that  $X$  and  $Y$  are independent random variables with identical Cauchy distributions given by

$$f_X(x) = \frac{2}{\pi(x^2 + 4)}; \quad f_Y(y) = \frac{2}{\pi(y^2 + 4)}$$

Substituting these values in Eq. (13.1.34), we have

$$f_Z(z) = \int_{-\infty}^\infty \frac{1}{|x|} \frac{2}{\pi(x^2 + 4)} \cdot \frac{2}{\pi\left(\frac{z^2}{x^2} + 4\right)} dx$$

Since  $f_Z(z)$  is an even function, we can rewrite this equation as

$$f_Z(z) = \frac{4 \cdot 2}{\pi^2} \int_0^\infty \frac{x}{(x^2 + 4)(z^2 + 4x^2)} dx$$

Substituting  $w = x^2$  and  $dw = 2x \, dx$  and expanding into partial fractions, we have

$$\begin{aligned}
 f_Z(z) &= \frac{4}{\pi^2} \int_0^\infty \frac{1}{(w+4)(z^2+4w)} dw \\
 &= \frac{4}{\pi^2} \int_0^\infty \frac{1}{(z^2-16)} \left[ \frac{1}{(w+4)} - \frac{1}{(z^2+4w)} \right] dw \\
 &= \frac{4}{\pi^2(z^2-16)} \left[ \ln(w+4) - \ln(z^2+4w) \right] \Big|_0^\infty = \frac{4}{\pi^2(z^2-16)} \left[ \ln\left(\frac{w+4}{z^2+4w}\right) \right] \Big|_0^\infty \\
 &= \frac{4}{\pi^2(z^2-16)} \left[ \ln\left(\frac{w+4}{z^2+4w}\right) \right] \Big|_0^\infty = \frac{4}{\pi^2(z^2-16)} \left[ \ln\left(\frac{1}{4}\right) - \ln\left(\frac{4}{z^2}\right) \right]
 \end{aligned}$$

or

$$f_Z(z) = \frac{4}{\pi^2(z^2-16)} \ln\left(\frac{z^2}{16}\right)$$

and  $f_Z(z)$  is shown in Fig. 13.1.10.

**Example 13.1.12** Given  $Z = g(X, Y) = Y/X$ , we have to find  $f_Z(z)$  from the knowledge of the joint density function  $f_{XY}(x, y)$ . The region  $D_{xy}(x, y) = \{x, y: y \leq xz\}$  is shown in Fig. 13.1.11. In the two regions of  $D_{xy}$  given by (1)  $x \leq 0$  and (2)  $x > 0$ ,  $F_Z(z)$  can be written as follows:

$$\begin{aligned}
 F_Z(z) &= \iint_{y \leq xz} f_{XY}(x, y) dy dx \\
 &= \int_{-\infty}^0 \int_{xz}^{\infty} f_{XY}(x, y) dy dx + \int_0^{\infty} \int_{-\infty}^{xz} f_{XY}(x, y) dy dx \quad (13.1.35)
 \end{aligned}$$

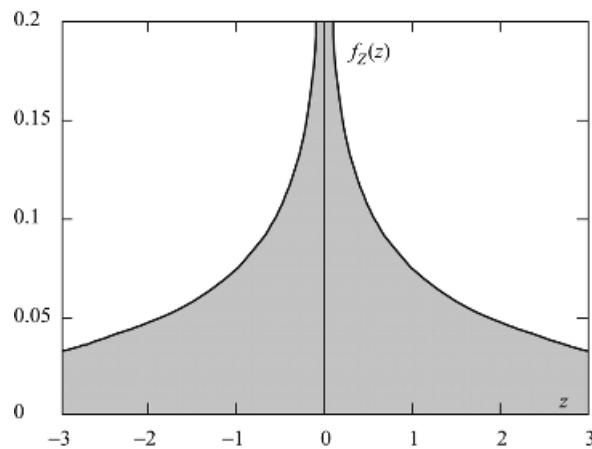


FIGURE 13.1.10

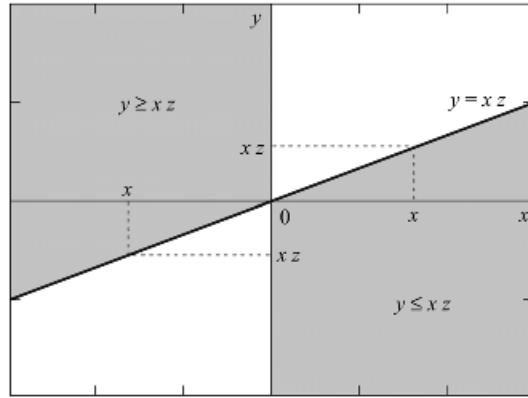


FIGURE 13.1.11

Using the chain rule of differentiation in Eq. (13.1.35), we can obtain  $f_Z(z)$  as

$$\begin{aligned} f_Z(z) &= - \int_{-\infty}^0 f_{XY}(x, xz) x \, dx + \int_0^\infty f_{XY}(x, xz) x \, dx \\ &= \int_{-\infty}^\infty |x| f_{XY}(x, xz) dx \end{aligned} \quad (13.1.36)$$

As an illustration, we will find  $f_Z(z)$  given that  $X$  and  $Y$  are jointly Gaussian random variables with means 0 and variances  $\sigma_X^2$  and  $\sigma_Y^2$  with correlation coefficient  $\rho$ , given by

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_X^2} - 2\rho\frac{xy}{\sigma_X\sigma_Y} + \frac{y^2}{\sigma_Y^2}\right)\right]$$

Substituting  $f_{XY}(x, y)$  into Eq. (13.1.36) and using the fact that  $f_{XY}(x, y)$  is an even function we have,

$$\begin{aligned} f_Z(z) &= \frac{2}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \int_0^\infty x \exp\left[-\frac{x^2}{2(1-\rho^2)}\left(\frac{1}{\sigma_X^2} - 2\rho\frac{z}{\sigma_X\sigma_Y} + \frac{z^2}{\sigma_Y^2}\right)\right] dx \\ &= \frac{1}{\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \frac{\sigma_X^2\sigma_Y^2(1-\rho^2)}{\sigma_Y^2 - 2\rho\sigma_X\sigma_Y + z^2\sigma_X^2} \\ &= \frac{\sigma_X\sigma_Y\sqrt{1-\rho^2}}{\pi\sigma_X^2\left(z^2 - 2\rho z\frac{\sigma_Y}{\sigma_X} + \frac{\sigma_Y^2}{\sigma_X^2}\right)} \end{aligned} \quad (13.1.37)$$

By completing the squares in the denominator of Eq. (13.1.37), we can write

$$f_Z(z) = \frac{\sigma_X\sigma_Y\sqrt{1-\rho^2}}{\pi\left(\sigma_X^2\left(z - \rho\frac{\sigma_Y}{\sigma_X}\right)^2 + \sigma_Y^2(1-\rho^2)\right)} \quad (13.1.38)$$

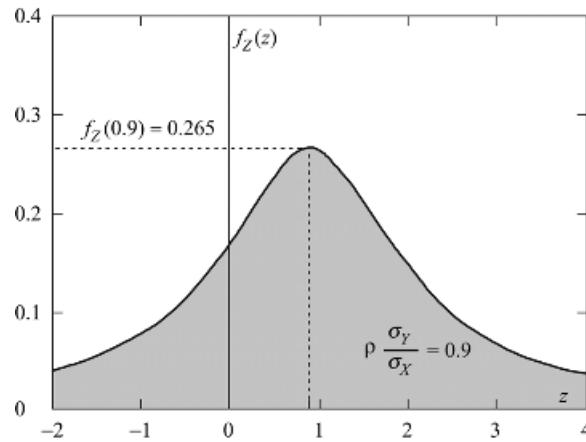


FIGURE 13.1.12

This is a Cauchy distribution centered at  $z = \rho(\sigma_Y/\sigma_X)$  and is shown in Fig. 13.1.12 with  $\sigma_X^2 = 4$ ,  $\sigma_Y^2 = 9$ , and  $\rho = \frac{3}{5}$ .

Integrating Eq. (13.1.38) with respect to  $z$ , we obtain the distribution function  $F_Z(z)$  as

$$F_Z(z) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \left[ \frac{\sigma_X z - \rho \sigma_Y}{\sigma_Y \sqrt{1 - \rho^2}} \right] \quad (13.1.39)$$

**Example 13.1.13** We are given two independent random variables,  $X$  and  $Y$ , which are exponentially distributed as

$$f_X(x) = 2e^{-2x}u(x); \quad f_Y(y) = e^{-y}u(y)$$

We have to find the density function  $f_Z(z)$  of the random variable given by  $Z = (Y/X)$ . From Eq. (13.1.36), we have

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} |x| f_{XY}(x, xz) dx \\ &= \int_0^{\infty} x \cdot 2e^{-2x} \cdot e^{-xz} dx = \int_0^{\infty} 2xe^{-x(2+z)} dx \\ &= \frac{2}{(2+z)^2} \end{aligned}$$

Integrating  $f_Z(z)$ , we obtain the distribution  $F_Z(z)$  as

$$F_Z(z) = \left( 1 - \frac{2}{2+z} \right) u(z)$$

These functions are shown in Fig. 13.1.13.

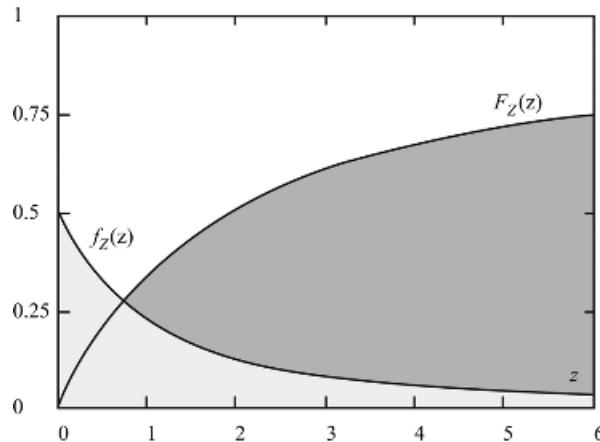


FIGURE 13.1.13

## 13.2 TWO FUNCTIONS OF TWO RANDOM VARIABLES, $Z = g(X, Y)$ , $W = h(X, Y)$

We are given two random variables,  $X$  and  $Y$ , and their joint density function  $f_{XY}(x, y)$ . We form the functions  $Z = g(X, Y)$  and  $W = h(X, Y)$  and determine the joint density  $f_{ZW}(z, w)$  from knowledge of  $f_{XY}(x, y)$ . The analysis is very similar to the one presented in the previous section.

The functions  $Z = g(X, Y)$  and  $W = h(X, Y)$  map the domain space of  $\{Z, W\}$  given by  $\{C_{xy}, \mathcal{F}_{XY}, P_{XY}\}$  into the range space  $\{C_{zw}, \mathcal{F}_{ZW}, P_{ZW}\}$  as shown in Fig. 13.2.1. From Fig. 13.2.1, we obtain

$$Z = g(X, Y), \quad W = h(X, Y); \quad \{C_{xy}, \mathcal{F}_{XY}, P_{XY}\} \rightarrow \{C_{zw}, \mathcal{F}_{ZW}, P_{ZW}\} \quad (13.2.1)$$

and the mapping of  $\{X, Y\}$  is from the probability space  $\{S, \mathcal{F}, P\}$  (where  $S$  is the product space  $S = S_X \times S_Y$ ) into the domain space of  $\{Z, W\}$  given by  $\{C_{xy}, \mathcal{F}_{XY}, P_{XY}\}$ :

$$X, Y : \{S, \mathcal{F}, P\} \rightarrow \{C_{xy}, \mathcal{F}_{XY}, P_{XY}\} \quad (13.2.2)$$

Alternately, if  $D_{xy} \subset C_{xy}$  and  $D_{zw} \subset C_{zw}$ , then  $D_{zw}$  is the *image* of  $D_{xy}$  under the transformation  $\{g, h\}$  and  $D_{xy}$  is the *inverse image* of  $D_{zw}$  under the transformation  $\{g^{-1}, h^{-1}\}$ . Hence,  $D_{zw}$  is the set of all points  $(x, y)$  belonging to  $D_{xy}$  such that  $g(x, y), h(x, y)$  belongs

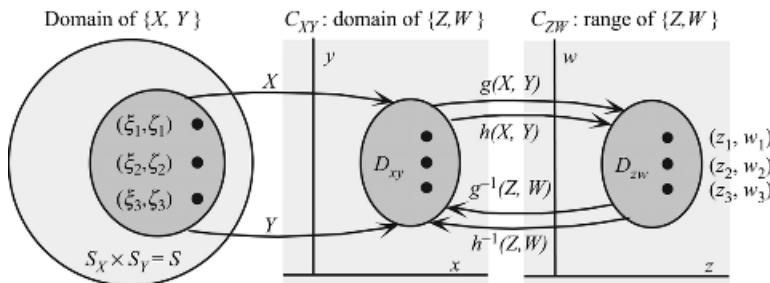


FIGURE 13.2.1

to  $D_{zw}$ . This is represented by

$$D_{xy} = \{x, y : g(x, y), h(x, y) \in D_{zw}\} \quad (13.2.3)$$

We also have conditions similar to the last section for  $Z$  and  $W$  to be a random variables.

### Conditions on $Z = g(X, Y)$ and $W = h(X, Y)$ to Be Random Variables

1. The domain of  $\{g(x, y), h(x, y)\}$  must include the range of the RV  $Z$  and  $W$ .
2. For each  $z, w \in D_{zw}$ , the set  $\{g(x, y) \leq z, h(x, y) \leq w\}$  must form a field  $\mathcal{F}_{zw}$ ; that is, it must consist of the unions and intersections of a countable number of intervals.
3. The probability of the events  $\{g(x, y) = \pm\infty\}$  and  $\{h(x, y) = \pm\infty\}$  is 0.

In essence, the transformation  $\{Z = g(X, Y), W = h(X, Y)\}$  induces a probability measure  $P_{ZW}$  on  $\mathcal{F}_{ZW}$ , and we have to find this probability measure. From Fig. 13.2.1 we can write

$$\begin{aligned} P_{ZW}(Z, W) &= P(Z, W \in D_{zw}) \\ &= P_{XY}(g^{-1}(Z, W), h^{-1}(Z, W) \in D_{xy}) = P_{XY}(X, Y \in D_{xy}) \end{aligned} \quad (13.2.4)$$

The set  $D_{zw}$  consists of all the countable unions and intersections of the type  $\{a < Z \leq b\}$ ,  $\{c < W \leq d\}$ . Without loss of generality, we will assume that  $D_{zw} = \{Z \leq z, W \leq w\}$ . Thus, substituting for  $D_{xy}$  from Eq. (13.2.3) in Eq. (13.2.4), we have

$$\begin{aligned} P_{ZW}(Z, W) &= P_{XY}[X, Y \in D_{xy}] = P_{XY}[x, y : g(x, y), h(x, y) \in D_{zw}] \\ &= P_{XY}[x, y : g(x, y) \leq z, h(x, y) \leq w] = F_{XY}(z, w) \\ &= P(Z \leq z, W \leq w) = F_{ZW}(z, w) \end{aligned} \quad (13.2.5)$$

Equation (13.2.5) can also be written as

$$P_{ZW}(Z, W) = P_{XY}[X, Y \in D_{xy}] = F_{ZW}(z, w) = \int \int_{D_{xy}} f_{XY}(\xi, \eta) d\xi d\eta \quad (13.2.6)$$

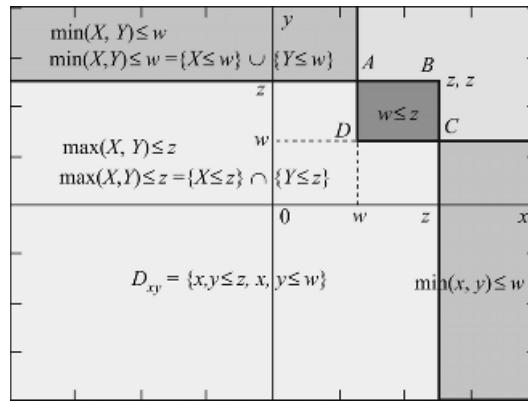
where  $f_{XY}(x, y)$  is the joint density of the random variables  $X$  and  $Y$ .

In conclusion, we can say that the probability of  $\{Z \leq z, W \leq w\}$  is given by the probability of the set of all values of  $x$  and  $y$  such that  $\{g(x, y) \leq z, h(x, y) \leq w\}$  or by integrating the joint density  $f_{XY}(x, y)$  over the domain  $D_{xy}$ . We will use both these interpretations in solving problems.

**Example 13.2.1** Here we are given that  $Z = \max(X, Y)$ ,  $W = \min(X, Y)$  and we have to find  $F_{ZW}(z, w)$  from the knowledge of the joint density  $F_{XY}(x, y)$ . Both  $Z$  and  $W$  are shown in Fig. 13.2.2.

The regions  $\min(X, Y) \leq w = \{X \leq w\} \cup \{Y \leq z\}$  and  $\max(X, Y) \leq z = \{X \leq z\} \cap \{Y \leq z\} = \{X \leq z\} \cap \{Y \leq z\}$  are shown in Fig. 13.2.2. There are two regions of interest:  $w > z$  and  $w \leq z$ . In the first region  $w > z$ ,  $D_{xy}$  equals the semiinfinite square defined by the point  $B = (z, z)$ . Thus

$$F_{Z,W}(z, w) = F_{XY}(z, z) = F_Z(z): \quad w > z \quad (13.2.7)$$



**FIGURE 13.2.2**

In the second region  $w \leq z$ ,  $D_{xy}$  equals the semiinfinite square defined by the point  $B = (z, z)$  minus the square  $ABCD$ . Hence

$$\begin{aligned}
F_{Z,W}(z,w) &= F_{XY}(z,z) - P(ABCD) \\
&= F_{XY}(z,z) - [F_{XY}(z,z) - F_{XY}(z,w) - F_{XY}(w,z) + F_{XY}(w,w)] \\
&= F_{XY}(z,w) + F_{XY}(w,z) - F_{XY}(w,w); \quad w \leq z
\end{aligned} \tag{13.2.8}$$

When  $w = z$ , Eq. (13.2.8) is identical to Eq. (13.2.7).

We can obtain  $F_W(w)$  by letting  $z$  tend to  $\infty$  in Eq. (13.2.8), yielding

$$\begin{aligned} F_{Z,W}(z,w) &= F_{XY}(z,w) + F_{XY}(w,z) - F_{XY}(w,w) \\ &= F_X(w) + F_Y(w) - F_{XY}(w,w) = F_W(w) \end{aligned} \quad (13.2.9)$$

which is similar to Eq. (13.1.7) derived in the previous section.

If  $X$  and  $Y$  are independent, then Eqs. (13.2.7) and (13.2.8) become

$$F_{Z,W}(z,w) = \begin{cases} F_X(z)F_Y(z): & w > z \\ F_X(z)F_Y(w) + F_X(w)F_Y(z) - F_X(w)F_Y(w): & w \leq z \end{cases} \quad (13.2.10)$$

**Example 13.2.2** To continue the previous example, the density functions of two independent random variables  $X$  and  $Y$  are given by

$$f_X(x) = 2e^{-2x}u(x); \quad f_Y(y) = e^{-y}u(y)$$

Hence the joint density function  $f_{XY}(x,y)$  is

$$f_{XY}(x,y) = 2e^{-(2x+y)}u(x)u(y)$$

Given random variables  $Z = \max(X, Y)$  and  $W = \min(X, Y)$ , we have to find the joint distribution  $F_{ZW}(z, w)$  and the marginal distributions  $F_Z(z)$  and  $F_W(w)$ .

The joint distribution  $F_{XY}(x,y)$  can be obtained by integrating  $f_X(x)$  and  $f_Y(y)$ :

$$F_{XY}(x,y) = (1 - e^{-2x})(1 - e^{-y})u(x)u(y)$$

$$= (1 - e^{2x} - e^{-y} + e^{-(2x+y)})u(x)u(y)$$

Using Eq. (13.2.10), we obtain

$$F_{ZW}(z, w) = \begin{cases} (1 - e^{-2z})(1 - e^{-z}); & z > 0, w > z \\ (1 - e^{-2z})(1 - e^{-w}) + (1 - e^{-2w})(1 - e^{-z}) \\ \quad - (1 - e^{-2w})(1 - e^{-w}); & z > 0, w \leq z \\ 1 - e^{-2z} - e^{-z} + e^{-3z}; & z > 0, w > z \\ 1 - e^{-2z} - e^{-z} + e^{-(2z+w)} + e^{-(z+2w)} - e^{-3w}; & z > 0, w \leq z \end{cases}$$

as the joint distribution function  $F_{ZW}(z, w)$ . The marginal distribution  $F_Z(z)$  is obtained by letting  $w = z$  in the second equation of  $F_{ZW}(z, w)$  since  $w$  cannot exceed  $z$ . Thus

$$F_Z(z) = 1 - e^{-2z} - e^{-z} + e^{-3z}, \quad z > 0$$

which is the same as the first equation. To obtain  $F_W(w)$ , we let  $z$  tend to  $\infty$  in the second equation of  $F_{ZW}(z, w)$  and obtain

$$F_W(w) = 1 - e^{-3w}, \quad w > 0$$

**Example 13.2.3** We are given  $Z = X^2 + Y^2$  and  $W = Y/X$ , where  $X$  and  $Y$  are zero mean independent Gaussian random variables with joint density function

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma^2} e^{-[(x^2+y^2)/2\sigma^2]}$$

The region  $D_{xy}$  is the set of all values of  $x, y$  such that  $x^2 + y^2 \leq z$  and  $y/x \leq w$  and is shown in Fig. 13.2.3. The joint distribution  $F_{ZW}(z, w)$  is found by integrating  $F_{ZW}(z, w)$  over the shaded region of the figure.

$$F_{ZW}(z, w) = \int \int_{D_{xy}} \frac{1}{2\pi\sigma^2} e^{-[(x^2+y^2)/2\sigma^2]} dx dy$$

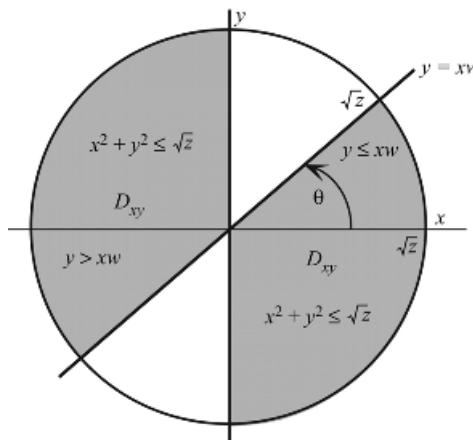


FIGURE 13.2.3

Using polar coordinates,  $x = \sqrt{z} \cos(\theta)$ ,  $y = \sqrt{z} \sin(\theta)$ ,  $dx dy = \sqrt{z} d\sqrt{z} d\theta$ , we can rewrite the preceding equation as

$$F_{ZW}(z, \theta) = 2 \int_{(\pi/2)}^{\theta} \int_0^{\sqrt{z}} \frac{1}{2\pi\sigma^2} e^{-(z/2\sigma^2)} \sqrt{z} d\sqrt{z} d\alpha$$

with the factor 2 appearing since there are two shaded regions. This last equation can be simplified as follows:

$$\begin{aligned} F_{ZW}(z, \theta) &= \frac{1}{\pi} \left( \theta + \frac{\pi}{2} \right) \int_0^z \frac{1}{\sigma^2} e^{-(z/2\sigma^2)} \sqrt{z} \frac{dz}{2\sqrt{z}} \\ &= \left( \frac{1}{2} + \frac{\theta}{\pi} \right) \left( 1 - e^{-(z/2\sigma^2)} \right); \quad z > 0, \quad -\frac{\pi}{2} < \theta \leq \frac{\pi}{2} \end{aligned}$$

Since  $w = (y/w) = \tan(\theta)$ , we can substitute  $\theta = \tan^{-1}(w)$  and obtain the marginal distributions  $F_Z(z)$  and  $F_W(w)$  as follows:

$$F_Z(z) = 1 - e^{-(z/2\sigma^2)}; \quad z > 0; \quad F_W(w) = \frac{1}{2} + \frac{\tan^{-1}(w)}{\pi}; \quad -\infty < w < \infty$$

The corresponding density functions are obtained by differentiation:

$$f_Z(z) = \frac{1}{2\sigma^2} e^{-(z/2\sigma^2)}; \quad z > 0; \quad f_W(w) = \frac{1}{\pi(1+w^2)}; \quad -\infty < w < \infty$$

The functions  $F_W(w)$  and  $f_W(w)$  are Cauchy-distributed and are shown in Figures 13.2.4 and 13.2.5.

The examples given above are some of the simple problems to solve with transformations of multiple functions of multiple random variables. In general, this method of solving for distribution functions of two functions of two random variables is more difficult. We now analyze alternate methods of finding the joint density functions.

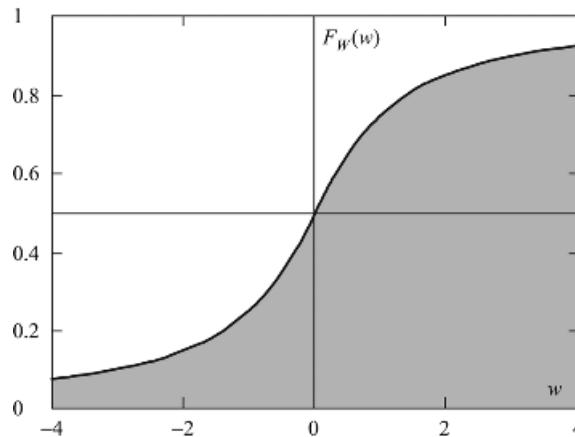


FIGURE 13.2.4

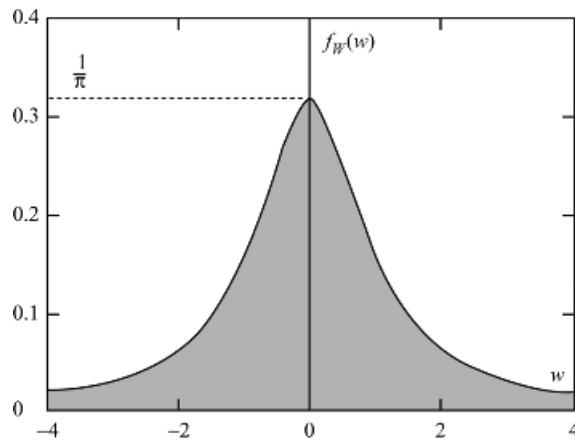


FIGURE 13.2.5

### 13.3 DIRECT DETERMINATION OF JOINT DENSITY $f_{ZW}(z,w)$ FROM $f_{XY}(x,y)$

It is easier to obtain  $f_{ZW}(z,w)$  directly from the joint density function  $f_{XY}(x,y)$  exactly as we did in the case of a single function of a single random variable. We are given the transformation  $Z = g(X,Y)$  and  $W = h(X,Y)$ , and the inverse images  $g^{-1}(Z,W)$  and  $h^{-1}(Z,W)$  exist. However, corresponding to the differential parallelogram  $\Delta D_{zw}$  in the  $Z-W$  plane, there may be several differential unconnected parallelograms  $\Delta D_{xy} = \sum dx_k dy_k$ ,  $k = 1, \dots, n$  in the  $X-Y$  plane as shown in Fig. 13.3.1. These differential regions are connected by the relation

$$dz dw = dx_1 dy_1 |J(x_1, y_1)| + \dots + dx_n dy_n |J(x_n, y_n)| \quad (13.3.1)$$

where  $|J(x,y)|$  is the determinant of the Jacobian matrix  $J(x,y)$ , defined by

$$J(x,y) = \begin{bmatrix} \frac{\partial z(x,y)}{\partial x} & \frac{\partial z(x,y)}{\partial y} \\ \frac{\partial w(x,y)}{\partial x} & \frac{\partial w(x,y)}{\partial y} \end{bmatrix} \quad (13.3.2)$$

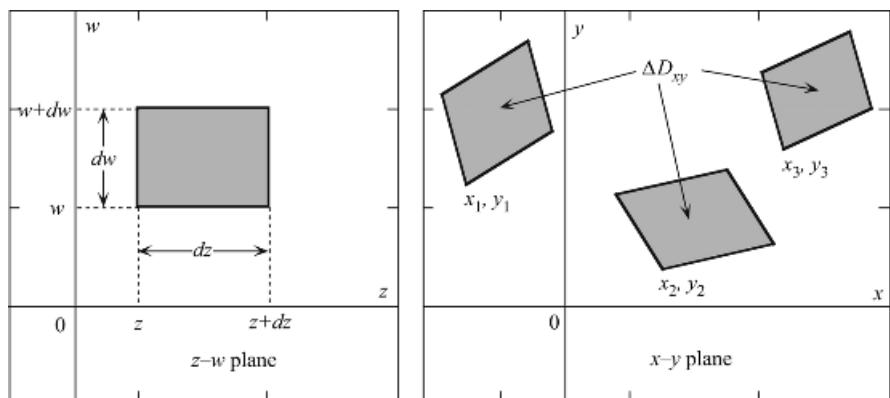


FIGURE 13.3.1

The joint density  $f_{ZW}(z,w)$  can be obtained from  $f_{XY}(x,y)$  as follows:

$$f_{ZW}(z,w)dz dw = f_{XY}(x_1,y_1)dx_1 dy_1 + \cdots + f_{XY}(x_n,y_n)dx_n dy_n \quad (13.3.3)$$

Substitution of Eq. (13.3.1) in Eq. (13.3.3) results in

$$\begin{aligned} f_{ZW}(z,w)dz dw &= \frac{f_{XY}(x_1,y_1)dz dw}{|J(x_1,y_1)|} + \cdots + \frac{f_{XY}(x_n,y_n)dz dw}{|J(x_n,y_n)|} \\ &= \sum_{k=1}^n \frac{f_{XY}(x_k,y_k)dz dw}{|J(x_k,y_k)|} \end{aligned} \quad (13.3.4)$$

or

$$f_{ZW}(z,w) = \sum_{k=1}^n \frac{f_{XY}(x_k,y_k)}{|J(x_k,y_k)|} \quad (13.3.5)$$

From this derivation it is clear that certain restrictions have to be satisfied for  $g(x,y)$  and  $h(x,y)$  if we have to determine  $f_{ZW}(z,w)$  directly from  $f_{XY}(x,y)$ :

1. All the partial derivative of  $g(x,y)$  and  $h(x,y)$  must exist.
2. There cannot be any planar solutions to  $g(x,y) = z$  and  $h(x,y) = w$ .

We can now formulate the methodology of finding  $f_{ZW}(z,w)$  from  $f_{XY}(x,y)$ .

#### Steps for Direct Determination of $f_{ZW}(z,w)$

1. Check whether the restrictions on  $g(x,y)$  and  $h(x,y)$  are satisfied.
2. Find the  $n$  real solutions to the system of equations

$$g(x,y) = z; \quad h(x,y) = w$$

and express them in terms of  $z$  and  $w$ . Let the  $n$  solutions be  $\{x_k, y_k, k = 1, \dots, n\}$ .

3. Find the Jacobian determinant  $|J|$  for each  $x_k, y_k$ , and express this also in terms of  $z$  and  $w$ .
4. Find the density function  $f_{ZW}(z,w)$  from Eq. (13.3.5):

$$f_{ZW}(z,w) = \sum_{k=1}^n \frac{f_{XY}(x_k,y_k)}{|J(x_k,y_k)|}$$

**Example 13.3.1** The functions  $Z = g(X,Y)$  and  $W = h(X,Y)$  are defined by

$$g(X,Y) = X + Y; \quad h(X,Y) = X - Y$$

where  $X$  and  $Y$  are independent positive random variables with joint density  $f_{XY}(x,y)$  given by  $6e^{-(3x+2y)} \cdot u(x)u(y)$ . We have to find  $f_{ZW}(z,w)$ .

1. All partials exist for  $g(x,y) = x + y$  and  $h(x,y) = x - y$ .
2. In the equations  $x + y = z$  and  $x - y = w$ , we note that  $z > 0$  and  $w$  can have negative values. Solving for  $x$  and  $y$  results in one real solution:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{-2} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix} \quad \text{or} \quad x = \frac{z+w}{2}; \quad y = \frac{z-w}{2}$$

Since  $x$  and  $y$  are positive, we must have  $z > |w|$  and  $z > w$ .

3. The Jacobian determinant is  $|J| = 2$ .

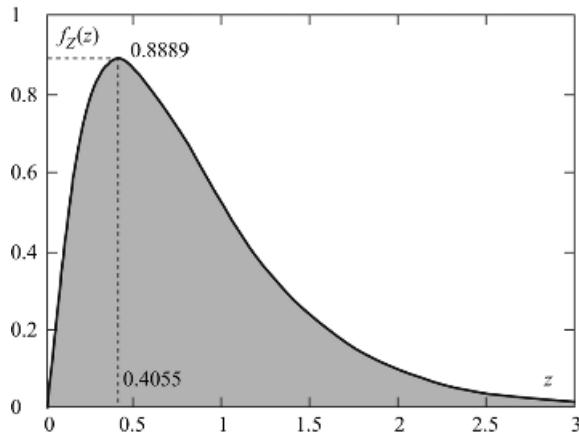


FIGURE 13.3.2

4. The joint density function  $f_{ZW}(z,w)$  from Eq. (13.3.5) is given by

$$f_{ZW}(z,w) = \frac{1}{2} 6 \exp \left[ -\left( 3 \frac{z+w}{2} + 2 \frac{z-w}{2} \right) \right] = 3e^{-[(5z+w)/2]}, \quad z > 0, \quad -z < w \leq z$$

Finding the marginal densities  $f_Z(z)$  and  $f_W(w)$  are a little tricky because of finding the limits of integrations with respect to  $z$  and  $w$ .

$$\begin{aligned} f_Z(z) &= \int_{-z}^z 3e^{-[(5z+w)/2]} dw; \quad z > 0, \quad -z < w \leq z \\ &= 6(e^{-2z} - e^{-3z})u(z) \end{aligned}$$

Function  $f_Z(z)$  is shown in Fig. 13.3.2.

Since  $w$  can take positive and negative values, we have the following integrals for  $f_W(w)$  for the two regions  $w > 0$  and  $w \leq 0$ . In the region  $w > 0$ , we have

$$\begin{aligned} f_W(w) &= \int_w^{\infty} 3e^{-[(5z+w)/2]} dz; \quad w > 0 \\ &= \frac{6}{5}(e^{-3w})u(z) \end{aligned}$$

and in the region  $w \leq 0$ , we have

$$\begin{aligned} f_W(w) &= \int_w^{\infty} 3e^{-[(5z+w)/2]} dz; \quad w \leq 0 \\ &= \frac{6}{5}(e^{2w})u(-w) \end{aligned}$$

The function  $f_W(w)$  is shown in Fig. 13.3.3.

The distribution function  $F_W(w)$  is obtained by integrating  $f_W(w)$  and is given by

$$F_W(w) = \begin{cases} \frac{3}{5}(e^{2w}) & w \leq 0 \\ \frac{3}{5} + \frac{2}{5}(1 - e^{-3w}) & w > 0 \end{cases}$$

It is shown in Fig. 13.3.4.

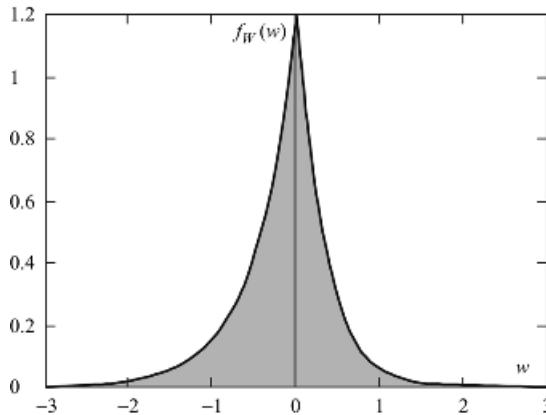


FIGURE 13.3.3

**Example 13.3.2** Two independent positive random variables  $X$  and  $Y$  are jointly exponentially distributed with density function  $f_{XY}(x,y) = 6e^{-(3x+2y)} u(x)u(y)$ , exactly as in the previous example. Two other random variables are  $Z$  and  $W$ , given by,  $Z = 2X/(X+Y)$  and  $W = X+Y$ . We have to find the joint density  $f_{ZW}(z,w)$  and the marginal densities  $f_Z(z)$  and  $f_W(w)$ .

1. All partials exist for  $g(x,y) = 2x/(x+y)$  and  $h(x,y) = x+y$ .
2. Solving for  $2x/(x+y) = z$  and  $x+y = w$ , we have  $x = zw/2$  and  $y = [1 - (z/2)]w$ . Since  $x$  and  $y$  are positive,  $w > 0$  and  $0 < z \leq 2$ .
3. The Jacobian matrix is

$$J = \begin{bmatrix} \frac{2y}{(x+y)^2} & \frac{-2x}{(x+y)^2} \\ 1 & 1 \end{bmatrix}$$

and the determinant is

$$|J| = \frac{2}{x+y} = \frac{2}{w}$$

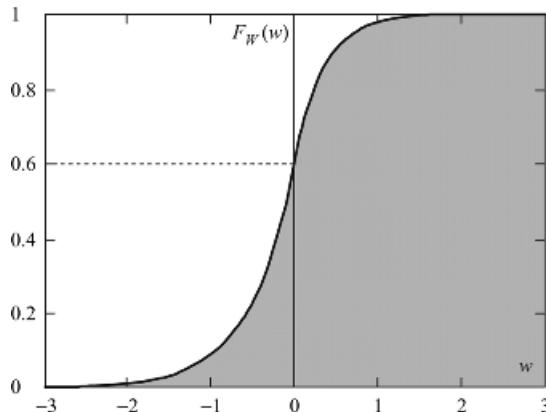


FIGURE 13.3.4

4. The joint density function  $f_{ZW}(z,w)$  from Eq. (13.3.5) is given by

$$f_{ZW}(z,w) = \frac{w}{2} 6 \exp - \left\{ \left[ 3 \frac{zw}{2} + 2(1 - \frac{z}{2})w \right] \right\} = 3we^{-w[2+(z/2)]}; \quad w > 0, 0 < z \leq 2$$

Integrating over all  $w$ , we obtain  $f_Z(z)$

$$f_Z(z) = \int_0^\infty 3we^{-w[2+(z/2)]} dw = \frac{12}{(z+4)^2}; \quad 0 < z \leq 2$$

and the corresponding distribution function  $F_Z(z)$  is

$$F_Z(z) = \begin{cases} 0: & z \leq 0 \\ 3\left(1 - \frac{4}{z+4}\right): & 0 < z \leq 2 \\ 1: & z > 2 \end{cases}$$

The functions  $f_Z(z)$  and  $F_Z(z)$  are shown in Fig. 13.3.5.

Integrating over all  $z$ , we obtain  $f_W(w)$

$$f_W(w) = \int_0^2 3w e^{-w[2+(z/2)]} dz = 6(e^{-2w} - e^{-3w}); \quad w > 0$$

and this result is the same one obtained in the previous example and shown in Fig. 13.3.2.

Integrating  $f_W(w)$ , we obtain  $F_W(w)$ :

$$F_W(w) = 1 + 2e^{-3w} - 3e^{-2w}; \quad w > 0$$

**Example 13.3.3** We will find  $f_{ZW}(z,w)$  from  $f_{XY}(x,y)$  for Example 13.2.3, where  $Z = X^2 + Y^2$  and  $W = Y/X$  and the joint density  $f_{XY}(x,y) = (1/2\pi\sigma^2)e^{-[(x^2+y^2)/2\sigma^2]}$ .

1. All partials exist for  $g(x,y) = x^2 + y^2$  and  $h(x,y) = y/x$ .
2. We solve  $x^2 + y^2 = z$  and  $y/x = w$ . Substituting  $y = wx$  in  $x^2 + y^2 = z$ , we obtain

$$x^2(1+w^2) = z \quad \text{or} \quad x = \pm \sqrt{\frac{z}{1+w^2}} \quad \text{and} \quad y = \pm \sqrt{\frac{w^2 z}{1+w^2}}$$

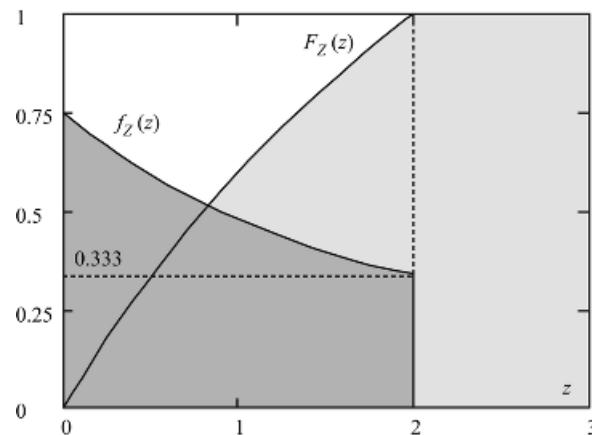


FIGURE 13.3.5

or

$$\begin{bmatrix} (x_1, y_1) \\ (x_2, y_2) \end{bmatrix} = \begin{bmatrix} \left( \sqrt{\frac{z}{1+w^2}}, \sqrt{\frac{w^2 z}{1+w^2}} \right); & z > 0, -\infty < w < \infty \\ \left( -\sqrt{\frac{z}{1+w^2}}, -\sqrt{\frac{w^2 z}{1+w^2}} \right); & z > 0, -\infty < w < \infty \end{bmatrix}$$

3. The Jacobian matrix is given by

$$J = \begin{bmatrix} 2x & 2y \\ -\frac{y}{x^2} & \frac{1}{x} \end{bmatrix}$$

and the Jacobian determinant

$$|J| = 2 \left( 1 + \frac{y^2}{x^2} \right) = 2(1 + w^2)$$

4. The joint density  $f_{ZW}(z,w)$  is given by

$$f_{ZW}(z,w) = \frac{1}{2(1+w^2)} \left\{ \frac{1}{2\pi\sigma^2} e^{-(z/2\sigma^2)} + \frac{1}{2\pi\sigma^2} e^{-(z/2\sigma^2)} \right\}$$

or

$$f_{ZW}(z,w) = \frac{1}{\pi(1+w^2)} \left\{ \frac{1}{2\sigma^2} e^{-(z/2\sigma^2)} \right\}; \quad z > 0, -\infty < w \leq \infty$$

Integrating this equation with respect to  $w$ , the marginal density  $f_Z(z)$  is

$$\begin{aligned} f_Z(z) &= \frac{1}{2\sigma^2} e^{-(z/2\sigma^2)} \int_{-\infty}^{\infty} \frac{dw}{\pi(1+w^2)} \\ &= \frac{1}{2\sigma^2} e^{-(z/2\sigma^2)} \frac{\tan^{-1}(w)}{\pi} \Big|_{-\infty}^{\infty} = \frac{1}{2\sigma^2} e^{-1(z/2\sigma^2)}; \quad z > 0 \end{aligned}$$

and by a similar integration with respect to  $z$  we obtain  $f_W(w)$  as

$$f_W(w) = \frac{1}{\pi(1+w^2)}$$

which is a Cauchy density. These results correspond to the results obtained in Example 13.2.3.

**Example 13.3.4** Two independent zero mean Gaussian random variables  $X$  and  $Y$  has joint density function  $f_{XY}(x,y)$  given by

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2} e^{-(1/2)[(x^2+y^2)/\sigma^2]}$$

Two other random variables  $Z$  and  $W$  are defined by

$$Z = \sqrt{X^2 + Y^2}; \quad W = \begin{cases} \tan^{-1}\left(\frac{Y}{X}\right) & \text{if } X > 0 \\ \pi + \tan^{-1}\left(\frac{Y}{X}\right) & \text{if } X \leq 0 \end{cases}$$

The angle  $W$  has been defined in this manner so that it covers the region  $-(\pi/2), (\pi/2)$  since  $\tan^{-1}(Y/X)$  covers only the region  $-(\pi/2), (\pi/2)$ . We have to find the joint density  $f_{ZW}(z, w)$  and the marginal densities  $f_Z(z)$  and  $f_W(w)$ . We apply the usual four steps to find a solution to this problem.

1. All partials exist for  $g(x, y) = \sqrt{x^2 + y^2}$  and  $h(x, y) = \tan^{-1}(y/x)$ .
2. We solve for  $\sqrt{x^2 + y^2} = z$  and  $\tan^{-1}(y/x) = w$  for  $x > 0$  and  $\pi + \tan^{-1}(y/x) = w$  for  $x \leq 0$ . We note that  $\tan(w) = y/x$  for both  $x \leq 0$  and  $x > 0$ . Squaring  $\sqrt{x^2 + y^2} = z$  and substituting  $y/x = \tan(w)$ , we can write

$$x^2 \left(1 + \frac{y^2}{x^2}\right) = z^2 \quad \text{or} \quad x^2 (1 + \tan^2(w)) = z^2 \quad \text{or} \quad x^2 = z^2 \cos^2(w)$$

However, for  $x > 0$ ;  $-(\pi/2) < w \leq (\pi/2)$  and  $\cos(w) > 0$  and for  $x \leq 0$ ;  $(\pi/2) < w \leq (3\pi/2)$  and  $\cos(w) \leq 0$ . Hence, the only solutions are  $x = z \cos(w)$  and  $y = z \sin(w)$  in the range  $-(\pi/2) < w \leq (3\pi/2)$ .

3. The Jacobian matrix and its determinant are given by

$$J = \begin{bmatrix} \frac{x}{\sqrt{x^2 + y^2}} & \frac{y}{\sqrt{x^2 + y^2}} \\ \frac{-y}{x^2 + y^2} & \frac{-x}{x^2 + y^2} \end{bmatrix} \quad \text{and} \quad |J| = \frac{1}{\sqrt{x^2 + y^2}} = \frac{1}{z}$$

4. The joint density  $f_{ZW}(z, w)$  is given by

$$f_{ZW}(z, w) = \frac{1}{2\pi} \frac{z}{\sigma^2} e^{-(1/2)(z^2/\sigma^2)}, \quad z > 0$$

Integrating with respect to  $w$  in the range  $-(\pi/2) < w \leq (3\pi/2)$ , we obtain

$$f_Z(z) = \frac{z}{\sigma^2} e^{-(1/2)(z^2/\sigma^2)}, \quad z > 0$$

which is a Rayleigh distribution, and the angle  $w$  is uniformly distributed in  $-(\pi/2) < w \leq (3\pi/2)$ ,  
or

$$f_W(w) = \frac{1}{2\pi}; \quad -\frac{\pi}{2} < w \leq \frac{3\pi}{2}$$

and  $Z$  and  $W$  are also independent random variables.

## 13.4 SOLVING $Z = g(X, Y)$ USING AN AUXILIARY RANDOM VARIABLE

The problem of finding the density function  $f_Z(z)$  of a function of two random variables  $Z = g(X, Y)$  can be considerably simplified by converting to two functions of two random variables with the use of an auxiliary random variable  $W = X$  or  $Y$ , provided  $g(x, y)$  satisfies the two restrictions of differentiability and no straight-line solutions.

**Example 13.4.1** We will now solve Example 13.1.10 using an auxiliary random variable. Here  $g(X, Y) = Z = XY$  and we introduce an auxiliary random variable  $W = h(X, Y) = X$ .

1. The restrictions on  $g(x,y) = xy$  and  $h(x,y) = x$  are satisfied.
2. Solving for  $xy = z$  and  $x = w$ , we obtain the real solution,  $x = w$  and  $y = z/w$ .
3. The Jacobian matrix and its absolute determinant are

$$J = \begin{bmatrix} y & x \\ 1 & 0 \end{bmatrix} \quad \text{or} \quad \|J\| = |x| = |w|$$

4. The joint density  $f_{ZW}(z,w)$  is given by

$$f_{ZW}(z,w) = \frac{1}{|w|} f_{XY}\left(w, \frac{z}{w}\right)$$

The function  $f_Z(z)$  is obtained by integrating this equation with respect to  $w$

$$F_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|w|} f_{XY}\left(w, \frac{z}{w}\right) dw$$

which is the same as Eq. (13.1.33).

**Example 13.4.2** This is the same as Example 13.1.7, where  $Z = g(X,Y) = \sqrt{X^2 + Y^2}$ . We introduce the auxiliary random variable  $W = h(X,Y) = X$  and proceed to solve it using the techniques of the previous section. The joint density  $f_{XY}(x,y)$  is given by

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2} e^{-(1/2)[(x^2+y^2)/\sigma^2]}$$

1. The restrictions on  $g(x,y) = \sqrt{x^2 + y^2}$  and  $h(x,y) = x$  are satisfied.
2. Solving for  $\sqrt{x^2 + y^2} = z$  and  $x = w$ , we obtain  $x^2 + y^2 = z^2$  or  $y^2 = z^2 - w^2$ , and the two solutions are  $x_1 = w$ ,  $y_1 = \sqrt{z^2 - w^2}$  and  $x_2 = w$ ,  $y_2 = -\sqrt{z^2 - w^2}$  with  $-z < w \leq z$  for real  $y_1$  and  $y_2$ .
3. The Jacobian matrix and its absolute determinant are

$$J = \begin{bmatrix} x & y \\ \sqrt{x^2 + y^2} & \sqrt{x^2 + y^2} \\ 1 & 0 \end{bmatrix} \quad \text{or} \quad \|J\| = \frac{y}{\sqrt{x^2 + y^2}} = \frac{\sqrt{z^2 - w^2}}{z}; \quad |w| \leq z$$

4. The joint density function  $f_{ZW}(z,w)$  is given by

$$f_{ZW}(z,w) = \frac{z}{\sqrt{z^2 - w^2}} \frac{2}{2\pi\sigma^2} e^{-(1/2)(z^2/\sigma^2)}; \quad z > 0, \quad -z < w \leq z$$

Integrating with respect to the variable  $w$ , we obtain

$$f_Z(z) = \frac{z}{\sigma^2} e^{-(1/2)(z^2/\sigma^2)} \frac{1}{\pi} \int_{-z}^z \frac{dw}{\sqrt{z^2 - w^2}}; \quad z > 0$$

Since the integrand is even, we can write this equation as

$$\begin{aligned} f_Z(z) &= \frac{z}{\sigma^2} e^{-(1/2)(z^2/\sigma^2)} \frac{2}{\pi} \int_0^z \frac{dw}{\sqrt{z^2 - w^2}}; \quad z > 0 \\ &= \frac{z}{\sigma^2} e^{-(1/2)(z^2/\sigma^2)} \frac{2}{\pi} \frac{\pi}{2} = \frac{z}{\sigma^2} e^{-(1/2)(z^2/\sigma^2)}; \quad z > 0 \end{aligned}$$

which is the same as the result obtained in Eq. (13.1.22).

**Example 13.4.3** This is the same as Example 13.1.8, where  $Z = g(X, Y) = X^2 + Y^2$ . We introduce the auxiliary random variable  $W = h(X, Y) = X$  and proceed to solve it using the techniques of the previous section. The joint density  $f_{XY}(x, y)$  is given by

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma^2} e^{-(1/2)[(x^2+y^2)/\sigma^2]}$$

1. The restrictions on  $g(x, y) = x^2 + y^2$  and  $h(x, y) = x$  are satisfied.
2. Solving for  $x^2 + y^2 = z$  and  $x = w$ , we obtain from  $Y^2 = z - x^2$ , the two solutions  $x_1 = w, y_1 = \sqrt{z - w^2}$  and  $x_2 = w, y_2 = -\sqrt{z - w^2}$ , with  $-\sqrt{z} < w \leq \sqrt{z}$  for real  $y_1$  and  $y_2$ .
3. The Jacobian matrix and its absolute determinant are

$$J = \begin{bmatrix} 2x & 2y \\ 1 & 0 \end{bmatrix} \quad \text{or} \quad \|J\| = 2y = 2\sqrt{z - w^2}; \quad z > 0, |w| \leq \sqrt{z}$$

4. The joint density function  $f_{ZW}(z, w)$  corresponding to the two solutions is given by,

$$f_{ZW}(z, w) = \frac{1}{2\sqrt{z - w^2}} \frac{2}{2\pi\sigma^2} e^{-(1/2)(z/\sigma^2)}; \quad z > 0, -\sqrt{z} < w \leq \sqrt{z}$$

Integrating with respect to the variable  $w$ , we obtain

$$f_Z(z) = \frac{1}{2\sigma^2} e^{-(1/2)(z/\sigma^2)} \int_{-\sqrt{z}}^{\sqrt{z}} \frac{dw}{\pi\sqrt{z - w^2}}; \quad z > 0$$

Substituting  $w = \sqrt{z} \sin(\theta)$  and  $dw = \sqrt{z} \cos(\theta)d\theta$ , we obtain

$$\begin{aligned} f_Z(z) &= \frac{1}{2\sigma^2} e^{-(1/2)(z/\sigma^2)} \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sqrt{z} \cos(\theta)d\theta}{\sqrt{z \cos^2(\theta)}}; \quad z > 0 \\ &= \frac{1}{2\sigma^2} e^{-(1/2)(z/\sigma^2)}; \quad z > 0 \end{aligned}$$

which is the same result obtained in Eq. (13.1.24).

**Example 13.4.4** The random variable  $Z$  is given by  $Z = (X + Y)^2$ , where  $X$  and  $Y$  are zero mean independent Gaussian random variables given by the joint density function

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma^2} e^{-(1/2)[(x^2+y^2)/\sigma^2]}$$

We have to find the density function  $f_Z(z)$  using auxiliary random variable  $W = X$ .

1. The restrictions on  $g(x, y) = (x + y)^2$  and  $h(x, y) = x$  are satisfied.
2. Solving for  $(x + y)^2 = z$  and  $x = w$ , we obtain two solutions:  $x_1 = w, y_1 = \sqrt{z} - w$  and  $x_2 = w, y_2 = -\sqrt{z} - w$ .
3. The Jacobian matrix and its absolute determinant are

$$J = \begin{bmatrix} 2(x + y) & 2(x + y) \\ 1 & 0 \end{bmatrix} \quad \text{or} \quad \|J\| = 2\sqrt{z}; \quad z > 0$$

4. The joint density  $f_{ZW}(z,w)$  corresponding to the two solutions is

$$f_{ZW}(z,w) = \frac{1}{2\sqrt{z}} \frac{1}{2\pi\sigma^2} e^{-(1/2\sigma^2)[w^2 + (\sqrt{z}-w)^2]} \\ + \frac{1}{2\sqrt{z}} \frac{1}{2\pi\sigma^2} e^{-(1/2\sigma^2)[w^2 + (-\sqrt{z}-w)^2]}, \quad z > 0$$

We will now simplify this expression. Completing the squares in the first and the second terms on the righthand side of the equation, we obtain

$$\frac{1}{2\sqrt{z}} \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}[w^2 + (\sqrt{z}-w)^2]\right\} \\ = \frac{1}{2\sqrt{z}} \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{2}{2\sigma^2} \left[ \left(w - \frac{\sqrt{z}}{2}\right)^2 + \frac{z}{4} \right]\right\} \\ \frac{1}{2\sqrt{z}} \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}[w^2 + (-\sqrt{z}-w)^2]\right\} \\ = \frac{1}{2\sqrt{z}} \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{2}{2\sigma^2} \left[ \left(w + \frac{\sqrt{z}}{2}\right)^2 + \frac{z}{4} \right]\right\}$$

Hence the joint density  $f_{ZW}(z,w)$  is given by

$$f_{ZW}(z,w) = \frac{1}{2\sqrt{z}} \frac{1}{2\pi\sigma^2} e^{-(z/4\sigma^2)} \left( \left\{ e^{-(1/\sigma^2)[w - (\sqrt{z}/2)]^2} \right. \right. \\ \left. \left. + e^{-(1/\sigma^2)[w + (\sqrt{z}/2)]^2} \right\}; \quad z > 0, \quad -\infty < w < \infty \right)$$

We integrate this equation over all  $w$  to obtain  $f_Z(z)$ :

$$f_Z(z) = \frac{1}{2\sqrt{z}} \frac{1}{2\pi\sigma^2} e^{-(z/4\sigma^2)} \int_{-\infty}^{\infty} \left\{ e^{-(1/\sigma^2)[w - (\sqrt{z}/2)]^2} + e^{-(1/\sigma^2)[w + (\sqrt{z}/2)]^2} \right\} dw \\ = \frac{1}{2\sqrt{z}} \frac{1}{2\pi\sigma^2} e^{-(z/4\sigma^2)} [\sigma\sqrt{\pi} + \sigma\sqrt{\pi}] = \frac{1}{2\sigma\sqrt{\pi}} \frac{e^{-(z/4\sigma^2)}}{\sqrt{z}}; \quad z > 0$$

The distribution function  $F_Z(z)$  is obtained by integrating  $f_Z(z)$  over  $z$ :

$$F_Z(z) = \frac{1}{2\sigma\sqrt{\pi}} \int_0^z \frac{e^{(\zeta/4\sigma^2)}}{\sqrt{\zeta}} d\zeta = \operatorname{erf}\left(\frac{\sqrt{z}}{2\sigma}\right)$$

where  $\operatorname{erf}(z)$  is defined by  $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-\xi^2} d\xi$  and

$$\operatorname{erf}(\sqrt{z}) = \frac{1}{\sqrt{\pi}} \int_0^z \frac{e^{-\xi}}{\sqrt{\xi}} d\xi$$

Functions  $f_Z(z)$  and  $F_Z(z)$  are shown in Fig.13.4.1.

**Example 13.4.5** In this example  $X$  is a zero mean unit variance random variable given by the density function

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

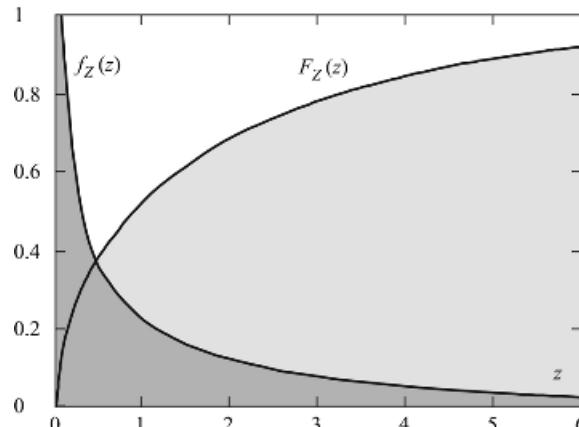


FIGURE 13.4.1

and  $Y$  is an independent chi-square random variable with  $v$  degrees of freedom whose density is given by Eq. (7.7.2):

$$f_Y(y) = \frac{y^{(v/2)-1} e^{-(y/2)}}{2^{v/2} \Gamma\left(\frac{v}{2}\right)} u(y)$$

The density function of the random variable  $Z = X\sqrt{Y/v}$  is called the *Student-t* distribution (Section 7.9) and finds extensive application in statistical analyses. We will use an auxiliary variable  $W = Y$ :

1. The restrictions on  $g(x, y) = x/\sqrt{y/v}$  and  $h(x, y) = y$  are satisfied.
2. Solving for  $x/\sqrt{y/v} = z$  and  $y = w$ , we obtain the solution  $x = z\sqrt{w/v}$  and  $y = w$ .
3. The Jacobian matrix and its absolute determinant are

$$J = \begin{bmatrix} \frac{1}{\sqrt{y/v}} & \frac{-x}{2v(y/v^{3/2})} \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad \|J\| = \frac{1}{\sqrt{y/v}} = \frac{1}{\sqrt{w/v}}; \quad w > 0$$

4. The density function  $f_{ZW}(z, w)$  is given by

$$\begin{aligned} f_{ZW}(z, w) &= \frac{\sqrt{w/v}}{\sqrt{2\pi}} e^{-(z^2 w/v^2)} \frac{w^{(v/2)-1} e^{-(w/2)}}{2^{v/2} \Gamma\left(\frac{v}{2}\right)} u(w) \\ &= \frac{\left(\frac{w}{2}\right)^{(v-1)/2} e^{-(w/2)[(z^2/v)+1]}}{2\sqrt{\pi v} \Gamma\left(\frac{v}{2}\right)} u(w) \end{aligned}$$

The density function  $f_Z(z)$  is found by integrating this equation over  $w$ ,

$$f_Z(z) = \frac{1}{2\sqrt{\pi v} \Gamma\left(\frac{v}{2}\right)} \int_0^\infty \left(\frac{w}{2}\right)^{(v-1)/2} e^{-(w/2)[(z^2/v)+1]} dw$$

This equation can be rewritten by substituting  $\xi = (w/2)[(z^2/v) + 1]$ , resulting in

$$f_Z(z) = \frac{1}{\sqrt{\pi v} \Gamma\left(\frac{v}{2}\right) \left(\frac{z^2}{v} + 1\right)^{(v+1)/2}} \int_0^\infty \xi^{(v-1)/2} e^{-\xi} d\xi$$

The integral in this equation is a gamma function given by  $\Gamma[(v+1)/2]$ . Thus

$$f_Z(z) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v} \Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{z^2}{v}\right)^{-(v+1)/2}$$

which is a Student-*t* density [Eq. (7.9.2)].

## 13.5 MULTIPLE FUNCTIONS OF RANDOM VARIABLES

We can now generalize the results of the previous section to  $n$  functions of  $n$  random variables. Let  $\{X_1, \dots, X_n\}$  be the  $n$  random variables defined by the joint density function  $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ . Let the  $n$  functions be

$$Z_1 = g_1(X_1, \dots, X_n), Z_2 = g_2(X_1, \dots, X_n), \dots, Z_n = g_n(X_1, \dots, X_n) \quad (13.5.1)$$

We assume that these functions satisfy the restrictions of existence of partial derivatives and no planar solutions. We now solve for

$$g_1(x_1, \dots, x_n) = z_1, g_2(x_1, \dots, x_n) = z_2, \dots, g_n(x_1, \dots, x_n) = z_n \quad (13.5.2)$$

Let the solutions be  $\{x_1, \dots, x_n\}$  be expressed as functions of  $\{z_1, \dots, z_n\}$ :

$$x_1 = h_1(z_1, \dots, z_n), x_2 = h_2(z_1, \dots, z_n), \dots, x_n = h_n(z_1, \dots, z_n) \quad (13.5.3)$$

Let the Jacobian matrix be given by

$$J(x_1, \dots, x_n) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_n}{\partial x_1} & \dots & \frac{\partial g_n}{\partial x_n} \end{bmatrix} \quad (13.5.4)$$

and its determinant be expressed in terms of  $\{z_1, \dots, z_n\}$ :

$$\|J\| = \|J(z_1, \dots, z_n)\| \quad (13.5.5)$$

The joint density function  $f_{Z_1, \dots, Z_n}(z_1, \dots, z_n)$  can now be given in terms of Eqs. (13.5.3) and (13.5.5):

$$\begin{aligned} f_{Z_1, \dots, Z_n}(z_1, \dots, z_n) &= \frac{f_{X_1, \dots, X_n}(x_1, \dots, x_n)}{\|J(z_1, \dots, z_n)\|} \\ &= \frac{f_{X_1, \dots, X_n}[h_1(z_1, \dots, z_n)_1, \dots, h_n(z_1, \dots, z_n)]}{\|J(z_1, \dots, z_n)\|} \end{aligned} \quad (13.5.6)$$

For general functions  $\{g_k(x_1, \dots, x_n, k = 1, \dots, n)\}$ ,  $\{h_k(z_1, \dots, z_n, k = 1, \dots, n)\}$  and  $\|J(z_1, \dots, z_n)\|$  are not that easy to solve. However, for linear functions

$\{g_k(x_1, \dots, x_n, k = 1, \dots, n)\}$ , the solutions are not that difficult, as shown by the following example.

**Example 13.5.1** Three independent zero mean unit variance Gaussian random variables  $X$ ,  $Y$ , and  $Z$  are given by the joint density function

$$f_{XYZ}(x, y, z) = \frac{1}{2\pi\sqrt{2\pi}} e^{-(1/2)(x^2+y^2+z^2)}$$

Three other random variables  $U$ ,  $V$ , and  $W$  are given by

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = \begin{bmatrix} 2 & 2 & 1 \\ 5 & 3 & 3 \\ 3 & 2 & 2 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

We have to find the joint density  $f_{UVW}(u, v, w)$  and the marginal densities  $f_U(u)$ ,  $f_V(v)$ , and  $f_W(w)$ .

The solution to

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 2 & 2 & 1 \\ 5 & 3 & 3 \\ 3 & 2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 & 2 & -3 \\ 1 & -1 & 1 \\ -1 & -2 & 4 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 2v - 3w \\ u - v + w \\ -u - 2v + 4w \end{bmatrix}$$

The Jacobian matrix and its determinant are given by

$$J = \begin{bmatrix} 2 & 2 & 1 \\ 5 & 3 & 3 \\ 3 & 2 & 2 \end{bmatrix}; \quad \|J\| = 1$$

Hence, using Eq. (13.5.6), we obtain the joint density function  $f_{UVW}(u, v, w)$

$$\begin{aligned} f_{UVW}(u, v, w) &= \frac{1}{12\pi\sqrt{2\pi}} e^{-(1/2)[(2v-3w)^2 + (u-v+w)^2 + (-u-2v+4w)^2]} \\ &= \frac{1}{2\pi\sqrt{2\pi}} e^{-(1/2)[(2u^2 - 9v^2 + 26w^2 + 2uv - 30vw - 6wu)]} \\ &= \frac{1}{(2\pi)^{3/2} |\mathbf{C}|^{1/2}} \exp\left\{-\frac{1}{2} \left( \begin{bmatrix} u & v & w \end{bmatrix} \mathbf{C}^{-1} \begin{bmatrix} u \\ v \\ w \end{bmatrix} \right) \right\} \end{aligned}$$

where the matrix  $\mathbf{C} = \begin{bmatrix} 9 & 19 & 12 \\ 19 & 43 & 27 \\ 12 & 27 & 17 \end{bmatrix}$  and  $|\mathbf{C}| = 1$ .

The random variables  $X$ ,  $Y$ ,  $Z$  are independent, whereas the transformed random variables  $U$ ,  $V$ ,  $W$  are not independent. Their correlation coefficients can be calculated as follows:

$$\rho_{UV} = \frac{19}{\sqrt{9 \cdot 43}} = 0.966, \quad \rho_{VW} = \frac{27}{\sqrt{43 \cdot 17}} = 0.999, \quad \rho_{WU} = \frac{12}{\sqrt{17 \cdot 9}} = 0.970$$

By successively integrating with respect to  $(v, w)$ ,  $(w, u)$  and  $(u, v)$ , we obtain the marginal density functions  $f_U(u)$ ,  $f_V(v)$ , and  $f_W(w)$  as

$$\begin{aligned}
 f_U(u) &= \frac{1}{2\pi\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{[-(1/2)(2u^2+9v^2+26w^2+2uv-30vw-6wu)]} dv dw \\
 &= \frac{1}{3\sqrt{2\pi}} e^{[-(1/2)(u^2/9)]} \\
 f_V(v) &= \frac{1}{2\pi\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{[-(1/2)(2u^2+9v^2+26w^2+2uv-30vw-6wu)]} dw du \\
 &= \frac{1}{43\sqrt{2\pi}} e^{[-(1/2)(v^2/43)]} \\
 f_W(w) &= \frac{1}{2\pi\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{[-(1/2)(2u^2+9v^2+26w^2+2uv-30vw-6wu)]} du dv \\
 &= \frac{1}{17\sqrt{2\pi}} e^{[-(1/2)(w^2/17)]}
 \end{aligned}$$

These results can be verified from the results of Example 13.1.5. Since  $\text{var}(ax) = a^2 \text{var}(X)$ , we have  $\text{var}(U) = 2^2 + 2^2 + 1^2 = 9$ ,  $\text{var}(V) = 5^2 + 3^2 + 3^2 = 43$ , and  $\text{var}(W) = 3^2 + 2^2 + 2^2 = 17$ , which checks with the results above.

# Inequalities, Convergences, and Limit Theorems

## 14.1 DEGENERATE RANDOM VARIABLES

In Section 10.2 it was stated that if the variance is zero, then the random variable will no longer be random. We can state this concept more fully as follows.

If  $E(X - a)^2 = 0$ , where  $a$  is a constant, then the random variable  $X$  equals  $a$  with probability 1. We shall show this result as follows. We construct a small hole of nonzero width  $2\epsilon$  with  $\epsilon > 0$  about the value  $a$  in the region of integration of  $x$  from  $-\infty$  to  $\infty$  as shown in Fig. 14.1.1.

We integrate  $(x - a)^2 f_X(x)$  from  $-\infty$  to  $\infty$  except for the hole about  $a$  and write

$$E(X - a)^2 = \int_{-\infty}^{\infty} (x - a)^2 f_X(x) dx \geq \int_{|x-a|>\epsilon} (x - a)^2 f_X(x) dx \quad (14.1.1)$$

By substituting the lowest value of  $\epsilon$  for  $|x-a|$  in Eq. (14.1.1), we can write

$$E(X - a)^2 \geq \int_{|x-a|>\epsilon} \epsilon^2 f_X(x) dx = \epsilon^2 \int_{|x-a|>\epsilon} f_X(x) dx = \epsilon^2 P(|X - a| > \epsilon)$$

Since  $E(X - a)^2 = 0$  by assumption, we must have  $\epsilon^2 P(|X - a| > \epsilon) = 0$ . Since  $\epsilon$  is a nonzero constant, we must have  $P(|X - a| > \epsilon) = 0$ . As a consequence

$$P(|X - a| = 0) = 1 \quad \text{and} \quad X = a \text{ with probability 1.} \quad (14.1.2)$$

What this result conveys is that if  $E(X - a)^2 = 0$ , then almost all realizations of  $X$  will be equal to a constant  $a$ . The random variable  $X$  is called a *degenerate random variable*. This method of analysis will be useful in showing some of the inequalities that will follow.

**Example 14.1.1** If the variance  $\sigma_X^2 = 0$ , then the random variable  $X = \mu_X$  almost certainly. Since variance of  $X$  equals  $E(X - \mu_X)^2 = 0$ , then, by Eq. (14.1.2)  $X = \mu_X$  with probability 1.

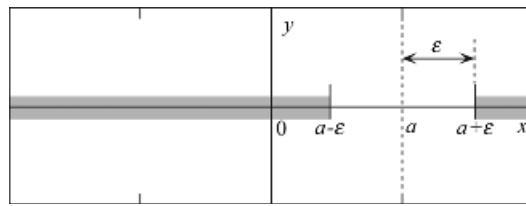


FIGURE 14.1.1

## 14.2 CHEBYSHEV AND ALLIED INEQUALITIES

In many problems in communications we are interested in finding bounds on certain performance measures. The idea in using a bound is to substitute a simpler expression for something more complicated, thus gaining some insight into the problem at the expense of accuracy. This process usually requires a fair amount of experience with the results that are to be expected.

One of the bounds that we are interested in is the tails probabilities of the type  $P(X > \varepsilon)$  or  $P(|X| > \varepsilon)$  (Fig. 14.2.1). Even if we cannot find these probabilities exactly, we should be able to find bounds on these quantities. These bounds are based on partial information about the random variable, such as mean and variance.

### Chebyshev Bound

The tails probability bound of the random variable  $X$  of the form  $P(|X - \mu_X| < \varepsilon)$  is based on the knowledge of only the mean and the variance of  $X$  and is known as the *Chebyshev inequality*, given by

$$P\{|X - \mu_X| \geq \varepsilon\} \leq \frac{\sigma_X^2}{\varepsilon^2} \quad \text{or} \quad CB(\varepsilon) = \frac{\sigma_X^2}{\varepsilon^2} \quad (14.2.1)$$

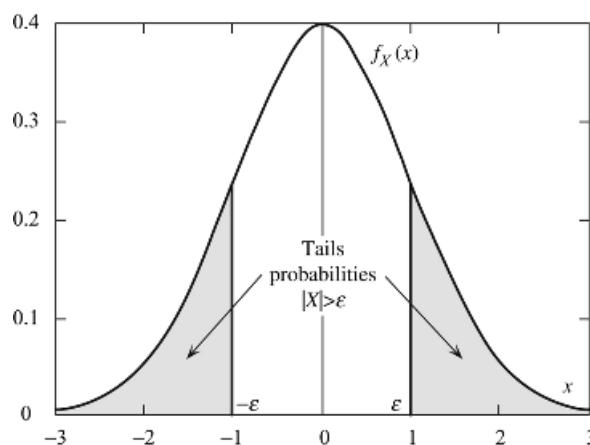


FIGURE 14.2.1

where  $\varepsilon$  is a positive constant  $>0$  and  $CB(\varepsilon)$  is the Chebyshev bound (this result can be shown using an analysis similar to that in the previous section):

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx \geq \int_{|x - \mu_X| > \varepsilon} (x - \mu_X)^2 f_X(x) dx \quad (14.2.2)$$

By substituting the lowest value of  $\varepsilon$  in Eq. (14.2.2), we obtain

$$\sigma_X^2 \geq \varepsilon^2 \int_{|x - \mu_X| > \varepsilon} f_X(x) dx = \varepsilon^2 P(|x - \mu_X| \geq \varepsilon) \quad (14.2.3)$$

and hence Eq. (14.2.1).

**Geometric Derivation of the Chebyshev Bound** We will now derive the Chebyshev bound from a different perspective. We will define a random variable  $Y = X - \mu_X$ . We will make use of the indicator functions defined in Eq. (12.5.7). Let  $A$  be a set on the real line such that  $A = \{-\infty, -\varepsilon\} \cup \{\varepsilon, \infty\}$ . We define an indicator function  $I_A$  as follows:

$$I_A = \begin{cases} 1 & \text{if } Y \in A \\ 0 & \text{if } Y \notin A \end{cases} \quad (14.2.4)$$

The indicator function  $I_A(y)$  is shown in Fig. 14.2.2.

Taking the expected value of  $I_A$ , we have

$$\begin{aligned} E[I_A] &= 1 \cdot P(Y \in A) + 0 \cdot P(Y \notin A) = P(Y \in A) \\ &= P(|Y| \geq \varepsilon) = P(|X - \mu_X| \geq \varepsilon) \end{aligned} \quad (14.2.5)$$

We will now bound  $I_A$  by another random variable  $Y^2/\varepsilon^2$  as shown in Fig. 14.2.2.

From the figure we can see that

$$I_A \leq \frac{Y^2}{\varepsilon^2} = \frac{|X - \mu_X|^2}{\varepsilon^2} \quad (14.2.6)$$

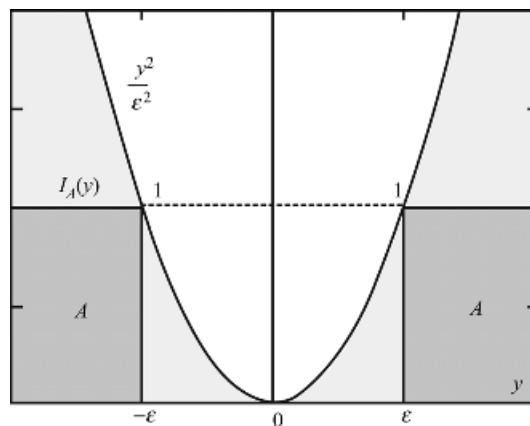


FIGURE 14.2.2

Taking expectation on both sides of Eq. (14.2.6), we have

$$E[I_A] = P(|X - \mu_X| \geq \varepsilon) \leq \frac{E[Y^2]}{\varepsilon^2} = \frac{E|X - \mu_X|^2}{\varepsilon^2} = \frac{\sigma_X^2}{\varepsilon^2} \quad (14.2.7)$$

which is Chebyshev inequality. The bound is  $E[Y^2/\varepsilon^2]$  is called the *Chebyshev bound*. We will use this technique for deriving the Chernoff bound.

**Alternate Forms of the Chebyshev Bound** By rearranging the terms in Eq. (14.2.1), we can rewrite Chebyshev's inequality in terms of the center probability:

$$P\{|X - \mu_X| < \varepsilon\} \geq 1 - \frac{\sigma_X^2}{\varepsilon^2} \quad (14.2.8)$$

By substituting  $\varepsilon = n\sigma_X$ , we can write Eq. (14.2.1) as follows:

$$P\{|X - \mu_X| \geq n\sigma_X\} \leq \frac{1}{n^2} \quad (14.2.9)$$

Thus, for any random variable  $X$ , the probability of a deviation from the mean  $\mu_X$  of more than  $n$  standard deviations is  $1/n^2$ . For example, in a standard Gaussian if  $n = 3, 1/n^2 = \frac{1}{9} = 0.1111$ . Chebyshev inequality tells us that the probability of deviation by more than 3 from 0 in a standard Gaussian is 0.1111. In actual cases this probability is 0.0027.

We can also rewrite Eq. (14.2.5) analogous to Eq. (14.2.4) as

$$P\{|X - \mu_X| < n\sigma_X\} \geq 1 - \frac{1}{n^2} \quad (14.2.10)$$

### **Generalized Chebyshev Inequality**

If  $\phi(x)$  is a convex function (Section 14.6) in the sense that its second derivative is positive for all  $x$ , then the inequality in Eq. (14.2.1) can be generalized to

$$P\{|X| > \varepsilon\} \leq \frac{E[\phi(X)]}{\phi(\varepsilon)} \quad (14.2.11)$$

This can be shown by using techniques similar to those used in demonstrating Chebyshev inequality, applying the fact that  $\phi(x)$  is convex:

$$E[\phi(X)] = \int_{-\infty}^{\infty} \phi(x)f_X(x)dx \geq \int_{|x|>\varepsilon} \phi(x)f_X(x)dx \quad (14.2.12)$$

Using the least value for  $\phi(x)$ , namely,  $\phi(\varepsilon)$  in Eq. (14.2.12), we have

$$E[\phi(X)] \geq \int_{|x|>\varepsilon} \phi(\varepsilon)f_X(x)dx = \phi(\varepsilon)P\{|X| > \varepsilon\} \quad (14.2.9)$$

and hence Eq. (14.2.11). If  $\phi(X) = (X - \mu_X)^2$  in Eq. (14.2.11), we get Chebyshev inequality.

**Example 14.2.1** We will apply the Chebyshev bounds to two different distributions, a Gaussian distribution and a Laplace distribution with zero means and unit variances.

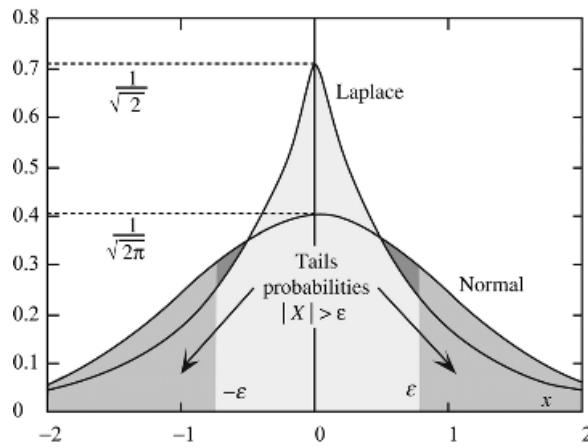


FIGURE 14.2.3

They are given by

$$f_N(x) = \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)}; \quad f_L(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}$$

and shown in Fig. 14.2.3.

We will find the tails probabilities  $P(|X| \geq \epsilon)$  for both the distributions and the Chebyshev bound, which will be the same for both since they have the same means and variances. Applying Eq. (14.2.1), the Chebyshev bounds are

$$P\{|X| \geq \epsilon\} \leq \frac{1}{\epsilon^2}$$

Thus the Chebyshev bound  $CB(\epsilon) = 1/\epsilon^2$ . Since we know both the density functions, we can calculate the exact tails probabilities  $P(|X| \geq \epsilon)$ , which are given by

$$F_N(\epsilon) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\epsilon} e^{-x^2/2} dx + \frac{1}{\sqrt{2\pi}} \int_{\epsilon}^{\infty} e^{-x^2/2} dx = 1 - \operatorname{erf}\left(\frac{\epsilon}{\sqrt{2}}\right)$$

$$F_L(\epsilon) = \frac{1}{\sqrt{2}} \int_{-\infty}^{-\epsilon} e^{\sqrt{2}x} dx + \frac{1}{\sqrt{2}} \int_{\epsilon}^{\infty} e^{-\sqrt{2}x} dx = e^{-\sqrt{2}\epsilon}$$

where

$$\operatorname{erf}\left(\frac{\epsilon}{\sqrt{2}}\right) = \sqrt{\frac{2}{\pi}} \int_0^{\epsilon} e^{-x^2/2} dx$$

The one-sided tails probabilities  $F_N(\epsilon)$  and  $F_L(\epsilon)$  for the Gaussian density and the Laplace density, and the Chebyshev bound  $CB(\epsilon)$  are shown in Fig. 14.2.4 and tabulated in Table 14.2.1 for  $\epsilon$  ranging from 1.5 to 5.

### Conclusions

It is clear from Fig. 14.2.4 and Table 14.2.1 that the Chebyshev bound is a very loose one. As  $\epsilon$  increases, the bound becomes somewhat tighter (Fig. 14.2.4). Table 14.2.1 also shows that the Chebyshev bound is the same for both density functions, which have the same mean and variance. In fact, the Chebyshev bound is applicable to *all* density

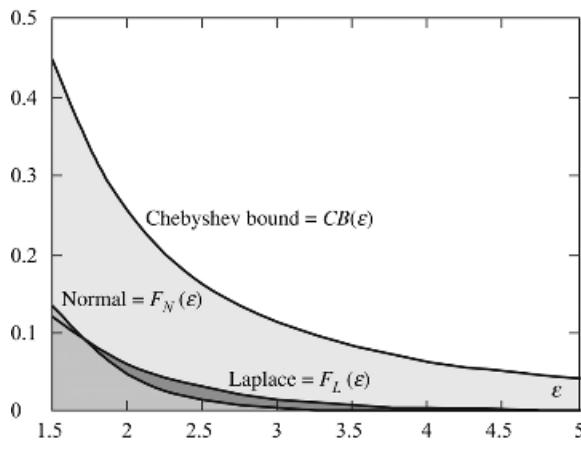


FIGURE 14.2.4

TABLE 14.2.1

$\varepsilon$	Normal $F_N(\varepsilon)$	Laplace $F_L(\varepsilon)$	Chebyshev $CB(\varepsilon) = 1/\varepsilon^2$
1.5	0.133614	0.119873	0.444444
1.75	0.080118	0.084174	0.326531
2	0.0455	0.059106	0.25
2.25	0.024449	0.041503	0.197531
2.5	0.012419	0.029143	0.16
2.75	0.00596	0.020464	0.132231
3	0.0027	0.01437	0.111111
3.25	0.001154	0.01009	0.094675
3.5	0.000465	0.007085	0.081633
3.75	0.000177	0.004975	0.071111
4	0.000063	0.003493	0.0625
4.25	0.000021	0.002453	0.055363
4.5	0.000007	0.001723	0.049383
4.75	0.000002	0.00121	0.044321
5	0.000001	0.000849	0.04

functions having the same means and variances. It is useful as a theoretical tool to determine the bounds of the tails probabilities, where only means and variances are known about the density functions.

### 14.3 MARKOV INEQUALITY

There are several different versions of Markov inequality. The usual version is defined for a positive random variable  $X$  with a finite mean  $\mu_X$ . It states that for a positive  $\varepsilon > 0$

$$P(X \geq \varepsilon) \leq \frac{\mu_X}{\varepsilon} \quad (14.3.1)$$

This result can be shown by the usual techniques developed in the previous section

$$\mu_X = \int_0^\infty xf_X(x)dx \geq \int_\varepsilon^\infty xf_X(x)dx \geq \varepsilon \int_\varepsilon^\infty f_X(x)dx = \varepsilon P(X \geq \varepsilon) \quad (14.3.2)$$

and hence Eq. (14.3.1). In Eq. (14.3.1), the quantity  $\mu_X/\varepsilon$  is called the *Markov bound*. Equation (14.3.1) can be slightly generalized to include both positive and negative random variables

$$P(|X - a| \geq \varepsilon) \leq \frac{E(|X - a|)}{\varepsilon} \quad (14.3.3)$$

where  $a$  is an arbitrary constant and  $\varepsilon > 0$ . Equation (14.3.3) can be further generalized to

$$P(|X - a| \geq \varepsilon) \leq \frac{E(|X - a|^k)}{\varepsilon^k} \quad (14.3.4)$$

where  $k$  is a positive integer. This inequality is called the *Bienayme inequality*.

We will derive the value of  $a$  for which  $P(|X - a| \geq \varepsilon)$  is a minimum in the equation

$$P(|X - a| \geq \varepsilon) \leq \frac{E[|X - a|^2]}{\varepsilon^2} \quad (14.3.5)$$

for which we have to minimize  $E[|X - a|^2]$  with respect to  $a$ :

$$E[|X - a|^2] = E[X^2 - 2Xa + a^2] = E[X^2] - 2aE[X] + a^2 \quad (14.3.6)$$

Differentiating Eq. (14.3.6) with respect to  $a$  and setting equal to 0, we have

$$\frac{d}{da} \{E[X^2] - 2aE[X] + a^2\} = 2a - 2E[X] = 0$$

from which we obtain  $a = E[X]$ . Thus we obtain Chebyshev inequality as follows:

$$P(|X - \mu_X| \geq \varepsilon) \leq \frac{E[|X - \mu_X|^2]}{\varepsilon^2} \quad (14.2.1)$$

**Example 14.3.1** A materials manager purchases 10,000 memory chips from a company where 3% of the chips are faulty. We have to find the probability that the number of good chips lies between 9675 and 9775. Without loss in generality, we will assume that the chips are independent.

Let  $X_k$  be the zero-one random variable

$$X_k = \begin{cases} 1 & \text{if the } k\text{th chip is good} \\ 0 & \text{if the } k\text{th chip is bad} \end{cases}$$

with the probability of  $X_k = 1$  is 0.97. Thus  $E[X_k] = 1 \times 0.97 + 0 \times 0.03 = 0.97$  and  $E[X_k^2] = 1 \times 0.97 + 0 \times 0.03 = 0.97$ . Hence  $\text{var}[X_k] = 0.97 - (0.97)^2 = 0.0291$ .

If  $Y$  is the sum of the random variables  $\{X_k, k = 1, \dots, 10,000\}$ , then the mean of  $Y$  is

$$E[Y] = \mu_Y = E\left[\sum_{k=1}^{10,000} X_k\right] = \sum_{k=1}^{10,000} E[X_k] = 9700$$

With the assumed independence of  $X_k$ , the variance of  $Y$  is

$$\text{var}[Y] = \sigma_Y^2 = \text{var}\left[\sum_{k=1}^{10,000} X_k\right] = \sum_{k=1}^{10,000} \text{var}[X_k] = 291$$

With  $a = (9675 + 9775)/2 = 9725$ ,  $\varepsilon = 50$ , and  $E[Y - a]^2$  given by

$$\begin{aligned} E[Y - a]^2 &= E[Y^2 - 2Ya + a^2] = EY^2 - 2\mu_Y a + a^2 \\ &= \sigma_Y^2 + \mu_Y^2 - 2\mu_Y a + a^2 = 291 + 9700^2 - 2 \cdot 9700 \cdot 9725 + 9725^2 \\ &= 916 \end{aligned}$$

we can use Eq. (14.3.5) and write

$$P(|Y - 9725| \geq 50) \leq \frac{916}{50^2} = 0.3664$$

What we want is the complementary event  $P(|Y - 9725| \leq 50)$ , and this is given by  $1 - 0.3664 = 0.6336$ . Thus, the probability that the number of good chips will lie between 9675 and 9775 is 63.36%. On the other hand, if we want the number of good chips to lie in the region  $9700 \pm 50$ , namely, 9650 and 9750, then the probability using Chebyshev inequality is

$$P(|Y - 9700| \leq 50) = 1 - P(|Y - 9700| \geq 50) = 1 - \frac{291}{2500} = 0.8836 = 88.36\%$$

**Example 14.3.2** In a binomially distributed random variable  $b(k; n, p) = \binom{n}{k} p^k q^{n-k}$ , we have  $n = 50$  and  $p = 0.1$ . Using Markov inequality, we want to find the probability of  $X > 10$  and compare it with the true value. The mean value is  $np = 5$  and  $\varepsilon = 10$ . Thus, from Eq. (14.3.1), we have the Markov bound

$$P(X \geq 10) \leq \frac{5}{10} = 0.5$$

and the true value is

$$P(X \geq 10) = \sum_{k=10}^{50} \frac{50!}{k!(50-k)!} 0.1^k 0.9^{50-k} = 0.02454$$

This Markov bound is very loose!

We will now find the Chebyshev bound from Eq. (14.2.7) with  $\phi(x) = x^2$  and compare the results. Equation (14.2.7) becomes  $P\{|X| > \varepsilon\} \leq E[X^2]/\varepsilon^2$  and  $E[X^2] = np + (np)^2 = 29.5$ . The Chebyshev bound is  $\{X > 10\} \leq (29.5/100) = 0.295$ , which is better than the Markov bound of 0.5, but nowhere close to the true value of 0.02454.

## 14.4 CHERNOFF BOUND

In determining the Chebyshev bound, the indicator function was bounded by a quadratic. In an effort to improve the tightness of the bound, we will use a positive exponential bound the indicator function as shown in Fig. 14.4.1.

The expectation of the indicator function yields

$$E[I_A] = P(|Y| \geq \varepsilon) = P(|X - \mu_X| \geq \varepsilon) \quad (14.4.1)$$

The indicator function is bounded by  $e^{\lambda(Y-\varepsilon)}$ ,  $\lambda > 0$ ,  $\varepsilon > 0$ , as shown in Fig. 14.4.1. From the figure,  $I_A \leq e^{\lambda(Y-\varepsilon)}$  and taking expectations on both sides, we obtain

$$E[I_A] = P(Y \geq \varepsilon) \leq E[e^{\lambda(Y-\varepsilon)}]$$

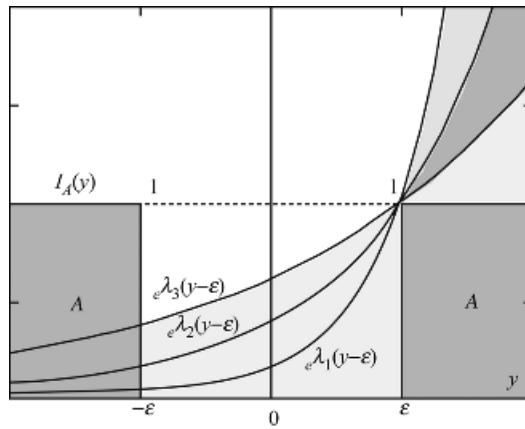


FIGURE 14.4.1

or

$$P(Y \geq \varepsilon) \leq e^{-\lambda\varepsilon} E[e^{\lambda Y}] = e^{-\lambda\varepsilon} M_Y(\lambda) \quad (14.4.2)$$

where  $|Y|$  has been replaced with  $Y$  since we are considering only the positive bound and  $M_Y(\lambda)$  is the moment generating function of  $Y$ .

The bound given above will hold for any  $\lambda > 0$ . However, we would like to choose  $\lambda$  so that  $e^{-\lambda\varepsilon} E[e^{\lambda Y}]$  is a minimum. Differentiating  $e^{-\lambda\varepsilon} E[e^{\lambda Y}]$  with respect to  $\lambda$  and setting it equal to zero, we obtain

$$\frac{d}{d\lambda} E[e^{\lambda(Y-\varepsilon)}] \Big|_{\lambda=\lambda_0} = E \left[ \frac{d}{d\lambda} e^{\lambda(Y-\varepsilon)} \right] \Big|_{\lambda=\lambda_0}$$

or

$$E[(Y - \varepsilon)e^{\lambda_0(Y-\varepsilon)}] = e^{-\lambda_0\varepsilon} E[(Y - \varepsilon)e^{\lambda_0 Y}] = 0 \quad (14.4.3)$$

Rearranging terms in Eq. (14.4.3), we obtain an implicit relation for the minimum  $\lambda_0$  in terms of the moment generating functions (MGFs):

$$\varepsilon = \frac{E[Ye^{\lambda_0 Y}]}{E[e^{\lambda_0 Y}]} = \frac{M'_Y(\lambda_0)}{M_Y(\lambda_0)} \quad (14.4.4)$$

The Chernoff bound is given by

$$P(Y \geq \varepsilon) \leq e^{-\lambda_0\varepsilon} M_Y(\lambda_0) \quad (14.4.5)$$

where  $\lambda_0$  is given by Eq. (14.4.4). A similar result holds for the bound on the negative side.

**Example 14.4.1** We will find the bounds on tails probability for the standard Gaussian random variable  $Y$  using Chernoff's formula and compare it to the Chebyshev bound found in Example 14.2.1. The MGF for any Gaussian random variable  $N$  is given by

$$M_N(t) = e^{t\mu + (t^2\sigma^2/2)} \quad \text{and for a standard Gaussian, } Y \sim N(0,1)$$

$$M_Y(t) = e^{t^2/2}$$

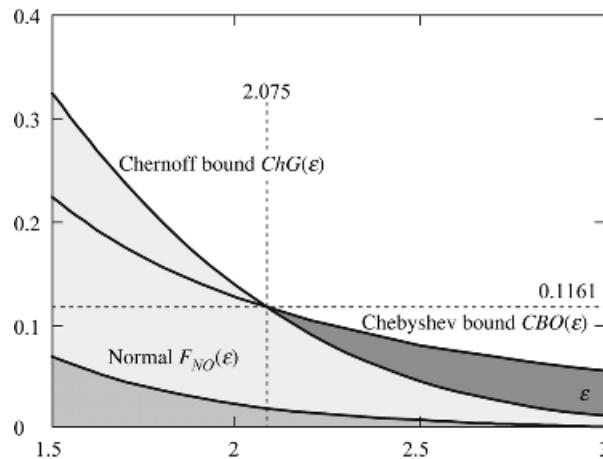


FIGURE 14.4.2

We will now find the optimum  $\lambda$  from Eq. (14.4.4):

$$\varepsilon = \frac{M'_Y(\lambda_0)}{M_Y(\lambda_0)} = \frac{\lambda_0 e^{(\lambda_0^2/2)}}{e^{(\lambda_0^2/2)}} \quad \text{or} \quad \lambda_0 = \varepsilon$$

The Chernoff bound from Eq. (14.4.5) is

$$P(Y \geq \varepsilon) \leq e^{-\lambda_0 \varepsilon} e^{\lambda_0^2/2} = e^{-\varepsilon^2} e^{\varepsilon^2/2} = e^{-\varepsilon^2/2} = ChG(\varepsilon)$$

The Chernoff bound,  $ChG(\varepsilon)$ , the one-sided Chebyshev bound,  $CBO(\varepsilon)$ , and the one-sided normal distribution,  $F_{NO}(\varepsilon) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\varepsilon} e^{-x^2/2} dx$  are shown in Fig. 14.4.2 and in Table 14.4.1 for  $\varepsilon = 1.5, \dots, 3$ .

TABLE 14.4.1

$\varepsilon$	Normal $F_{NO}(\varepsilon)$	Chernoff $ChG(\varepsilon)$	Chebyshev $CBO(\varepsilon) = 1/2\varepsilon^2$
1.5	0.066807	0.324652	0.222222
1.6	0.0548	0.278037	0.195312
1.7	0.044566	0.235746	0.17301
1.8	0.035931	0.197899	0.154321
1.9	0.028717	0.164474	0.138504
2	0.02275	0.135335	0.125
2.1	0.017864	0.110251	0.113379
2.2	0.013903	0.088922	0.103306
2.3	0.010724	0.071005	0.094518
2.4	0.008197	0.056135	0.086806
2.5	0.00621	0.043937	0.08
2.6	0.004661	0.034047	0.073964
2.7	0.003467	0.026121	0.068587
2.8	0.002555	0.019841	0.063776
2.9	0.001866	0.014921	0.059453
3	0.00135	0.011109	0.055556

From the figure and the table we see that for  $\epsilon > 2$  the Chernoff bound is better than the Chebyshev bound. The crossover point occurs when  $\epsilon = 2.075$ .

## 14.5 CAUCHY–SCHWARTZ INEQUALITY

An important inequality used in communication problems is the Cauchy–Schwartz inequality. Let  $g(x)$  and  $h(x)$  be two complex functions with finite norm defined by

$$\begin{aligned}\langle g(x), g(x) \rangle &= \|g(x)\| = \sqrt{\int_{-\infty}^{\infty} g(x)g^*(x)dx} < \infty \\ \langle h(x), h(x) \rangle &= \|h(x)\| = \sqrt{\int_{-\infty}^{\infty} h(x)h^*(x)dx} < \infty\end{aligned}\tag{14.5.1}$$

The Cauchy–Schwartz inequality is given by

$$\left| \int_{-\infty}^{\infty} g(x)h^*(x)dx \right| \leq \sqrt{\int_{-\infty}^{\infty} |g(x)|^2 dx} \cdot \sqrt{\int_{-\infty}^{\infty} |h(x)|^2 dx}\tag{14.5.2}$$

We can also state this inequality in terms of the inner product defined by

$$\langle g(x), h(x) \rangle = \int_{-\infty}^{\infty} g(x)h^*(x)dx = \langle h(x), g(x) \rangle^*\tag{14.5.3}$$

and the inequality becomes

$$|\langle g(x), h(x) \rangle| \leq \|g(x)\| \cdot \|h(x)\|\tag{14.5.4}$$

If we define two real-valued functions,  $a(x)$  and  $b(x)$ , as

$$a(x) = |g(x)|, \quad b(x) = |h(x)|$$

then we have to show that

$$\int_{-\infty}^{\infty} a(x)b(x)dx \leq \sqrt{\int_{-\infty}^{\infty} a^2(x)dx} \sqrt{\int_{-\infty}^{\infty} b^2(x)dx}\tag{14.5.5}$$

Let us represent  $a(x)$  and  $b(x)$  in a function space spanned by the basis vectors  $\phi_1(x)$  and  $\phi_2(x)$ . These vectors have the following orthonormal property:

$$\int_{-\infty}^{\infty} \phi_1^2(x)dx = \int_{-\infty}^{\infty} \phi_2^2(x)dx = 1; \quad \int_{-\infty}^{\infty} \phi_1(x)\phi_2(x)dx = 0\tag{14.5.6}$$

Thus

$$a(x) = a_1\phi_1(x) + a_2\phi_2(x); \quad b(x) = b_1\phi_1(x) + b_2\phi_2(x)\tag{14.5.7}$$

where  $a_1, a_2, b_1, b_2$  are coordinates in the function space of  $\{\phi_1(x), \phi_2(x)\}$  and the vectors,  $a = [a_1, a_2]$ ;  $b = [b_1, b_2]$  as shown in Fig. 14.5.1.

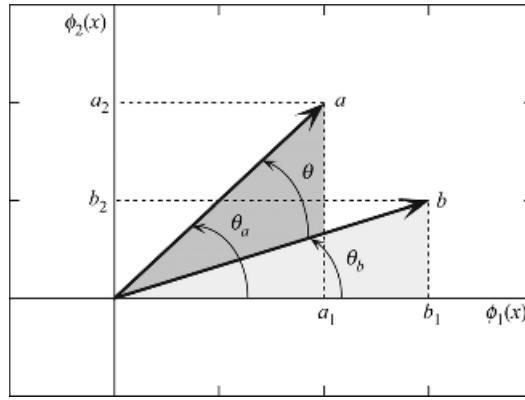


FIGURE 14.5.1

We will find the inner product,  $\langle a(x), b(x) \rangle = \int_{-\infty}^{\infty} a(x)b(x)dx$ . From Eq. (14.5.7) we can write

$$\begin{aligned}
 \int_{-\infty}^{\infty} a(x)b(x)dx &= \int_{-\infty}^{\infty} [a_1\phi_1(x) + a_2\phi_2(x)][b_1\phi_1(x) + b_2\phi_2(x)]dx \\
 &\quad [\text{using orthonormal properties of } \phi_1(x) \text{ and } \phi_2(x)] \\
 &= \int_{-\infty}^{\infty} [a_1b_1\phi_1^2(x) + a_2b_2\phi_2^2(x)] dx = a_1b_1 + a_2b_2
 \end{aligned} \tag{14.5.8}$$

From Fig. 14.5.1 we can write

$$\begin{aligned}
 \cos(\theta) &= \cos(\theta_a - \theta_b) = \cos(\theta_a)\cos(\theta_b) + \sin(\theta_a)\sin(\theta_b) \\
 &= \frac{a_1}{|a|} \cdot \frac{b_1}{|b|} + \frac{a_2}{|a|} \cdot \frac{b_2}{|b|} = \frac{a \cdot b}{|a||b|}
 \end{aligned} \tag{14.5.9}$$

Since  $|\cos(\theta)| \leq 1$ , we have

$$\frac{|a \cdot b|}{|a||b|} \leq 1 \quad \text{and} \quad |a \cdot b| \leq |a||b| \tag{14.5.10}$$

Substituting for  $|a| = \sqrt{\int_{-\infty}^{\infty} a^2(x)dx}$ ;  $|b| = \sqrt{\int_{-\infty}^{\infty} b^2(x)dx}$  in Eq. (14.5.10), we obtain Eq. (14.5.5) and hence Eq. (14.5.2). The equality is obtained when we substitute  $a = Kb$  in Eq. (14.5.10) or  $a(x) = Kb(x)$  and hence  $g(x) = Kh(x)$ .

If  $g(X)$  and  $h(X)$  are real random variables, then we define the norms and the inner product given by Eqs. (14.5.1) and (14.5.3) as

$$\begin{aligned}
 \langle g(X), g(X) \rangle &= E[g^2(X)] = \int_{-\infty}^{\infty} g^2(x)f(x)dx < \infty \\
 \langle h(X), h(X) \rangle &= E[h^2(X)] = \int_{-\infty}^{\infty} h^2(x)f(x)dx < \infty
 \end{aligned} \tag{14.5.11}$$

$$\langle g(X), h(X) \rangle = E[g(X)h(X)] = \int_{-\infty}^{\infty} g(x)h(x)f(x)dx = \langle h(X), g(X) \rangle \tag{14.5.12}$$

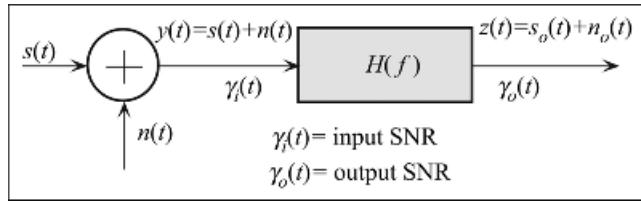


FIGURE 14.5.2

and Cauchy–Schwartz inequality of Eq. (14.5.2) becomes

$$\left| \int_{-\infty}^{\infty} g(x)h(x)f(x)dx \right| \leq \sqrt{\int_{-\infty}^{\infty} g^2(x)f(x)dx} \cdot \sqrt{\int_{-\infty}^{\infty} h^2(x)f(x)dx} \quad (14.5.13)$$

or

$$|E[g(X)h(X)]| \leq \sqrt{E[g^2(X)]E[h^2(X)]}$$

**Example 14.5.1** A communication system is shown in Fig. 14.5.2, where  $s(t)$  is the input signal and  $n(t)$  is a random variable representing input noise,  $s_o(t)$  and  $n_o(t)$  are the output signal and output noise,  $H(f)$  is the transfer function.  $\gamma_i(t)$  the input signal-to-noise ratio (SNR), and  $\gamma_o(t)$  the output signal-to-noise ratio are given by

$$\gamma_i(t) = \frac{s^2(t)}{E[n^2(t)]}; \quad \gamma_o(t) = \frac{s_o^2(t)}{E[n_o^2(t)]} \quad (14.5.14)$$

The output  $s_o(t)$  is given by the inverse Fourier transform of  $S(f)H(f)$ , or

$$s_o(t) = \int_{-\infty}^{\infty} s(\tau)h(t - \tau)d\tau = \int_{-\infty}^{\infty} S(f)H(f)e^{j2\pi f}df \quad (14.5.15)$$

and the output noise is characterized by

$$E[n_o^2(t)] = \int_{-\infty}^{\infty} |H(f)|^2N(f)df \quad (14.5.16)$$

where  $N(f)$  is the frequency-domain characterization of the input noise  $n(t)$ . With these characterizations of Eqs. (14.5.15) and (14.5.16), the output SNR  $\gamma_o(t)$  can be given by

$$\gamma_o(t) = \frac{s_o^2(t)}{E[n_o^2(t)]} = \frac{\left| \int_{-\infty}^{\infty} S(f)H(f)e^{j2\pi f}df \right|^2}{\int_{-\infty}^{\infty} |H(f)|^2N(f)df} \quad (14.5.17)$$

The Cauchy–Schwartz inequality of Eq. (14.5.2) is modified to suit the frequency-domain characterization as follows

$$\left| \int_{-\infty}^{\infty} A(f)B^*(f)df \right|^2 \leq \int_{-\infty}^{\infty} |A(f)|^2df \cdot \int_{-\infty}^{\infty} |B(f)|^2df \quad (14.5.18)$$

with the equality holding for  $A(f) = KB^*(f)$ . We substitute

$$A(f) = H(f)\sqrt{N(f)}$$

and

$$B(f) = \frac{S(f)e^{j2\pi ft}}{\sqrt{N(f)}} \quad (14.5.19)$$

Thus

$$A(f)B(f) = S(f)H(f)e^{j2\pi ft} \quad (14.5.20)$$

Applying Eqs. (14.5.19) and (14.5.20) in Eq. (14.5.18), the Cauchy–Schwartz inequality becomes

$$\left| \int_{-\infty}^{\infty} S(f)H(f)e^{j2\pi ft} df \right|^2 \leq \int_{-\infty}^{\infty} |H(f)|^2 N(f) df \cdot \int_{-\infty}^{\infty} \frac{|S(f)|^2}{N(f)} df \quad (14.5.21)$$

Substituting Eq. (14.5.21) in Eq. (14.5.17), the output SNR is given by

$$\frac{s_o^2(t)}{E[n_o^2(t)]} = \gamma_o(t) \leq \frac{\int_{-\infty}^{\infty} |H(f)|^2 N(f) df \cdot \int_{-\infty}^{\infty} \frac{|S(f)|^2}{N(f)} df}{\int_{-\infty}^{\infty} |H(f)|^2 N(f) df} \quad (14.5.22)$$

or

$$\gamma_o(t) \leq \int_{-\infty}^{\infty} \frac{|S(f)|^2}{N(f)} df \quad (14.5.22)$$

The equality will hold for  $A(f) = KB^*(f)$ , or

$$H(f)\sqrt{N(f)} = K \frac{S^*(f)e^{-j2\pi fT}}{\sqrt{N(f)}} \quad \text{or} \quad H_{\text{opt}}(f) = K \frac{S^*(f)e^{-j2\pi fT}}{|N(f)|} \quad (14.5.23)$$

where we have substituted  $T$  the time of observation for  $t$ . The maximum output SNR is

$$\gamma_{o, \text{max}} \leq \int_{-\infty}^{\infty} \frac{|S(f)|^2}{|N(f)|} df \quad (14.5.24)$$

## 14.6 JENSEN'S INEQUALITY

### Convex Functions

Before we enunciate Jensen's inequality, we want to make clear the idea of a convex function. We are interested in a convex function because convexity guarantees a unique minimum over the interval of convexity. A function  $h(x)$  is convex in the interval  $(a, b)$  if for every  $x_1$  belonging to the interval  $(a, b)$ , there exists a constant  $m$  such that the following inequality is satisfied for all  $x$ :

$$h(x) \geq h(x_1) + m(x - x_1) \quad (14.6.1)$$

As seen in Fig. 14.6.1,  $h(x_1) + m(x - x_1)$  is the equation to a straight line passing tangentially to the point  $h(x_1)$  and  $m$  is the slope of the straight line at the point  $x_1$ . It is clear from the figure that the inequality is satisfied geometrically. For the sake of completeness, we will give a more formal definition of convexity.

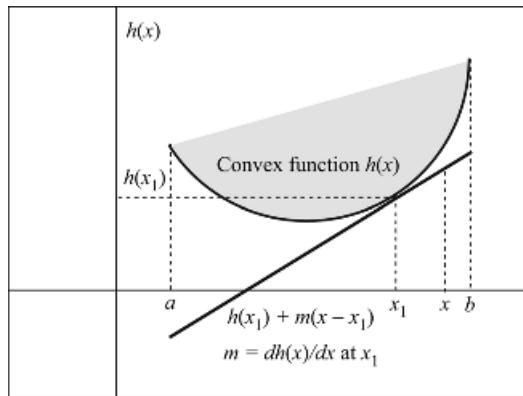


FIGURE 14.6.1

A function  $h(x)$  is convex over an interval  $(a, b)$ , if every  $x_1$  and  $x_2$  belonging to  $(a, b)$  and for  $0 \leq \lambda \leq 1$ , the following equation holds.

$$h[\lambda x_1 + (1 - \lambda)x_2] \leq \lambda h(x_1) + (1 - \lambda)h(x_2) \quad (14.6.2)$$

If this equation is a strict inequality, then  $h(x)$  is strictly convex. Geometrically, this equation conveys that if the function  $h(x)$  lies below the line segment connecting two points in the interval  $(a, b)$ , then it is convex.

It can also be shown that it is necessary and sufficient that the second derivative  $d^2h(x)/dx^2 \geq 0$  for all  $x$  in the interval  $(a, b)$ , for Eq. (14.6.2) to be satisfied, in which case  $h(x)$  is convex.

**Example 14.6.1** The following functions  $h(x)$  are convex because the second derivative is nonnegative:

1.  $x^2$ , since  $h''(x) = 2$ , positive for all  $x$
2.  $|x|$ , since  $h''(x) = 0$ , nonnegative for all  $x$
3.  $e^x$ , since  $h''(x) = e^x$ , positive for all  $x$
4.  $-\ln(x)$ , for  $0 < x < \infty$ , since  $h''(x) = 1/x^2$ , positive in the range of  $x$
5.  $x \ln(x)$ , for  $0 < x < \infty$ , since  $h''(x) = 1/x$ , positive in the range of  $x$

### Jensen's Inequality

Jensen's inequality states that if  $h(x)$  is convex and if  $X$  is a random variable with  $E|X| < \infty$ , then

$$E[h(X)] \geq h(E[X]) \quad (14.6.3)$$

Substituting  $E[X] = x_1$  in the convexity [Eq. (14.6.1)], we obtain

$$h(x) \geq h(E[X]) + m(x - E[X]) \quad (14.6.4)$$

Taking expectations on both sides of Eq. (14.6.4) gives

$$\begin{aligned} E[h(X)] &\geq h(E[X]) + m(E[X] - E[X]) \\ &= 0 \end{aligned}$$

and hence Jensen's inequality [Eq. (14.6.3)].

**Example 14.6.2** Since  $E[X^{2n}] = E[(X^n)^2]$  is a convex function from Example 14.6.1, it follows from Jensen's inequality that

$$E[X^{2n}] \geq \{E[X^n]\}^2 \quad (14.6.5)$$

and as a consequence

$$E[X^2] \geq \{E[X]\}^2 \quad (14.6.6)$$

which is a restatement of the fact that the variance of a random variable is nonnegative.

## 14.7 CONVERGENCE CONCEPTS

We saw in Section 14.1 that if  $E(X - a)^2 = 0$ , then  $X = a$  with probability 1 or  $X$  converges to  $a$  with probability 1. There are a few other convergence concepts, and we shall make these a little clearer.

A machinist wants to make piston rings of diameter  $a$  inches to a tolerance limit of  $\pm \varepsilon$ . Let us assume that the variability of the measurement is  $\sigma_X^2$ . According to the Chebyshev inequality of Eq. (14.2.8), we have

$$P\{|X - a| < \varepsilon\} \geq 1 - \frac{\sigma_X^2}{\varepsilon^2} \quad (14.7.1)$$

If  $\sigma_X$  is very much smaller than  $\varepsilon$ , then the observed random variable  $X$  is between  $a - \varepsilon$  and  $a + \varepsilon$  is almost certain, and one measurement of  $X$  is sufficient. However, if  $\sigma_X$  is not sufficiently small compared to  $\varepsilon$ , then the estimate  $a$  will not give results of sufficient accuracy. To improve the accuracy of the estimate  $a$ , we take  $n$  measurements corresponding to  $n$  random variables  $\{X_i, i = 1, \dots, n\}$  with mean  $a$  and noise random variables  $W_i$  given by

$$X_i = a + W_i, \quad i = 1, \dots, n$$

where we will assume that the random variables  $W_i$  are independent with zero mean and variance  $\sigma^2$ . We will represent the average of these  $n$  random variables by

$$\hat{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (14.7.2)$$

where the mean of  $\hat{X}$  is  $a$  and the variance is  $n\sigma_X^2$  and the corresponding Chebyshev inequality is given by

$$P\{|X - a| < \varepsilon\} \geq 1 - \frac{\sigma_X^2}{n\varepsilon^2} \quad (14.7.3)$$

We can see from Eq. (14.7.3) that as the number of samples  $n$  becomes very large, the probability that the  $|X - a|$  is less than  $\varepsilon$  becomes almost certain even though  $\sigma_X^2$  may not be small compared to  $\varepsilon$ .

### Pointwise Convergence

A discrete sequence of random variables

$$\{X_1, X_2, \dots, X_n, \dots\} \quad (14.7.4)$$

converges to a limiting random variable  $X$  if and only if for any  $\varepsilon > 0$ , however small, we can find a number  $n_0$  such that

$$|X_n(\xi) - X(\xi)| < \varepsilon \quad (14.7.5)$$

for every  $n > n_0$  and every  $\xi$ . This type of convergence is also called *pointwise* or *everywhere convergent*. This may be all right for a real variable, but it is highly restrictive for a random variable. It is possible that the sequence  $\{X_n\}$  may converge for  $\xi$  points belonging to a subset of the sample space  $S$ . Thus, convergences using probability measures are more useful than this type of convergence.

### Almost Sure Convergence (a.s.)

A sequence of random variables  $\{X_n\}$  converges *almost surely* (a.s.), or *almost everywhere* (a.e.), or *strongly* to the random variable  $X$ , if for every  $\xi$  point in the sample space  $S$  satisfies the following criterion:

$$\lim_{n \rightarrow \infty} |X_n(\xi) - X(\xi)| \rightarrow 0 \text{ with probability 1} \quad (14.7.6)$$

The criterion can also be written as

$$P(X_n \rightarrow X) = 1 \quad \text{as } n \rightarrow \infty \quad (14.7.7)$$

In Eq. (14.7.7) the quantity  $X_n \rightarrow X$  is to be interpreted as the set of all outcomes  $\xi$  such that the number  $X_n(\xi)$  tends to the number  $X(\xi)$ . In some cases the limit  $X$  may not be known a priori. In this case the Cauchy criterion of convergence comes handy to define *mutual convergence*.

### Cauchy Convergence Criterion

A sequence of random variables  $\{X_n\}$  on the probability space  $\{S, \mathcal{F}, P\}$  is convergent almost surely if the set of points  $\xi$  for which the real sequence  $\{X_n(\xi)\}$  is convergent with probability 1. Or, in other words

$$P\{\xi \in S : X_{n+m}(\xi) - X_n(\xi) \rightarrow 0\} \quad \text{as } n \rightarrow \infty \quad (14.7.8)$$

for every  $m > 0$ . Using this criterion, the sequence  $\{X_n\}$  converges *mutually almost surely* if

$$P\{X_{n+m} - X_n \rightarrow 0\} \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad (14.7.9)$$

### Convergence in Probability (p)

A sequence of random variables  $\{X_n\}$  converges in *probability* or *weakly converges* to the random variable  $X$  if for every  $\epsilon > 0$ , however small

$$\lim_{n \rightarrow \infty} P\{|X_n - X| \geq \epsilon\} = 0 \quad (14.7.10)$$

or equivalently

$$\lim_{n \rightarrow \infty} P\{|X_n - X| \leq \epsilon\} = 1 \quad (14.7.11)$$

Using the Cauchy criterion, the sequence  $\{X_n\}$  converges *mutually in probability* to  $X$  if

$$\lim_{n \rightarrow \infty} P\{|X_{n+m} - X_n| \geq \epsilon\} \rightarrow 0 \quad (14.7.12)$$

Convergence in probability is widely used in probability theory. In particular, it is used to determine the weak law of large numbers and the consistency of estimators.

**Convergence in Mean Square (m.s.)**

A sequence of random variables  $\{X_n\}$  converges in the *mean square* or *limit in the mean* to the random variable  $X$  if

$$\lim_{n \rightarrow \infty} E\{|X_{n+m} - X|^2\} \rightarrow 0 \quad (14.7.13)$$

This is also called *quadratic mean convergence*.

Mutual convergence in the mean square is defined by

$$\lim_{n \rightarrow \infty} E\{|X_{n+m} - X_n|^2\} \rightarrow 0 \quad (14.7.14)$$

for all  $m > 0$ .

**Convergence in Distribution**

The widest form of convergence is the *convergence in distribution*. If we denote the distribution functions of  $X_n$  and  $X$  by  $F_n(x)$  and  $F(x)$ , respectively, then the sequence  $\{X_n\}$  converges in *distribution* to the random variable  $X$  if

$$\lim_{n \rightarrow \infty} F_n(x) \rightarrow F(x) \quad (14.7.15)$$

for every continuity point  $x$  of  $F(x)$ . However, the sequence  $\{X_n(\xi)\}$  need not converge for any  $\xi$ .

**Properties of Convergences**

1. *Almost-sure convergence* implies convergence in probability. The converse is not true.
2. *Mean-square convergence* also implies convergence in probability. Again the converse is not true.
3. If  $\{X_n\}$  converges in probability, then there exists a subsequence  $\{X_{nk}\}$  of  $\{X_n\}$  that converges almost surely.
4. Convergences almost surely, probability, and mean square all imply convergence in distribution. Of course, the converse is not true.

All the four convergences are shown schematically in Fig. 14.7.1.

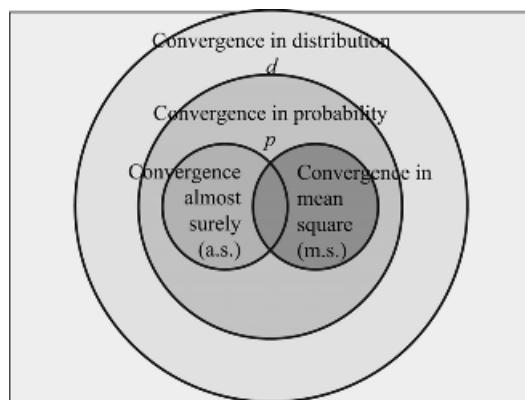


FIGURE 14.7.1

## 14.8 LIMIT THEOREMS

### Weak Law of Large Numbers (WLLN)

Armed with the convergence concepts, we will now explore some limit theorems. Let  $\{X_n\}$  be a sequence of independent identically distributed random variables with finite means  $\mu$  and finite variances  $\sigma^2$ . Let  $S_n$  be the sum of the  $n$  random variables given by

$$S_n = X_1 + X_2 + \cdots + X_n \quad (14.8.1)$$

Let the random variable  $Z_n$ , called the *sample mean*, be defined by

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \quad (14.8.2)$$

Clearly the expected value of  $Z_n$  is

$$E[Z_n] = E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] = \frac{n\mu}{n} = \mu \quad (14.8.3)$$

and because of the independence of the sequence  $\{X_n\}$ , the variance of  $Z_n$  is

$$\text{var}[Z_n] = \text{var}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \quad (14.8.4)$$

The weak law of large numbers states that for every  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P[|Z_n - \mu| > \varepsilon] \rightarrow 0 \quad (14.8.5)$$

To show this result, we apply Chebyshev's inequality [Eq. (14.2.1)]

$$P[|Z_n - \mu| > \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2} \quad (14.8.6)$$

and since  $\varepsilon > 0$ , the limit of Eq. (14.8.6) as  $n$  tends to  $\infty$  is indeed 0. In other words, the probability of the sample mean  $Z_n$  for a specific  $n$  of a sequence  $\{X_n\}$  of random variables deviating from the true mean  $\mu$  by an arbitrary  $\varepsilon > 0$  becomes vanishingly small as  $n$  tends to  $\infty$ . Or the sample mean  $Z_n$  converges in probability to the true mean.

**Example 14.8.1** In the design of a gyroscope for a navigation system we want the probability of the sample drift rate  $Z_n$  calculated for  $n$  production gyros differing from the unknown true drift rate  $\mu$  by more than 0.1 be less than 2%. We want to find the number  $n$  of the gyros needed to establish this criterion expressed in terms of the unknown variance  $\sigma^2$ . Applying Chebyshev inequality given by Eq. (14.2.8), we can write

$$P\left[|Z_n - \mu| < \frac{1}{10}\right] \geq 1 - \frac{\sigma^2 \cdot 100}{n} > 0.98$$

From  $1 - (100\sigma^2/n) > 0.98$  we have

$$\begin{aligned} 1 - 0.98 &\geq \frac{100\sigma^2}{n} \\ n &\geq \frac{100\sigma^2}{0.02} = 5000\sigma^2 \end{aligned}$$

If the quality control is very good,  $\sigma$  will be small, in which case the number  $n$  will be small. However, since we are using Chebyshev inequality that gives very loose bounds,

we expect  $n$  to be overly conservative. We will revisit this example after discussing the central limit theorem in the next section.

**Example 14.8.2** We will assume that there are a sequence of independent events  $\{A_i\}$  that will be denoted by the sequence of indicator functions  $I_{A_i}(\xi)$ , given by

$$I_{A_i}(\xi) = \begin{cases} 1 & \text{if } \xi \in A_i \\ 0 & \text{if } \xi \in \bar{A}_i \end{cases}$$

$$Z_n = \frac{1}{n} \sum_{i=1}^n I_{A_i}$$

We will assume that  $P(A_i) = p$  and  $P(\bar{A}_i) = 1 - p$  for  $i = 1, \dots, n$ . Thus

$$\begin{aligned} E[I_{A_i}] &= p \quad \text{and} \quad E[I_{A_i}^2] = p \quad \text{and} \quad \text{var}[I_{A_i}] = p(1 - p), \quad i = 1, \dots, n \\ E[Z_n] &= p \quad \text{and} \quad \text{var}[Z_n] = \frac{p(1 - p)}{n} \end{aligned}$$

We can now apply WLLN and write

$$P\{|Z_n - p| < \varepsilon\} \geq 1 - \frac{p(1 - p)}{n\varepsilon^2}$$

Assuming  $p = 0.4$ , we will estimate  $p$  within 0.1. If  $n = 500$ , then this equation becomes

$$P\{|Z_n - p| < 0.1\} \geq 1 - \frac{0.4 \times 0.6}{500 \times 0.01} = 1 - \frac{8}{125} = \frac{117}{125}$$

What this equation tells us is that if we perform 500 trials, then 468 of these will result in an error of  $\{Z_n - p\}$  of less than 0.1.

### **Strong Law of Large Numbers**

The weak law of large numbers given by Eq. (14.8.5) states that the sample mean  $Z_n$  converges in probability to the true mean  $\mu$  as  $n$  tends to  $\infty$ . The strong law of large numbers states that this convergence is not only in probability but also with probability 1, which is almost surely (a.s.). Or

$$P\left\{\lim_{n \rightarrow \infty} (Z_n - \mu) = 0\right\} = 1 \quad (\text{a.s.}) \quad (14.8.7)$$

Unlike the weak law of large numbers, which is easy to show with Chebyshev's inequality, the strong law of large numbers is difficult to show and will not be attempted in this book. Even though Eqs. (14.8.5) and (14.8.7) look similar, they are dramatically different. Equation (14.8.7) states that the sequence of the sample means  $Z_n$  for every  $n$  will tend to the true mean  $\mu$  and will stay close to it. The strong law gives us confidence that sample means calculated from a number of realizations will always tend to the true mean as the number increases and isolated discrepancies become unimportant.

### **Central Limit Theorem (CLT)**

Perhaps the most important theorem in statistics is the central limit theorem, which in its most general form states that if we sum a sufficient number of random variables, the resulting probability is a Gaussian. Let  $\{X_n\}$  be a sequence of  $n$  independent

identically distributed random variables with mean  $\mu_X$  and variance  $\sigma_X^2$ . Their sum  $S_n$  is given by

$$S_n = X_1 + X_2 + \cdots + X_n \quad (14.8.8)$$

The mean  $\mu$  and the variance  $\sigma^2$  of  $S_n$  are

$$\begin{aligned} E[S_n] &= \mu = n\mu_X \\ \text{var}[S_n] &= \sigma^2 = n\sigma_X^2 \end{aligned} \quad (14.8.9)$$

We will now define a normalized random variable  $Z_n$  with mean 0 and variance 1 as follows:

$$Z_n = \frac{S_n - \mu}{\sigma} = \frac{S_n - n\mu_X}{\sigma_X\sqrt{n}} \quad (14.8.10)$$

The central limit theorem states that as  $n$  tends to  $\infty$ ,  $Z_n$  tends to a standard normal distribution:

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} dx \quad (14.8.11)$$

We can show this result heuristically by using moment generating functions (MGFs) in

$$Z_n = \frac{S_n - n\mu_X}{\sigma_X\sqrt{n}} = \frac{1}{\sigma_X\sqrt{n}} \sum_{i=1}^n (X_i - \mu_X) \quad (14.8.12)$$

Taking MGF on both sides, we obtain

$$\begin{aligned} M_{Z_n}(t) &= E(e^{tZ_n}) = E\left[e^{(t/\sigma_X\sqrt{n}) \sum_{i=1}^n (X_i - \mu_X)}\right] \\ &= E\left[\prod_{i=1}^n e^{(t/\sigma_X\sqrt{n})(X_i - \mu_X)}\right] \\ &= \prod_{i=1}^n E\left[e^{(t/\sigma_X\sqrt{n})(X_i - \mu_X)}\right] = \left\{E\left[e^{(t/\sigma_X\sqrt{n})(X - \mu_X)}\right]\right\}^n \end{aligned} \quad (14.8.13)$$

In Eq. (14.8.13) we have used the independence property of  $\{X_i\}$  and  $X$  has the same pdf as  $X_n$ . We can now expand the last exponential in the above equation and write

$$\begin{aligned} E\left[e^{(t/\sigma_X\sqrt{n})(X - \mu_X)}\right] &= E\left[1 + \frac{1}{\sigma_X\sqrt{n}}(X - \mu_X) + \frac{t^2}{2n\sigma_X^2}(X - \mu_X)^2 + h_n(X; t)\right] \\ &= 1 + 0 + \frac{t^2}{2n} + E[h_n(X; t)] \end{aligned}$$

where  $h_n(X; t)$  represents the higher-order terms. Hence

$$\begin{aligned} M_{Z_n}(t) &= \left\{1 + 0 + \frac{t^2}{2n} + E[h_n(X; t)]\right\}^n \\ \lim_{n \rightarrow \infty} M_{Z_n}(t) &= \left\{1 + \frac{t^2}{2n}\right\}^n = e^{t^2/2} \end{aligned} \quad (14.8.14)$$

Since  $E[h_n(X; t)]$  is of the order of  $1/n^2$ , it is zero in the limit as  $n$  tends to  $\infty$ . The term  $e^{t^2/2}$  is the MGF of a standard Gaussian, and we have shown the central limit theorem. The number  $n$  is usually assumed to be 30 for CLT to hold, but as we shall see in the following example, it holds for  $n$  much less than 30.

**Example 14.8.3** Random variables  $X_i, i = 1, 2, 3, 4$  are independent and identically distributed with a uniform distribution in the interval  $(0, 1]$  with  $f_{X_i}(x) = u(x) - u(x - 1)$ . The mean values are  $\mu_{X_i} = \frac{1}{2}$  and the variances are  $\sigma_{X_i}^2 = \frac{1}{12}$ . We form the random variables  $Y_1 = X_1, Y_2 = X_1 + X_2, Y_3 = X_1 + X_2 + X_3, Y_4 = X_1 + X_2 + X_3 + X_4$ , and their density functions are  $f_{Y_1}(y) = f_{X_1}(y), f_{Y_2}(y) = f_{X_1}(y)^* f_{X_2}(y), f_{Y_3}(y) = f_{Y_2}(y)^* f_{X_1}(y), f_{Y_4}(y) = f_{Y_3}(y)^* f_{X_1}(y)$ . These densities are given by

$$f_{Y_1}(y) = \begin{cases} 1 & \text{if } 0 < y \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad f_{Y_2}(y) = \begin{cases} y & \text{if } 0 < y \leq 1 \\ 2 - y & \text{if } 1 < y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

$$f_{Y_3}(y) = \begin{cases} \frac{y^2}{2} & \text{if } 0 < y \leq 1 \\ -y^2 + 3y - \frac{3}{2} & \text{if } 1 < y \leq 2 \\ \frac{y^2}{2} - 3y + \frac{9}{2} & \text{if } 2 < y \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

$$f_{Y_4}(y) = \begin{cases} \frac{y^3}{6} & \text{if } 0 < y \leq 1 \\ -\frac{y^3}{2} + 2y^2 - 2y + \frac{2}{3} & \text{if } 1 < y \leq 2 \\ \frac{y^3}{2} - 4y^2 + 10y - \frac{22}{3} & \text{if } 2 < y \leq 3 \\ -\frac{y^3}{6} + 2y^2 - 8y + \frac{32}{3} & \text{if } 3 < y \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

These densities are compared to the corresponding normal densities given by

$$f_{N_i}(y) \sim N(\mu_{N_i}, \sigma_{N_i}^2) = N\left(i \frac{1}{2}, i \frac{1^2}{12}\right), \quad i = 1, 2, 3, 4$$

or

$$f_{N_1}(y) = \frac{1}{\sqrt{\frac{\pi}{6}}} e^{-[x-(1/2)]^2/(1/6)} \quad f_{N_2}(y) = \frac{1}{\sqrt{\frac{\pi}{3}}} e^{-(x-1)^2/(1/3)}$$

$$f_{N_3}(y) = \frac{1}{\sqrt{\frac{\pi}{2}}} e^{-[x-(3/2)]^2/(1/2)} \quad f_{N_4}(y) = \frac{1}{\sqrt{\frac{2\pi}{3}}} e^{-(x-2)^2/(2/3)}$$

These densities are shown in all four panels of Fig. 14.8.1.

We see that the approximation progressively improves as  $n$  increases from 1 to 4, and it is very good when  $n = 4$  (Fig. 14.8.1d).

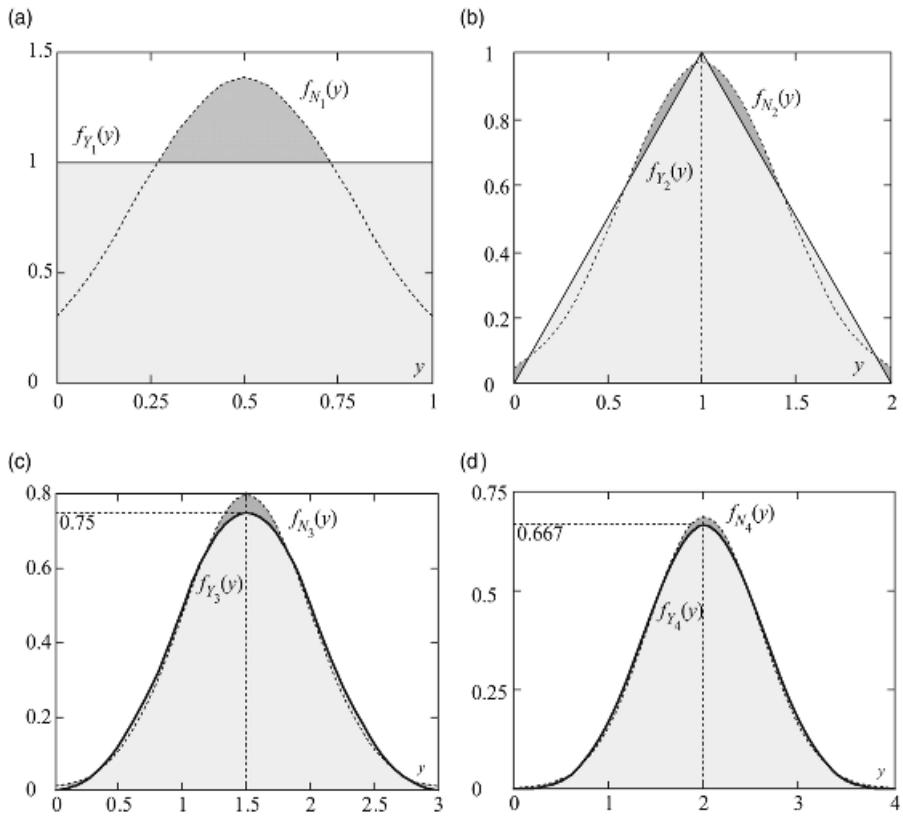


FIGURE 14.8.1

**Example 14.8.4** We will revisit Example 14.8.1 in light of the CLT and determine whether we can get a better bound on the number  $n$  of gyros needed. In order to use the Gaussian tables, we have to convert the random variable  $Z_n$  into the standard Gaussian form by substituting

$$W_n = \frac{Z_n - \mu}{\sqrt{\text{var}(Z_n)}} = \frac{\sqrt{n}(Z_n - \mu)}{\sigma}$$

We can substitute  $W_n$  in  $P[|Z_n - \mu| < \frac{1}{10}] \geq 0.98$  and arrive at

$$P\left[\frac{\sigma|W_n|}{\sqrt{n}} < \frac{1}{10}\right] \geq 0.98 \quad \text{or} \quad P\left[|W_n| < \frac{\sqrt{n}}{10\sigma}\right] \geq 0.98$$

Referring to Gaussian tables, we have the value for a probability of 0.98 is  $x = 2.3263$ . Thus,  $\sqrt{n}/10\sigma \geq 2.3263$  or  $n \geq 2.3263^2 \times 100\sigma^2 = 541.17\sigma^2$  and  $n \geq 542\sigma^2$ . This is a much better number than the  $5000\sigma^2$  that we obtained in Example 14.8.1.

# Computer Methods for Generating Random Variates

If we have an actual probabilistic system, we can obtain the pdf by analyzing the data from the system. On the other hand, if we are designing a system such as transmission of packets through a communication channel, we must get some idea as to how the system will behave before it is actually implemented. We resort to simulation of the system on a computer. The probability distributions may be known *a priori*, and we have to generate random data for the given probability distributions. We call such generated values *random variates*. It is not feasible to develop techniques for generating random variates for every possible distribution function. As it turns out, generating a uniform distribution in the interval  $(0,1]$  is fast on a computer. Starting with the uniform distribution, we can generate random variates for any given distribution function by a number of techniques that will be discussed.

## 15.1 UNIFORM-DISTRIBUTION RANDOM VARIATES

Almost every computer will have a uniform number generator built into the system using some form of the modulus function. One such generator uses the *linear congruential* or *power residue* method that produces a sequence of integers  $\{x_i, i = 0, \dots, m-2\}$ . It is given by the recursive relationship

$$x_{i+1} = (\alpha x_i + c) \bmod m \quad (15.1.1)$$

where  $\alpha$  is a constant multiplier,  $c$  is the increment, and  $m$  is the modulus function. The initial value  $x_0$  is called the “seed.” Equation (15.1.1) can be recursively solved to yield

$$x_n = \left[ \alpha^n x_0 + c \left( \frac{\alpha^n - 1}{\alpha - 1} \right) \right] \bmod m \quad (15.1.2)$$

All these parameters affect the mean, variance, and cycle length of the sequences. It is called a *mixed congruential generator* if  $c \neq 0$  and *multiplicative congruential generator* if  $c = 0$ . The sequence repeats after  $x_{m-2}$  and the period of repetition is less than or equal to  $m - 1$  depending on the parameters.

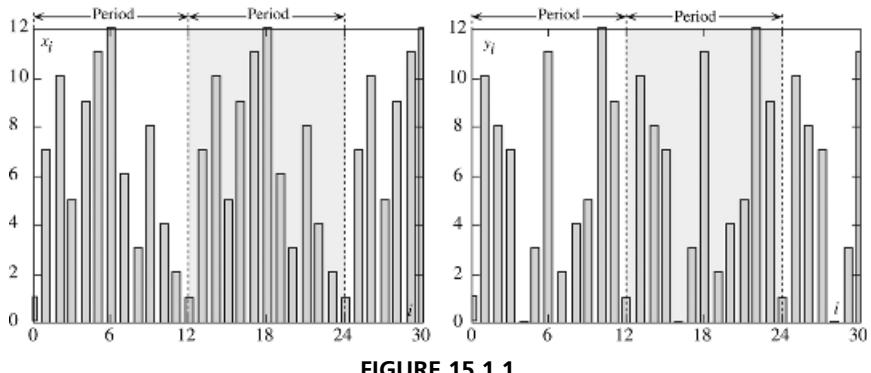


FIGURE 15.1.1

**Example 15.1.1** We will form two sequences with

1.  $\{x_i\} : x_0 = 1, \alpha = 7, c = 0, m = 13$
2.  $\{y_i\} : y_0 = 1, \alpha = 7, c = 3, m = 13$

Using the recursive relationship of Eq. (15.1.1), these two sequences  $\{x_i\}$  and  $\{y_i\}$  are shown in Fig. 15.1.1 and Table 15.1.1.

As we can see from Fig. 15.1.1 and Table 15.1.1, the repetition period is  $13 - 1 = 12$ . Also changing  $c$  from 0 to 3 has radically changed the sequence. We can conclude that almost all random-number generators are not truly random but *pseudorandom* sequences that have the appearance in the statistical sense of being random. Thus, when we refer to *random numbers* we really mean pseudorandom sequences where the period is very high.

A random-number generator must have the following properties:

1. The numbers generated must have good statistical properties. They must be independent and identically distributed. In other words, the values may not be correlated.
2. The repetition period must be long for any simulation to be useful.
3. Random numbers must be repeatable. For each specified  $x_0$ ,  $\alpha$ ,  $c$ , and  $m$ , the generator must produce the same sequence of numbers.
4. Many simulations will require thousands of random numbers, and hence the generation of these numbers must be fast and thus must also have computational efficiency.
5. It must be easy to generate separate sequences of random numbers.

TABLE 15.1.1

$i$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>Series 1: <math>x_0 = 1, \alpha = 7, c = 0, m = 13</math></i>																
$x_i$	1	7	10	5	9	11	12	6	3	8	4	2	1	7	10	5
<i>Series 2: <math>x_0 = 1, \alpha = 7, c = 3, m = 13</math></i>																
$y_i$	1	10	8	7	0	3	11	2	4	5	12	9	1	10	8	7

The linear congruential generator may satisfy all these conditions provided  $\alpha$  and  $m$  are carefully chosen. Good statistical properties have been obtained with  $\alpha = 7^5 = 16,807$  and  $m = 2^{31} - 1 = 2,147,483,647$ . Under these conditions the period of the sequence is  $2^{31} - 2 = 2,147,483,646$ . Sometimes  $\alpha = 742,938,285$  and  $m = 2^{48}$  are also used.

## 15.2 HISTOGRAMS

Having generated random variates, we have to see how well they approximate the desired probability density. From the definition of the density function  $f_X(x)$ , we have

$$f_X(x) \approx \frac{P(X \leq x + \Delta x) - P(X \leq x)}{\Delta x}$$

or

$$f_X(x)\Delta x \approx P(x < X \leq x + \Delta x) \quad (15.2.1)$$

which gives the value of the probability mass at the point  $x$  and width  $\Delta x$ . Hence with the generated random variates  $\{x_i\}$  we will be able to construct a *histogram* that will approximate the given mass function  $f_X(x)\Delta x$ . The purpose of the histogram is to construct the model of the desired distribution from the generated data. The usual form of the histogram is to divide the range of data points  $N$  into equal-size bins or subintervals  $\Delta x$ . Or the necessary number of bins  $K$  needed will be given by

$$\text{Number of bins } K = \frac{\text{range of data } N}{\text{bin width } \Delta x} \quad (15.2.2)$$

Next, the number of data points  $\{n_i, i = 1, \dots, K\}$  falling into each bin is counted. The vertical axis (ordinate,  $y$ ) represents the counts in each bin, and the horizontal axis (abscissa,  $x$ ) represents  $K \cdot \Delta x$ , the product of the number of bins  $K$  and the bin interval  $\Delta x$ . If  $f_X(x)$  is finite, then the product  $K \cdot \Delta x$  must cover the entire range of  $x$ . If it is infinite or semiinfinite, then the product can cover a reasonable range. Finally, the histogram must be *normalized* so that it can be compared to the desired density function. The normalization proceeds as follows. The normalized count is the count in each bin  $\{n_i, i = 1, \dots, K\}$ , divided by the total number  $N$  of observations multiplied by the bin width  $\Delta x$  so that the area under the histogram equals one. This type of normalization results in a histogram that is similar to the probability density function. Even though this may be less intuitive, it is the one to be used if the histogram is to model the desired probability density function.

We will now model a uniform distribution  $U(0,1)$  from the random variates generated by the linear congruential generator of a computer. For this model,  $N = 10,000$  points were generated and  $K = 100$  bins were chosen with each bin interval  $\Delta x = 0.01$ . Note that  $K \cdot \Delta x = 1$  covers the  $x$  axis of the  $U(0,1)$  distribution. The histogram subroutine is called that places the counts  $\{n_i, i = 1, \dots, K\}$  in each of the  $K$  bins. The number of points falling in the first 10 bins is shown in Table 15.2.1. The expected number in each bin is 100.

The resulting histogram is divided by  $N \cdot \Delta x = 10,000 \times 0.01 = 100$  to yield the model for  $U(0,1)$  distribution. The normalized histogram is shown in Fig. 15.2.1.

Superimposed over the histogram is the actual uniform distribution. By “eyeballing” the two plots in Fig. 15.2.1, we see that the modeling is indeed good. More formally, we can also test the goodness of fit by a number of mathematical tests, and we will use the chi-square test to ascertain whether the histogram fits the true pdf sufficiently well. We will now enumerate the procedure for the chi-square test as applied to this case.

TABLE 15.2.1

Bins	0	1	2	3	4	5	6	7	8	9
Count	102	100	109	100	117	107	86	93	86	81
Expected	100	100	100	100	100	100	100	100	100	100

### Chi-Square Test

1. From the histogram find the number of observations  $n_i$  in each of the  $i = 1, \dots, K$  bins.
2. From the probability density function  $f_X(x)$  calculate the occupancy  $m_i = f_X(x_i)N\Delta x$  in each of the  $i = 1, \dots, K$  bins.
3. Calculate the deviation  $D_i$  from  $D_i = n_i - m_i$ ,  $i = 0, \dots, K - 1$ .
4. Form the chi-square statistic from  $\chi = \sum_{i=0}^{K-1} (D_i/m_i)^2$ . This will have  $K-2$  degrees of freedom.
5. Set the probability level  $p_\gamma$  or the significance level  $\alpha = 1 - p_\gamma$ . Usually  $p_\gamma$  is set at 95% or 99% or  $\alpha$  is set at 5% or 1%. The higher the significance level, the tighter will be the bounds.
6. From the chi-square tables find the threshold value  $t_\alpha$  for the given significance level and  $K-2$  degrees of freedom.
7. If the  $\chi$  value is less than  $t_\alpha$ , the hypothesis that the fit is good is accepted. Otherwise it is rejected.

We can use the  $\chi^2$  test to determine the goodness of the fit to the uniform distribution that has been achieved. The  $\chi$  value is found to be  $\chi = 91.39$  at the 5% significance level. The corresponding  $t_\alpha$  value for  $100 - 2 = 98$  degrees of freedom at the 5% significance level is obtained from Table 15.2.2 as  $t_\alpha = 122.10773$ . The chi-square density and distribution for 98 degrees of freedom is shown in Figs. 15.2.2 and 15.2.3 with the probability and significance levels. The hypothesis that the histogram is a good fit to the true normal distribution is accepted since  $91.39 < 122.10773$ .

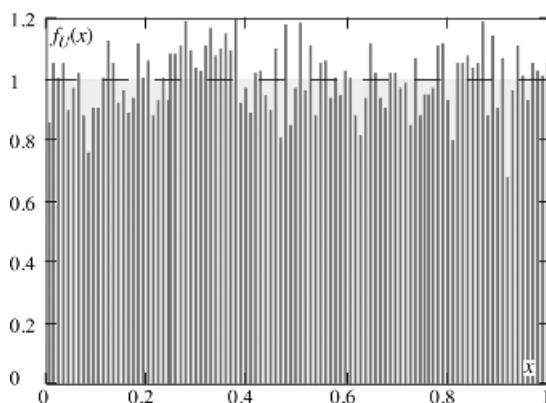


FIGURE 15.2.1

TABLE 15.2.2

$K$	$p_{\gamma} 95\%$	$p_{\gamma} 97.5\%$	$p_{\gamma} 99\%$	$K$	$p_{\gamma} 95\%$	$p_{\gamma} 97.5\%$	$p_{\gamma} 99\%$
1	3.84146	5.02389	6.63490	35	49.80185	53.20335	57.34207
2	5.99146	7.37776	9.21034	40	55.75848	59.34170	63.69074
3	7.81473	9.34840	11.34487	45	61.65623	65.41016	69.95683
4	9.48773	11.14329	13.27670	50	67.50481	71.42019	76.15389
5	11.07050	12.83250	15.08627	55	73.31149	77.38046	82.29212
6	12.59159	14.44938	16.81189	60	79.08194	83.29767	88.37942
7	14.06714	16.01276	18.47531	65	84.82064	89.17714	94.42208
8	15.50731	17.53455	20.09024	70	90.53122	95.02318	100.42518
9	16.91898	19.02277	21.66599	75	96.21667	100.83934	106.39292
10	18.30704	20.48306	23.20918	80	101.87947	106.62857	112.32879
12	21.02601	23.33660	26.21693	82	104.13874	108.93729	114.69489
14	23.68475	26.11891	29.14122	84	106.39484	111.24226	117.05654
16	26.29620	28.84532	31.99992	86	108.64789	113.54360	119.41390
18	28.86928	31.52636	34.80530	88	110.89800	115.84143	121.76711
20	31.41042	34.16959	37.56623	90	113.14527	118.13589	124.11632
22	33.92443	36.78070	40.28936	92	115.38979	120.42708	126.46165
24	36.41502	39.36407	42.97982	94	117.63165	122.71511	128.80325
26	38.88513	41.92316	45.64168	96	119.87094	125.00007	131.14122
28	41.33713	44.46079	48.27823	98	122.10773	127.28207	133.47567
30	43.77297	46.97924	50.89218	100	124.34211	129.56120	135.80672

Starting from uniformly distributed random variates, we can generate other random variates given the desired probability distribution. We will discuss three widely used techniques for generating these variates:

1. Inverse transformation
2. Convolution
3. Acceptance–rejection

In all the simulations that follow we have used  $u_i \sim U(0,1)$  distribution shown in Section 15.2 with  $N = 10,001$  points.

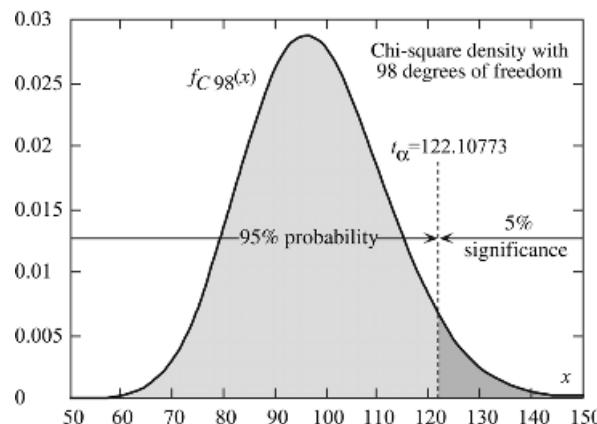


FIGURE 15.2.2

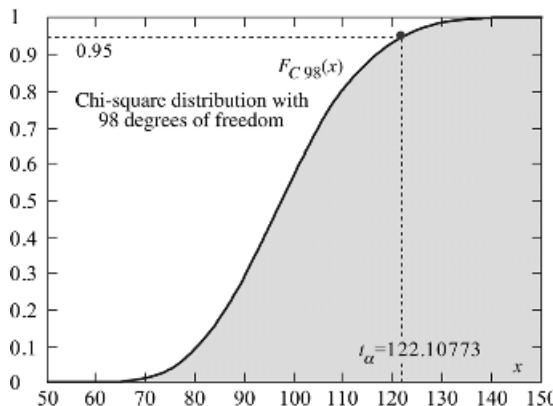


FIGURE 15.2.3

## 15.3 INVERSE TRANSFORMATION TECHNIQUES

One of the more useful ways of generating random variates is through the inverse transformation technique, which is based on the following result. If  $U$  is a random variable uniformly distributed in  $(0,1]$  with distribution function  $F_U(u) = u, u \in (0,1]$ , and  $X$  is a random variable with distribution function  $F_X(x)$ , then the random variates for  $X$  can be generated from  $X = F_X^{-1}(U)$ . We will show that  $F_X^{-1}(U)$  is a random variable with distribution  $F_X(x)$ . Assuming that  $X = F_X^{-1}(U)$ , we have

$$P\{X \leq x\} = P\{F_X^{-1}(U) \leq x\} \quad (15.3.1)$$

Since  $F_X(x)$  is monotonically increasing, we can write Eq. (15.3.1) as follows:

$$P\{F_X^{-1}(U) \leq x\} = P\{U \leq F_X(x)\} = F_U\{F_X(x)\} \quad (15.3.2)$$

However,  $U$  is uniformly distributed in  $(0,1]$  and hence  $F_U(u) = u, 0 < u \leq 1$ . Applying this fact in Eq. (15.3.2), we have

$$F_U\{F_X(x)\} = F_X(x) \quad (15.3.3)$$

This result can be used to obtain the random variates for any distribution function  $F_X(x)$  only if  $F_X^{-1}$  can be solved in closed form or  $F_X$  can be accurately and efficiently inverted numerically. The methodology of finding random variates  $\{X_i\}$  by the inversion technique can be enumerated as follows:

1.  $F_X(x)$  is known a priori.
2. Uniform random variates  $\{u_i, i = 1, \dots, N\}$  in the interval  $(0,1]$  are generated.
3. The inverse  $F_X^{-1}(u)$  of the cdf  $F_X(x)$  can be analytically or numerically determined.
4. The variates  $\{x_i\}$  are generated from  $x_i = F_X^{-1}(u_i), i = 1, \dots, N$ .
5. The histogram for  $\{x_i\}$  is constructed and tested using chi-square goodness of fit.

We will use this procedure to find random variates for a number of discrete and continuous distributions starting from 10,001 points of the uniform random variates in  $(0,1]$  generated in the previous section.

### Discrete Distributions

Since  $F_X(x)$  is a monotonically increasing function, the inverse for a finite-dimensional integer-valued discrete distribution functions can be obtained in a simple manner as shown in the following two examples.

**Example 15.3.1** The distribution of packets in bytes in a communication system is given by

$$F_X(x) = \begin{cases} 0, & 0 < x \leq 64 \\ 0.2, & 64 < x \leq 128 \\ 0.4, & 128 < x \leq 256 \\ 0.6, & 256 < x \leq 320 \\ 0.8, & 320 < x \leq 384 \\ 1, & x > 384 \end{cases}$$

and is shown in Fig. 15.3.1.

It is not difficult to find the inverse function  $F_X^{-1}(u)$ . From Fig. 15.3.1, for values of the uniform random variate  $u_i$  given by  $0.4 < u_i \leq 0.6$ ,  $F_X^{-1}(u) = 256$  bytes. Similarly, we can obtain the complete inverse function as

$$x_i = F_X^{-1}(u_i) = \begin{cases} 64, & 0 < u_i \leq 0.2 \\ 128, & 0.2 < u_i \leq 0.4 \\ 256, & 0.4 < u_i \leq 0.6 \\ 320, & 0.6 < u_i \leq 0.8 \\ 384, & 0.8 < u_i \leq 1 \end{cases}$$

and the sequence  $\{x_i\}$  represents the random variates for this discrete distribution. The generated uniform random variates  $u_i$  are assigned values 64, 128, 256, 320, and 384 depending on the ranges in which they fall along the vertical axis. They are then sorted in ascending order and plotted with  $x = F_X^{-1}(u)$  along the horizontal axis and  $i$  scaled by  $1/10,000$  along the vertical axis so that the range is from 0 to 1. The plot is shown in Fig. 15.3.2. As an example, the uniform random variates the falling in the shaded region between 0.4 and 0.6 represent 256 bytes.

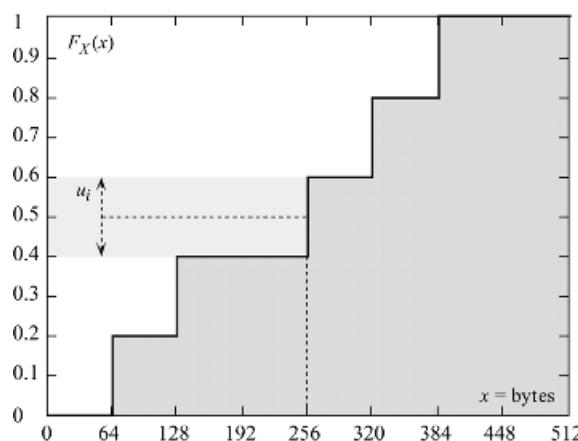


FIGURE 15.3.1

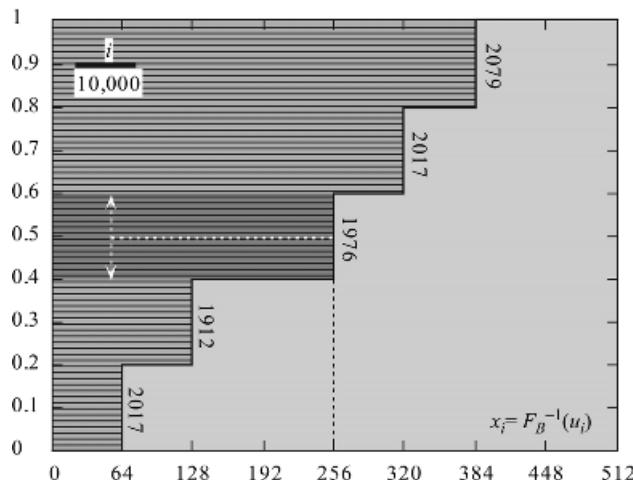


FIGURE 15.3.2

Out of the 10,001 uniform random variates generated, 2017 fall in the range 0–0.2, 1912 fall in the range 0.2–0.4, 1976 fall in the range 0.4–0.6, 2017 fall in the range 0.6–0.8, and 2079 fall in the range 0.8–1.0 as shown in Fig. 15.3.2, which is exactly the same as Fig. 15.3.1.

**Example 15.3.2** In this example, we will simulate a binomial distribution  $F_B(x)$  with  $n = 5$  and  $p = 0.4$  given by

$$F_B(x) = \sum_{k=0}^x \frac{5!}{k!(5-k)!} (0.4)^k (0.6)^{5-k}$$

and shown below and graphed in Fig. 15.3.3:

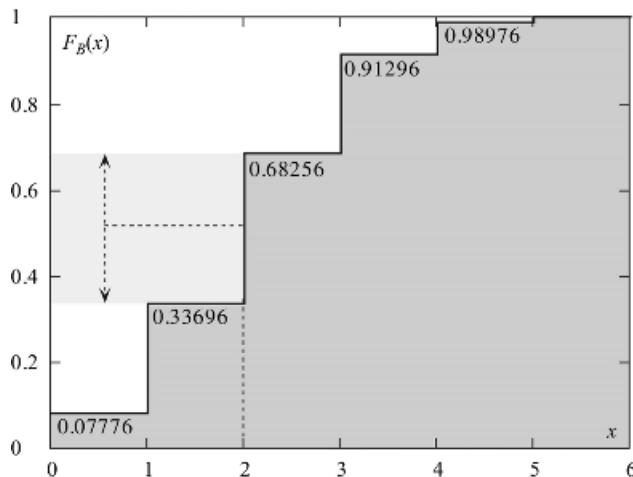


FIGURE 15.3.3

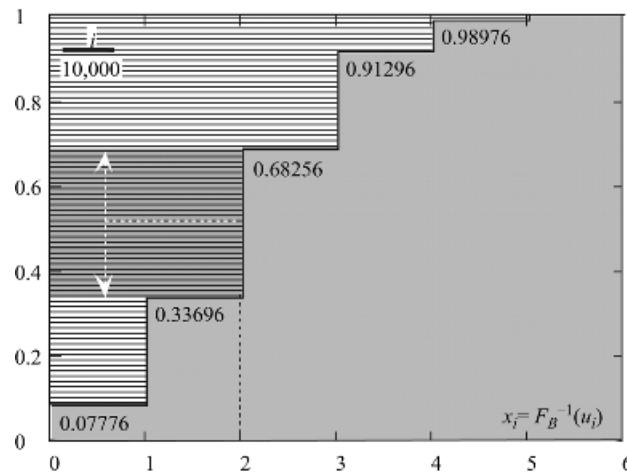


FIGURE 15.3.4

$$F_B(x) = \begin{cases} 0.07776, & 0 < x \leq 1 \\ 0.33696, & 1 < x \leq 2 \\ 0.68256, & 2 < x \leq 3 \\ 0.91296, & 3 < x \leq 4 \\ 0.98976, & 4 < x \leq 5 \\ 1, & x > 5 \end{cases}$$

The inverse function  $F_B^{-1}(U)$  can be found in a manner similar to that in the previous example. In Fig. 15.3.3, for values of the uniform random variate  $u_i$  between 0.33696 and 0.68256,  $F_B^{-1}(u) = 2$ . Similarly, we can obtain the complete inverse function as

$$F_B^{-1}(u_i) = \begin{cases} 0, & 0 < u_i \leq 0.07776 \\ 1, & 0.07776 < u_i \leq 0.33696 \\ 2, & 0.33696 < u_i \leq 0.68256 \\ 3, & 0.68256 < u_i \leq 0.91296 \\ 4, & 0.91296 < u_i \leq 0.98976 \\ 1, & 0.98976 < u_i \leq 1 \end{cases}$$

which is shown in Fig. 15.3.4.

The vertical axis has been scaled by  $\frac{1}{10,000}$  for the range to correspond to  $(0,1]$ .

Out of the 10,001 uniform random variates generated, 797 fall in the range 0–0.07776, 2555 fall in the range 0.07776–0.33696, 3404 fall in the range 0.33696–0.68256, 2312 fall in the range 0.68256–0.91296, 801 fall in the range 0.91296–0.98976, and 132 fall in the range 0.98976–1. Figure 15.3.4 is the same as Fig. 15.3.3.

### Continuous Distributions

If the cumulative distribution function of a random variable can be expressed in a closed form, then it is not difficult to find the inverse in a closed form. We will give several examples to clarify the inverse transformation method.

**Example 15.3.3 (Exponential Distribution)** In this example we will obtain random variates for an exponential distribution given by

$$F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0 \quad (15.3.4)$$

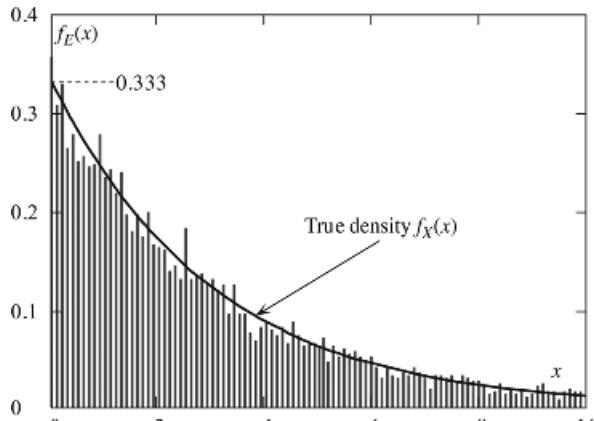


FIGURE 15.3.5

The inverse is given by

$$1 - e^{-\lambda x} = u \quad \text{or} \quad e^{-\lambda x} = 1 - u \quad \text{and} \quad x = F_X^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u), \quad x \geq 0$$

The exponential random variates are generated from this equation with  $\lambda = \frac{1}{3}$ , resulting in the following equation:

$$x_i = -3 \ln(1 - u_i), \quad x \geq 0, \quad i = 0, 1, \dots, 10,000$$

The histogram subroutine `hist(K, Δx, x)` is called with  $K = 100$  bins and bin width  $Δx = 0.1$ . The resulting histogram is divided by  $N \cdot Δx = 1000$  to yield the normalized histogram  $f_E(x_k)$ ,  $k = 1, \dots, 100$ . This is shown in Fig. 15.3.5. Superimposed on top of the histogram is the true density function,  $f_X(x) = \frac{1}{3} e^{-(x/3)}$ ,  $x > 0$ . We can see from figure that the fit is very good. We can also apply the chi-square goodness of fit, which it easily passes.

**Example 15.3.4 (Weibull Distribution)** The cdf of a Weibull distribution is given by

$$F_X(x) = 1 - e^{-[\lambda(x-\mu)]^\alpha}, \quad x > \mu \quad (15.3.5)$$

This distribution, as mentioned in Chapter 7, is a generalization of the exponential with parameters, scale  $\lambda$ , shape  $\alpha$ , and location  $\mu$ . Since this is in a closed form, the inverse is given by

$$1 - e^{-[\lambda(x-\mu)]^\alpha} = u \quad \text{or} \quad e^{-[\lambda(x-\mu)]^\alpha} = 1 - u \quad \text{and} \quad F_X^{-1}(u) = \mu + \frac{1}{\lambda} \sqrt[\alpha]{-\ln(1 - u)}$$

with  $u < 1$ . The Weibull random variates are generated with  $\lambda = \frac{1}{3}$ ,  $\alpha = 3$  and  $\mu = 2$  from the following equation:

$$x_i = 2 + 3 \sqrt[3]{-\ln(1 - \mu_i)}, \quad i = 0, \dots, 10,000$$

The histogram subroutine `hist(K, Δx, x)` is called with  $K = 100$  bins and bin width  $Δx = 0.1$ . The resulting histogram is divided by  $N \cdot Δx = 1000$  to yield the normalized histogram  $f_W(x_k)$ ,  $k = 1, \dots, 100$ . This is shown in Fig. 15.3.6. Superimposed on top of the histogram is the true Weibull density function,  $f_X(x) = 3 \left(\frac{1}{3}\right)^3 (x-2)^2 e^{-[(x-2)/3]^3}$ ,  $x > 2$ .

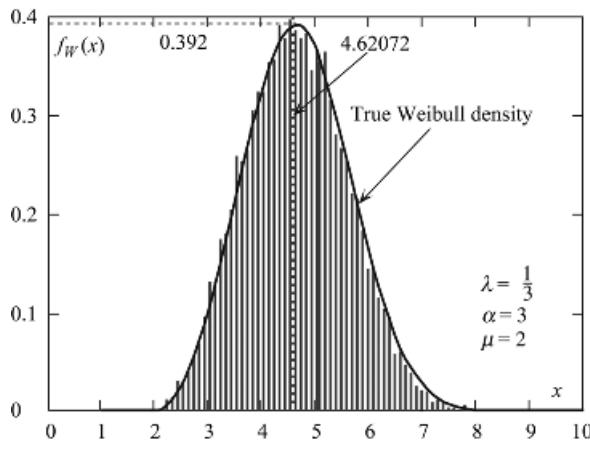


FIGURE 15.3.6

We can see from the figure that the fit is very good. We can also apply the chi-square goodness of fit, which it readily passes.

**Example 15.3.5 (Rayleigh Distribution)** The random variates corresponding to a Rayleigh distribution can be generated by taking (1) the direct inverse of the distribution or (2) the square root of the sum of the squares of two independent Gaussian distributions. The cdf of a Rayleigh distribution is given by

$$F_Z(z) = 1 - e^{-(z^2/2\sigma^2)}, \quad z > 0 \quad (15.3.6)$$

Since this is in a closed form, the inverse is found as follows:

$$1 - e^{-(z^2/2\sigma^2)} = u \quad \text{or} \quad e^{-(z^2/2\sigma^2)} = 1 - u \quad \text{and} \quad z = F_Z^{-1}(u) = \sqrt{-2\sigma^2 \ln(1 - u)}$$

The Rayleigh random variates are generated from this equation with  $\sigma^2 = 1$ , resulting in the following equation:

$$z_i = \sqrt{-2\sigma^2 \ln(1 - u_i)}, \quad z > 0, \quad i = 0, 1, \dots, 10,000$$

The histogram subroutine `hist(K, Δx, x)` is called with  $K = 200$  bins and bin width  $Δx = 0.05$ . The resulting histogram is divided by  $N \cdot Δx = 500$  to yield the normalized histogram  $f_R(x_k)$ ,  $k = 1, \dots, 200$ . This is shown in Fig. 15.3.7. Superimposed on top of the histogram is the true density function,  $f_Z(z) = ze^{-(z^2/2)}$ ,  $z > 0$ . We can see from figure that the fit is very good. We can also apply the chi-square goodness of fit, which it easily passes.

### Special Methods

Even though some distributions may not have a closed form, random variates can be obtained by special transformations. The very important Gaussian random variates can be generated by these special transformation techniques. Two independent random variables  $U$  and  $V$  are both uniformly distributed in  $(0,1]$  and whose joint density is given by

$$f_{UV}(u, v) = \begin{cases} 1, & 0 < u \leq 1, 0 < v \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (15.3.7)$$

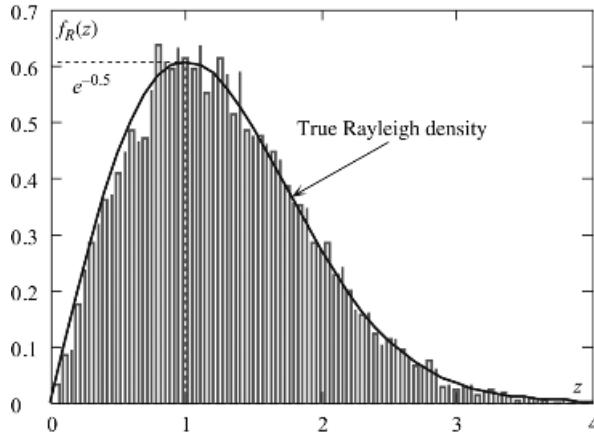


FIGURE 15.3.7

The transformation used to generate standard Gaussian random variables is

$$X = \sqrt{-2 \ln(U)} \cos(2\pi V) \\ Y = \sqrt{-2 \ln(U)} \sin(2\pi V) \quad (15.3.8)$$

We can find the joint density  $f_{XY}(x,y)$  from the transformation. Solving for  $u$  and  $v$  in Eq. (15.3.8), we obtain

$$u = e^{-(x^2+y^2)/2} : v = \frac{1}{2\pi} \tan^{-1}\left(\frac{y}{x}\right) \quad (15.3.9)$$

The Jacobian matrix of transformation is

$$J = \begin{bmatrix} \frac{dx}{du} & \frac{dx}{dv} \\ \frac{dy}{du} & \frac{dy}{dv} \end{bmatrix} = \begin{bmatrix} -\cos(2\pi v) & -2\pi \sin(2\pi v) \sqrt{-2 \ln(u)} \\ \frac{u \sqrt{-2 \ln(u)}}{u^2 + y^2} & \frac{-\sin(2\pi v)}{u^2 + y^2} \sqrt{-2 \ln(u)} \\ \frac{-\sin(2\pi v)}{u^2 + y^2} & \frac{2\pi \cos(2\pi v) \sqrt{-2 \ln(u)}}{u^2 + y^2} \end{bmatrix} \quad (15.3.10)$$

and the absolute value of the Jacobian determinant is

$$\|J\| = \begin{vmatrix} -\cos(2\pi v) & -2\pi \sin(2\pi v) \sqrt{-2 \ln(u)} \\ \frac{u \sqrt{-2 \ln(u)}}{u^2 + y^2} & \frac{-\sin(2\pi v)}{u^2 + y^2} \sqrt{-2 \ln(u)} \\ \frac{-\sin(2\pi v)}{u^2 + y^2} & \frac{2\pi \cos(2\pi v) \sqrt{-2 \ln(u)}}{u^2 + y^2} \end{vmatrix} = \frac{2\pi}{u} = 2\pi e^{(x^2+y^2)/2} \quad (15.3.11)$$

Hence, from Eqs. (15.3.7) and (15.3.11) we have

$$f_{XY}(x,y) = \frac{f_{UV}(u,v)}{\|J\|} = \frac{1}{2\pi} e^{-(x^2+y^2)/2} = \left( \frac{1}{\sqrt{2\pi}} e^{(x^2/2)} \right) \left( \frac{1}{\sqrt{2\pi}} e^{-(y^2/2)} \right) \quad (15.3.12)$$

which yields two independent standard Gaussian random variables. This transformation is called the *Box–Mueller transformation*.

**Example 15.3.6 (Gaussian Distribution)** We will now use the Box–Mueller transformation to generate two independent standard Gaussian random variates. In this example, 10,001 points of two independent uniformly distributed random variates  $\{u_i\}$  and  $\{v_i\}$  are generated. They are passed through the transformation of Eq. (15.3.8) to form the

Gaussian random variates  $\{z_i\}$  and  $\{w_i\}$ . Theoretically, these must be zero mean and unit variance. If they are not, they can be transformed into exactly zero mean and unit variance by another transformation

$$x_i = \frac{z_i - \bar{z}}{\bar{\sigma}_z}, \quad y_i = \frac{w_i - \bar{w}}{\bar{\sigma}_w}, \quad i = 0, \dots, \leq 10,000$$

where  $\bar{z} = (1/N) \sum_{i=0}^N z_i$  and  $\bar{\sigma}_z = \sqrt{(1/N) \sum_{i=0}^N (z_i - \bar{z})^2}$ ,  $N = 10,000$ , with similar definitions for  $\bar{w}$  and  $\bar{\sigma}_w$ . The histogram subroutines are called with  $K = 100$  bins and bin width  $\Delta x = \Delta y = 0.08$ . The resulting histogram is normalized by  $N\Delta x = N\Delta y = 800$  to yield histograms  $f_{X_k}$  and  $f_{Y_k}$ ,  $k = 1, \dots, 100$  for the Gaussian random variates. Both plots are shown in Fig. 15.3.8 with the true standard Gaussian  $f_X(x) = (1/\sqrt{2\pi})e^{-(x^2/2)}$  and  $f_Y(y) = (1/\sqrt{2\pi})e^{-(y^2/2)}$  superimposed on them. From these figures we find that the fit is very good. These random variates are stored for later use in simulating stochastic systems.

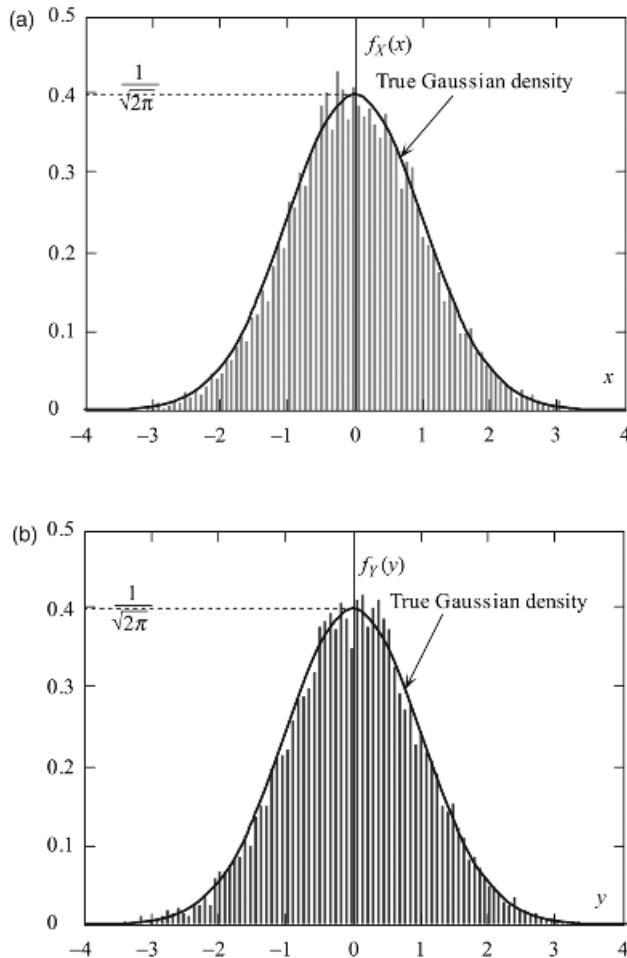


FIGURE 15.3.8

**Example 15.3.7 (Rayleigh Distribution with Phase)** We have already found random variates for a Rayleigh distribution using a direct transformation. We will now apply the relationship to two independent standard Gaussian random variables by

$$Z = \sqrt{X^2 + Y^2}$$

to generate random variates for the Rayleigh distribution. The Gaussian random variates are generated by the Box–Mueller transformation as explained in the previous section. The Rayleigh random variates and the corresponding phase variates are given by

$$\left. \begin{aligned} z_i &= \sqrt{x_i^2 + y_i^2} \\ w_i &= \tan^{-1}\left(\frac{y_i}{x_i}\right) \end{aligned} \right\}, \quad i = 0, \dots, N$$

The sequences  $\{z_i\}$  and  $\{w_i\}$  are both passed through histogram subroutines with the number of bins  $K = 200$  and the bin width  $\Delta z = \Delta w = 0.05$ . Both these histograms are normalized with division by  $N\Delta z = N\Delta w = 500$ . The resulting histograms  $f_{Zk}$  and  $f_{Wk}$  are shown in Fig. 15.3.9 with the true Rayleigh density  $f_Z(z) = ze^{-(z^2/2)}$ ,  $z > 0$  and uniform density  $U(-\pi/2, \pi/2]$  superimposed on them. We can see from both plots that the fit is very good.

Table 15.3.1 shows the densities with their closed-form distribution functions and their inverses.

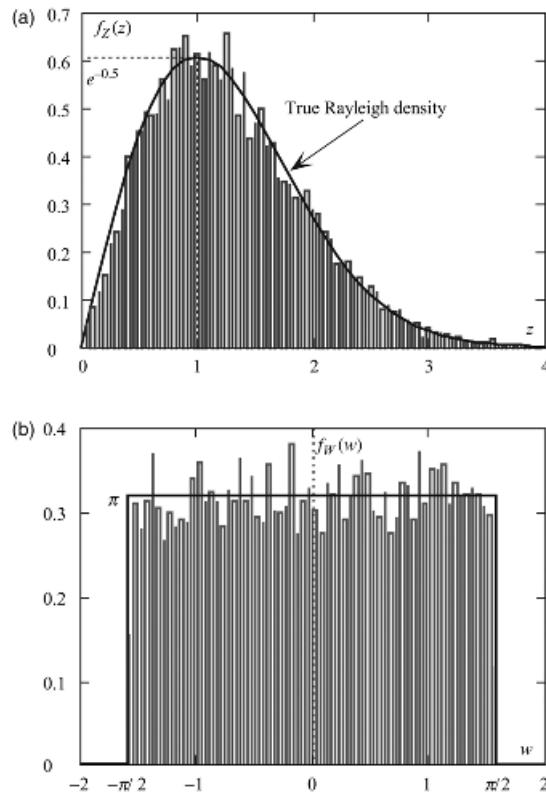


FIGURE 15.3.9

TABLE 15.3.1

Type	pdf $f(x)$	cdf $F(x)$	Inverse $F^{-1}(u)$
Uniform $(a,b)$	$\frac{1}{b-a}$	$\frac{x-a}{b-a}$	$a + u(b-a)$
Triangular	$\frac{1}{2} \left( \frac{x-a}{b-a} \right)^2 \quad a < x \leq b$ $1 - \frac{1}{2} \left( \frac{x+a-2b}{b-a} \right)^2 \quad b < x \leq 2b-a$	$a + (b-a)\sqrt{2u}, \quad u \leq \frac{1}{2}$ $b + (b-a) \left[ 1 - \sqrt{2(1-u)} \right] \quad u > \frac{1}{2}$	
Exponential	$\frac{1}{\lambda} e^{-\lambda x} u(x)$		$-\frac{1}{\lambda} \ln(1-u), \quad u < 1$
Laplace	$\frac{\lambda}{2} e^{-\lambda x }, \quad \lambda \geq 0$	$\frac{1}{2} e^{\lambda x}, \quad x \leq 0 \quad 1 - \frac{1}{2} e^{-\lambda x}, \quad x > 0$	$\frac{1}{\lambda} \ln(2u), \quad u \leq \frac{1}{2} - \frac{1}{\lambda} \ln[2(1-u)], \quad u > \frac{1}{2}$
Weibull	$\alpha \lambda^\alpha (x - \mu)^{\alpha-1} e^{-[\lambda(x-\mu)]^\alpha} \quad x > \mu$	$1 - e^{-[\lambda(x-\mu)]^\alpha}, \quad x > \mu$	$\mu + \frac{1}{\lambda} \sqrt{-\ln(1-u)}, \quad u < 1$
Rayleigh	$\frac{z}{\sigma^2} e^{-(z^2/\sigma^2)}, \quad z > 0$	$1 - e^{-(z^2/2\sigma^2)}, \quad z > 0$	$\sqrt{-2\sigma^2 \ln(1-u)}, \quad u < 1$
Pareto	$\frac{\alpha}{\beta} \left( \frac{\beta}{x+\beta} \right)^{\alpha+1}, \quad x > 0$	$1 - \left( \frac{\beta}{x+\beta} \right)^\alpha, \quad x > 0$	$\beta \left[ \frac{1}{\sqrt[alpha]{1-u}} - 1 \right], \quad u < 1$
Cauchy	$\frac{\beta}{\pi[(x-\alpha)^2 + \beta^2]}$		$\frac{1}{2} + \frac{1}{\pi} \tan^{-1} \left( \frac{x-\alpha}{\beta} \right)$
Gaussian	$\frac{1}{2\pi} e^{-(x^2+y^2)/2}$		$x = \sqrt{-2 \ln(u_1)} \cos(2\pi u_2)$ $y = \sqrt{-2 \ln(u_1)} \sin(2\pi u_2)$

## 15.4 CONVOLUTION TECHNIQUES

In cases where it may not be possible to obtain inverses, we can use the convolution technique. From Eq. (13.1.15) we recall that the density of the sum  $Z$  of two independent random variables  $X$  and  $Y$  is the convolution of the densities of  $X$  and  $Y$ :

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx \quad (15.4.1)$$

We can generate random variates from independent random variates that have already been generated. We will give two examples.

**Example 15.4.1 (Triangular Distribution)** We want to generate random variates for a triangular distribution. Two independent uniformly distributed random variates,  $\{a_i\}$  and  $\{b_i\}$ , are generated. We form the sum

$$x_i = a_i + b_i, \quad i = 0, \dots, 10,000$$

The sequence  $\{x_i\}$  is passed through the histogram subroutine with the number of bins  $K = 100$  and the bin width  $\Delta x = 0.02$ . The resulting histogram is normalized with  $N\Delta x = 10,000 \times 0.02 = 200$ . The histogram  $f_{X_k}$  is graphed in Fig. 15.4.1. The true triangular density given by

$$f_X(x) = \begin{cases} x, & 0 < x \leq 1 \\ 2-x, & 1 < x \leq 2 \end{cases}$$

is superimposed on the histogram. The fit is very close and passes the chi-square test easily.

**Example 15.4.2 (Erlang Distribution)** The Erlang distribution is obtained by the  $n$ -fold convolution of  $n$  independent exponential distributions as given in Eq. (7.4.3). Hence we can generate Erlang random variates with  $n$  degrees of freedom by summing  $n$  independent exponential random variates. We will now generate Erlang random variates with 2 degrees

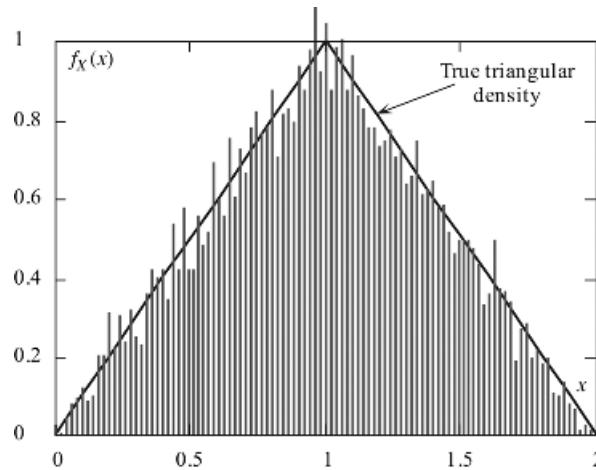


FIGURE 15.4.1

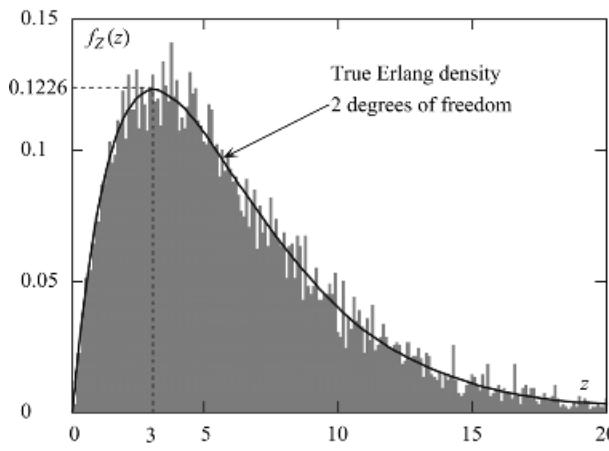


FIGURE 15.4.2

of freedom. From two sets of uniform random variates  $\{a_i\}$  and  $\{b_i\}$ , we generate two identically distributed exponential random variates  $\{x_i\}$  and  $\{y_i\}$ , with  $\lambda = \frac{1}{3}$  given by

$$x_i = -3 \ln(a_i); \quad y_i = -3 \ln(b_i)$$

Erlang random variates  $\{z_i\}$  are formed by summing  $\{x_i\}$  and  $\{y_i\}$ . The variates  $\{z_i\}$  are then passed through the histogram subroutine with number of bins  $K = 200$  and bin width  $\Delta z = 0.1$ . The resulting histogram is normalized by division with  $N\Delta x = 1000$ . The histogram  $f_{Z_k}$  is graphed for  $k = 0, \dots, 200$  and shown in Fig. 15.4.2. The true Erlang density for 2 degrees of freedom given by

$$f_Z(z) = \left(\frac{1}{3}\right)^2 z e^{-(z/3)}, \quad z \geq 0$$

is superimposed on the histogram as shown in the figure. The fit is very good as seen in the diagram.

## 15.5 ACCEPTANCE–REJECTION TECHNIQUES

The rejection method does *not* require that the cumulative distribution function [indefinite integral of  $p(x)$ ] be readily computable, much less the inverse of that function—which was required for the transformation method in the previous section.

Usually the inverse transformation technique uses the cumulative distribution function in closed form, in which case it is the preferred method of generating random variates. However, in many cases the cdf may not be expressible in closed form or it may be very inefficient to calculate it numerically. On the other hand, the density functions may be available in closed form such as the beta or the gamma density. The *rejection method* is a technique for generating random deviates  $\{x_i\}$  whose density function  $f_X(x)$ , if known in closed form, is amenable directly to determine these variates. We will now outline the *acceptance–rejection* method of generating variates.

Consider a function  $g(x)$  that upper-bounds the density  $f_X(x)$  or  $g(x) \geq f_X(x)$  for all  $x$ . This is called the *bounding* or *comparison* function  $g(x)$  which, is *not* a density function

since  $\int_{-\infty}^{\infty} g(x)dx \geq \int_{-\infty}^{\infty} f_X(x)dx = 1$ . We define a normalization constant  $c = \int_{-\infty}^{\infty} g(x)dx \geq 1$ , and another function  $w_X(x) = g(x)/c$  for all  $x$ , thus rendering  $w_X(x)$  a proper density function.

We can now state the acceptance–rejection algorithm:

1. Generate random variates  $\{y_i, i = 0, \dots, N\}$  having the density function  $w_Y(y)$ .
2. Generate random variates  $\{u_i, i = 0, \dots, N\}$  uniformly distributed in  $(0,1]$  and independent of  $\{y_i\}$ .
3. If  $u_i \leq [f_X(y_i)]/[g(y_i)] \leq 1$ , accept the point  $y_i$  and set  $\{x_i = y_i\}$ . Otherwise, reject  $y_i$ . The number of accepted points  $\{x_i\}$  will be equal to  $N/c$ .
4. The sequence  $\{x_i\}$  are the desired random variates with density function  $f_X(x)$ .

We will now show from conditional probability considerations that the generated variates  $\{x_i\}$  do indeed have the distribution  $F_X(x)$ . From the definition of conditional probability, we obtain

$$P(\text{generated } X \leq x) = P(Y \leq x \mid \text{accept}) = \frac{P(Y \leq x, \text{accept})}{P(\text{accept})} \quad (15.5.1)$$

We need the following result for any  $y$ :

$$P(\text{accept} \mid Y = y) = P\left(U \leq \frac{f_X(y)}{g(y)}\right) \quad (15.5.2)$$

Since  $U$  is uniformly distributed in  $(0,1]$ , we have  $F_U(u) = u$ , and it is independent of  $Y$ . Hence Eq. (15.5.2) can be written as follows:

$$P\left(U \leq \frac{f_X(y)}{g(y)}\right) = F_U\left(\frac{f_X(y)}{g(y)}\right) = \left(\frac{f_X(y)}{g(y)}\right) \quad (15.5.3)$$

Incorporating Eq. (15.5.3) into Eq. (15.5.2), we obtain

$$P(\text{accept} \mid Y = y) = \frac{f_X(y)}{g(y)} \quad (15.5.4)$$

We can evaluate the denominator in Eq. (15.5.1) from Eq. (15.5.4) with substitution of  $w_Y(y) = g(y)/c$ .

$$\begin{aligned} P(\text{accept}) &= \int_{-\infty}^{\infty} P(\text{accept} \mid Y = y)w_Y(y)dy \\ &= \int_{-\infty}^{\infty} \frac{f_X(y)}{g(y)} \frac{g(y)}{c} dy = \frac{1}{c} \end{aligned} \quad (15.5.5)$$

We will now evaluate the numerator of Eq. (15.5.1):

$$P(Y \leq x, \text{accept}) = \int_{-\infty}^{\infty} P(\text{accept}, Y \leq x \mid Y = y)w_Y(y)dy \quad (15.5.6)$$

Since  $Y = y$  in Eq. (15.5.6) we can rewrite it as

$$P(Y \leq x, \text{accept}) = \int_{-\infty}^x P(\text{accept}, y \leq x \mid Y = y)w_Y(y)dy \quad (15.5.7)$$

where the integration is terminated at  $x$  because  $y \leq x$ . Applying Eq. (15.5.4) and substituting  $w_Y(y) = g(y)/c$  in Eq. (15.5.7), we obtain

$$\begin{aligned} P(Y \leq x, \text{ accept}) &= \int_{-\infty}^x \frac{f_X(y)g(y)}{c} dy \\ &= \frac{F_X(x)}{c} \end{aligned} \quad (15.5.8)$$

Combining Eqs. (15.5.5) and (15.5.8), we obtain the result that

$$P(\text{generated } X \leq x) = \frac{F_X(x)}{c} = F_X(x) \quad (15.5.9)$$

thus showing that the acceptance-rejection method yields the desired probability distribution.

**Example 15.5.1** We will give an example of generating beta density variates by the method described above. The general beta density is given by

$$f_\beta(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 \leq x \leq 1$$

and there is no closed-form solution for the cdf. For  $\alpha$  and  $\beta$  integers, we have

$$f_\beta(x) = \frac{(\alpha + \beta - 1)!}{(\alpha - 1)!(\beta - 1)!} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 \leq x \leq 1$$

Random variates will be generated for a beta(3,6) density (Fig. 15.5.1) given by

$$f_\beta(x) = 168x^2(1-x)^5, \quad 0 \leq x \leq 1$$

The maximum of the density function occurs at  $x = \frac{2}{7} = 0.2857$  and  $f_\beta(\frac{2}{7}) = 2.55$ . We can bound this density with a rectangle of width 2.55 so that  $g(x) = 2.55$ ,  $0 < x \leq 1$ . Hence the normalizing constant  $c = 2.55$  and  $w_X(x) = g(x)/c$  is a  $U(0,1]$  density function.

### Algorithm for Acceptance–Rejection

1. Generate a sequence  $\{y_i, i = 0, \dots, N\}$  with uniform density  $w_Y(y) = g(y)/c$ .
2. Generate uniform random variates  $\{u_i, i = 0, \dots, N\}$  independent of  $\{y_i\}$ .
3. If  $u_i \leq [168y_i^2(1-y_i)^5]/2.55$ , return  $x_i = y_i, i = 0, \dots, N$ .

Sequences  $\{y_i, i = 0, \dots, 10,000\}$  and  $\{u_i, i = 0, \dots, 10,000\}$  are generated, and if the condition 3 (above) is satisfied,  $x_i$  was set equal to  $y_i$ . The sequence  $\{x_i\}$  has 3936 points out of the original total of 10,001 points. This corresponds to the expected  $10,001/2.55 = 3922$  points. The sequence  $\{x_i\}$  consisting of 3936 points was passed through a histogram subroutine with  $K = 100$  bins and bin width  $\Delta x = 0.01$ . The histogram was normalized by division with  $3936 \times 0.01 = 39.36$ . The normalized histogram is graphed in Fig. 15.5.1 with the true beta(3,6) density superimposed for comparison. The bounding function  $g(x)$  is also shown. From the figure it is seen that the fit is good.

### Discussion

We have discussed three methods for generation of random variates for a given distribution function. The inversion method is used where the cumulative distribution

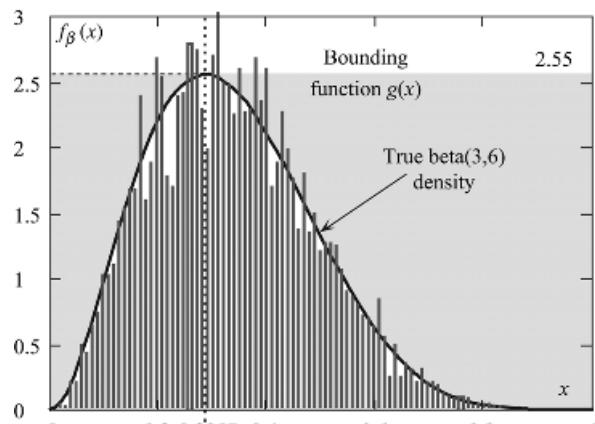


FIGURE 15.5.1

function can be expressed in a closed form or can be computed efficiently. The convolution method can be used when the density function is expressible as the resultant of a sum of independent random variates. The acceptance-rejection method is used directly on the probability density function. In generating histograms the number of bins  $K$  and the bin width  $\Delta x$  have to be manipulated first by “eyeballing” and later confirming that the  $\chi$  value is less than the  $t_\alpha$  value in the chi-square test at the significance level  $\alpha$ . Note also that higher significance levels yield tighter bounds.

# Elements of Matrix Algebra\*

A number of results depend on an understanding of matrices and manipulations of matrices. We will now discuss relevant aspects of matrix algebra. We will assume familiarity with linear vector spaces [56].

## 16.1 BASIC THEORY OF MATRICES

### Definitions

A *matrix*  $\mathbf{X}$  is defined as an array of  $m$  column vectors  $\{\mathbf{x}_j, j = 1, \dots, m\}$  with each column vector  $\mathbf{x}_j$  having  $n$  real or complex elements  $\mathbf{x}_j = \{x_{1j}, x_{2j}, \dots, x_{ij}, \dots, x_{nj}\}^T$ . Thus the matrix  $\mathbf{X}$  is an  $n \times m$  rectangular array of real or complex numbers and written as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_j & \cdots & \mathbf{x}_m \\ x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{im} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{nm} \end{bmatrix} \quad (16.1.1)$$

*Notation:* In the double subscript  $(i j)$  the first subscript denotes the row, and the second subscript denotes the column. Thus, an  $n \times m$  array consists of  $n$  rows and  $m$  columns. Sometimes the matrix  $\mathbf{X}$  is also written as  $\{x_{ij}\}$ . A *square matrix* is a matrix with equal number of rows and columns. It is denoted as  $n \times n$  square matrix or an  $n$  matrix. If the column vectors  $\{\mathbf{x}_j = [x_{1j}, x_{2j}, \dots, x_{nj}]^T, j = 1, 2, \dots, m\}$  are linearly independent, then the matrix  $\mathbf{X}$  is column-independent. On the other hand, if the row vectors denoted by  $\{\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{im}], i = 1, 2, \dots, n\}$  are linearly independent, then the matrix  $\mathbf{X}$  is row-independent. The dots in the column-vector  $\mathbf{x}_j$  and the row-vector  $\mathbf{x}_i$  represent the integers  $1, \dots, n$ . An arbitrary rectangular matrix cannot be both column- and row-independent unless it is a square matrix. However, not all square matrices need be row-independent and column-independent.

\*Parts of this chapter are based, with permission, on Chapters 30, 31, and 33 from the book *Linear Systems Properties—a Quick Reference*, by Venkatarama Krishnan, published by CRC Press, 1988.

### Matrix Addition

Addition is defined for two matrices  $\mathbf{X}$  and  $\mathbf{Y}$  of similar dimensions:

$$\mathbf{X} + \mathbf{Y} = \{x_{ij}\} + \{y_{ij}\} = \{x_{ij} + y_{ij}\} \quad (16.1.2)$$

The addition rule is commutative:  $\mathbf{X} + \mathbf{Y} = \mathbf{Y} + \mathbf{X}$ .

#### Example 16.1.1

$$\mathbf{X} = \begin{bmatrix} 1 & 5 & 2 \\ 3 & 4 & 7 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} 2 & 3 & 4 \\ 8 & 6 & 9 \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} 4 & 6 \\ 3 & 2 \end{bmatrix}$$

$$\mathbf{X} + \mathbf{Y} = \begin{bmatrix} 3 & 8 & 6 \\ 11 & 10 & 16 \end{bmatrix}, \quad \mathbf{X} + \mathbf{Z} \text{ and } \mathbf{Y} + \mathbf{Z} \text{ are not defined}$$

### Matrix Multiplication

Matrix multiplication is a little complex. If we denote  $\mathbf{X} \times \mathbf{Y} = \mathbf{Z}$ , then the number of columns of  $\mathbf{X}$  must be equal to the number of rows of  $\mathbf{Y}$ . Otherwise, multiplication is not defined. If  $\mathbf{X}$  is  $? \times n$ , then  $\mathbf{Y}$  has to be  $n \times ?$  In this case the dimension of  $\mathbf{Z}$  is  $? \times ?$ . The product  $\mathbf{X} \times \mathbf{Y}$  is defined by

$$\mathbf{Z} = \mathbf{X} \times \mathbf{Y} = \sum_{k=1}^n x_{ik} y_{kj} = \{z_{ij}\} \quad (16.1.3)$$

#### Example 16.1.2

$$\mathbf{X} = \begin{bmatrix} 3 & 1 & 2 \\ 4 & 5 & 1 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} 4 & 1 \\ 3 & 2 \\ 1 & 5 \end{bmatrix}$$

$$\mathbf{X} \times \mathbf{Y} = \begin{bmatrix} 3 \times 4 + 1 \times 3 + 2 \times 1 & 3 \times 1 + 1 \times 2 + 2 \times 5 \\ 4 \times 4 + 5 \times 3 + 1 \times 1 & 4 \times 1 + 5 \times 2 + 1 \times 5 \end{bmatrix} = \begin{bmatrix} 17 & 15 \\ 32 & 19 \end{bmatrix}$$

$$\mathbf{Y} \times \mathbf{X} = \begin{bmatrix} 4 \times 3 + 1 \times 4 & 4 \times 1 + 1 \times 5 & 4 \times 2 + 1 \times 1 \\ 3 \times 3 + 2 \times 4 & 3 \times 1 + 2 \times 5 & 3 \times 2 + 2 \times 1 \\ 1 \times 3 + 5 \times 4 & 1 \times 1 + 5 \times 5 & 1 \times 2 + 5 \times 1 \end{bmatrix} = \begin{bmatrix} 16 & 9 & 9 \\ 17 & 13 & 18 \\ 23 & 26 & 7 \end{bmatrix}$$

In general, matrix multiplication is not commutative, that is,  $\mathbf{XY} \neq \mathbf{YX}$  as shown above. As an example, consider the following symmetric matrices:

$$\mathbf{X} = \begin{bmatrix} a & b \\ b & d \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} e & f \\ f & h \end{bmatrix}$$

$$\mathbf{XY} = \begin{bmatrix} ae + bf & af + bh \\ be + df & bf + dh \end{bmatrix}, \quad \mathbf{YX} = \begin{bmatrix} ae + bf & be + df \\ af + bh & bf + dh \end{bmatrix}$$

The matrices  $\mathbf{X}$  and  $\mathbf{Y}$  commute only if  $af + bh = be + df$ .

We usually denote in the product  $\mathbf{XY}$  that  $\mathbf{Y}$  is premultiplied by  $\mathbf{X}$  or  $\mathbf{X}$  is postmultiplied by  $\mathbf{Y}$ . Unlike real numbers, division of one matrix by another matrix is not defined.

### Transpose of Matrix

The transpose of a matrix  $\mathbf{X} = \{x_{ij}\}$  is obtained by interchanging the rows and columns and is denoted by

$$\mathbf{X}^T = \{x_{ij}\}^T = \{x_{ji}\} \quad (16.1.4)$$

The transpose of a matrix product  $(\mathbf{XY})^T$  is given by  $\mathbf{Y}^T \mathbf{X}^T$  whereas  $(\mathbf{X} + \mathbf{Y})^T = \mathbf{X}^T + \mathbf{Y}^T$ . If  $\mathbf{X} = \mathbf{X}^T$ , then we say that the matrix  $\mathbf{X}$  is symmetric.

### Example 16.1.3

$$\mathbf{X} = \begin{bmatrix} 2 & 5 & 4 \\ 1 & 3 & 6 \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} 3 & 2 \\ 1 & 1 \\ 4 & 3 \end{bmatrix}; \quad \mathbf{X}^T = \begin{bmatrix} 2 & 1 \\ 5 & 3 \\ 4 & 6 \end{bmatrix} \quad \mathbf{Y}^T = \begin{bmatrix} 3 & 1 & 4 \\ 2 & 1 & 3 \end{bmatrix}$$

$$(\mathbf{XY})^T = \begin{bmatrix} 27 & 30 \\ 21 & 23 \end{bmatrix}; \quad \mathbf{Y}^T \mathbf{X}^T = \begin{bmatrix} 27 & 30 \\ 21 & 23 \end{bmatrix}$$

### Determinant of a Matrix

The nonzero determinant of a square matrix is one of the two most important *invariants* of a matrix. It is usually denoted by  $\det \mathbf{X}$  or  $|\mathbf{X}|$ . It is computed as

$$\det \mathbf{X} = \sum_{i=1}^n x_{ij} \Delta_{ij} = x_{1j} \Delta_{1j} + x_{2j} \Delta_{2j} + \cdots + x_{nj} \Delta_{nj} \quad (16.1.5)$$

with no implied summation over  $j$  and  $\Delta_{ij}$  is the *cofactor* of the matrix  $\mathbf{X}$  defined by

$$\Delta_{ij} = (-1)^{i+j} M_{ij} \quad (16.1.6)$$

where  $M_{ij}$  is the minor of the matrix  $\mathbf{X}$  obtained by taking the determinant after deleting the  $i$ th row,  $i = 1, \dots, n$ , and the fixed  $j$ th column. Equation (16.1.5) is called a *cofactor expansion* about the  $j$ th column. The difference between the cofactor and the minor is the sign as shown in Eq. (16.1.6).

**Example 16.1.4** The determinant of the following matrix is evaluated as follows:

$$\mathbf{X} = \begin{bmatrix} 2 & 5 & 3 & 2 \\ 1 & 2 & -4 & 5 \\ -4 & 1 & 3 & -2 \\ 3 & -4 & 2 & 1 \end{bmatrix}$$

Expanding in terms of minors about the second column, the expression for the determinant  $|\mathbf{X}|$  is

$$|\mathbf{X}| = -5 \cdot \begin{vmatrix} 1 & -4 & 5 \\ -4 & 3 & -2 \\ 3 & 2 & 1 \end{vmatrix} + 2 \cdot \begin{vmatrix} 2 & 3 & 2 \\ -4 & 3 & -2 \\ 3 & 2 & 1 \end{vmatrix} \\ + 1 \cdot \begin{vmatrix} 2 & 3 & 2 \\ 1 & -4 & 5 \\ 3 & 2 & 1 \end{vmatrix} + -4 \cdot \begin{vmatrix} 2 & 3 & 2 \\ 1 & -4 & 5 \\ -4 & 3 & -2 \end{vmatrix}$$

Each determinants in the above equation can again be expanded in terms of minors about the second column as shown by the expansion of the first term:

$$-5 \cdot \begin{vmatrix} 1 & -4 & 5 \\ -4 & 3 & -2 \\ 3 & 2 & 1 \end{vmatrix} = -5 \times 4 \cdot \begin{vmatrix} -4 & -2 \\ 3 & 1 \end{vmatrix} - 5 \times 3 \cdot \begin{vmatrix} 1 & 5 \\ 3 & 1 \end{vmatrix} \\ - 5 \times (-2) \cdot \begin{vmatrix} 1 & 5 \\ -4 & -2 \end{vmatrix} = 350$$

Evaluating the determinant by this expansion, we have  $|\mathbf{X}| = 350 - 52 - 42 + 376 = 632$ .

If only the determinant of the matrix is zero, we call that it a *singular* or *simply degenerate matrix*. In addition to the determinant, if the first- and higher-order minors are also zero, then we call the matrix *multiply degenerate*.

### Properties of Determinants (X and Y Nonsingular)

1.  $\det \mathbf{X} = \det \mathbf{X}^T$ : The following properties (2,3,4,5,6) are very useful in the manipulation of matrices.
2. If any row (or column) of  $\mathbf{X}$  is multiplied by a constant  $a$  to yield a new matrix  $\mathbf{Y}$ , then  $\det \mathbf{X} = a \cdot \det \mathbf{Y}$ .
3. If any row (or column) of  $\mathbf{X}$  is proportional to the other one, then the rows (or columns) are linearly dependent and  $\det \mathbf{X} = 0$ .
4. If any two rows (or columns) of  $\mathbf{X}$  are interchanged an odd number of times to yield  $\mathbf{Y}$ , then  $\det \mathbf{Y} = -\det \mathbf{X}$ . If they are interchanged an even number of times to yield  $\mathbf{Z}$ , then  $\det \mathbf{Z} = \det \mathbf{X}$ .
5. If any row (or column) of  $\mathbf{X}$  is modified by adding to it  $\alpha$  times the corresponding elements of another row (or column) to yield  $\mathbf{Y}$ , then  $\det \mathbf{Y} = \det \mathbf{X}$ . (The determinant remains unchanged.)
6. The determinant of a triangular matrix  $\mathbf{X}$  is the product of its diagonal terms  $\det \mathbf{X} = x_{11} \cdot x_{22} \cdots \cdot x_{nn}$ .
7. If the matrices  $\mathbf{X}$  and  $\mathbf{Y}$  are both  $n \times n$ , then  $\det (\mathbf{XY}) = (\det \mathbf{X})(\det \mathbf{Y})$ .
8.  $\det (\alpha \mathbf{X}) = \alpha^n \det \mathbf{X}$ .
9. Even if  $\mathbf{X}$  and  $\mathbf{Y}$  are  $n \times n$ ,  $\det (\mathbf{X} + \mathbf{Y}) \neq \det \mathbf{X} + \det \mathbf{Y}$  in general.

**Trace of a Matrix**

The trace of a square matrix  $\text{Tr}(\mathbf{X})$  is the second most important invariant of a matrix. The trace is the sum of the diagonal terms of a matrix. The trace of the matrix given in Example 16.1.4 is  $2 + 2 + 3 + 1 = 8$ .

**Rank of a Matrix**

The column rank of a rectangular matrix is the number of linearly independent column vectors. Similarly, the row rank of a matrix is the number of linearly independent rows. If the column rank and the row rank are the same, then we call that number the *rank* of the matrix. Thus in a nonsingular square matrix the column rank equals the row rank and the rank is the dimension of the matrix. In general, rank of any  $n \times m$  matrix can be defined as the number of elements  $2, 3, \dots, n$  or  $m$  required to form a nonzero determinant. Forming all possible combinations of determinants from any given matrix is a difficult job. Finding a rank is not always easy, but some methods are easier than others. We show one method for finding the rank and the determinant of any matrix using pivotal  $2 \times 2$  determinants.

**Example 16.1.5** Find the rank and determinant of the following nonsingular matrix:

$$\mathbf{X} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & -2 & 3 \\ -1 & 1 & 1 & -2 & 8 \\ 2 & 1 & -2 & 2 & -1 \\ 1 & 2 & 2 & -1 & 0 \\ 2 & -1 & 1 & -2 & 5 \end{bmatrix} \quad (16.1.7)$$

*Step 1.* Remove  $a_{11} = 1$  outside the matrix in Eq. (16.1.7) using property 2 as shown in the previous section. We form a  $4 \times 4$  matrix  $\mathbf{X}_1$  from  $2 \times 2$  determinants of the form  $\{a_{11} \cdot a_{jk} - a_{j1} \cdot a_{1k}, j = 2, \dots, 5; k = 2, \dots, 5\}$  as shown below:

$$\mathbf{X}_1 = 1 \cdot \begin{bmatrix} 3 & 4 & -4 & 11 \\ -3 & -8 & 6 & -7 \\ 0 & -1 & 1 & -3 \\ -5 & -5 & 2 & -1 \end{bmatrix} \quad (16.1.8)$$

*Step 2.* Normalize the first element = 3 as in step 1, and the resultant matrix is shown below.

$$\mathbf{X}_1 = 1.3 \cdot \begin{bmatrix} 1 & \frac{4}{3} & -\frac{4}{3} & \frac{11}{3} \\ -3 & -8 & 6 & -7 \\ 0 & -1 & 1 & -3 \\ -5 & -5 & 2 & -1 \end{bmatrix} \quad (16.1.9)$$

*Step 3.* Form a  $3 \times 3$  matrix  $\mathbf{X}_2$  in an exactly analogous fashion by forming  $2 \times 2$  determinants as in step 1 and normalize the first element in  $\mathbf{X}_2$  as shown in Eq. (16.1.10)

$$\mathbf{X}_2 = 1.3 \cdot \begin{bmatrix} -4 & 2 & 4 \\ -1 & 1 & -3 \\ \frac{5}{3} & \frac{-14}{3} & \frac{52}{3} \end{bmatrix} = 1.3 \cdot -4 \cdot \begin{bmatrix} 1 & \frac{-1}{2} & -1 \\ -1 & \frac{1}{2} & -3 \\ \frac{5}{3} & \frac{-14}{3} & \frac{52}{3} \end{bmatrix} \quad (16.1.10)$$

*Step 4.* Form a  $2 \times 2$  matrix  $\mathbf{X}_3$  in an exactly analogous fashion by forming  $2 \times 2$  determinants as in step 1 and normalize the first element in  $\mathbf{X}_3$  as shown in Eq. (16.1.11):

$$\mathbf{X}_3 = 1.3 - 4 \cdot \begin{bmatrix} \frac{1}{2} & -4 \\ \frac{-23}{6} & 19 \end{bmatrix} = 1.3 - 4 \cdot \frac{1}{2} \begin{bmatrix} 1 & -8 \\ \frac{-23}{6} & 19 \end{bmatrix} \quad (16.1.11)$$

*Step 5.* Continue this process of reducing the order of the matrix until we end up with either a  $1 \times 1$  matrix or a matrix whose elements are all zero. Finally we form a  $1 \times 1$  matrix  $\mathbf{X}_4$  in an exactly analogous fashion by forming  $2 \times 2$  determinants as in step 1:

$$\mathbf{X}_4 = 1.3 \cdot -4 \cdot \frac{1}{2} \cdot \frac{-35}{3} = 70 \quad (16.1.12)$$

Since we have come to a scalar, the operations stop at this point. The number of cycles is 5 and hence the rank of this matrix is 5. The determinant is given by  $\mathbf{X}_4 = 70$ . Thus, we can obtain both the rank and the determinant by this method.

**Example 16.1.6** In this example we will determine the rank and the maximum possible determinant of a singular matrix given by

$$\mathbf{X} = \begin{bmatrix} 6 & 11 & 5 & 4 & 1 \\ 5 & 3 & -2 & -3 & 4 \\ -8 & -12 & -4 & 2 & -5 \\ 6 & 7 & 1 & 0 & 3 \\ 3 & 0 & -3 & -1 & 2 \end{bmatrix}$$

We form the various matrices as in the previous example, and the results are shown below:

$$\mathbf{X} = \begin{bmatrix} 6 & 11 & 5 & 4 & 1 \\ 5 & 3 & -2 & -3 & 4 \\ -8 & -12 & -4 & 2 & -5 \\ 6 & 7 & 1 & 0 & 3 \\ 3 & 0 & -3 & -1 & 2 \end{bmatrix} = 6 \cdot \begin{bmatrix} 1 & \frac{11}{6} & \frac{5}{6} & \frac{4}{6} & \frac{1}{6} \\ 5 & 3 & -2 & -3 & 4 \\ -8 & -12 & -4 & 2 & -5 \\ 6 & 7 & 1 & 0 & 3 \\ 3 & 0 & -3 & -1 & 2 \end{bmatrix}$$

$$\mathbf{X}_1 = 6 \cdot \begin{bmatrix} \frac{-37}{6} & \frac{-37}{6} & \frac{-19}{3} & \frac{19}{6} \\ \frac{8}{3} & \frac{8}{3} & \frac{22}{3} & \frac{-11}{3} \\ -4 & -4 & -4 & 2 \\ \frac{-11}{2} & \frac{-11}{2} & -3 & \frac{3}{2} \end{bmatrix} = 6 \cdot \frac{-37}{6} \begin{bmatrix} 1 & 1 & \frac{38}{37} & \frac{-19}{37} \\ \frac{8}{3} & \frac{8}{3} & \frac{22}{3} & \frac{-11}{3} \\ -4 & -4 & -4 & 2 \\ \frac{-11}{2} & \frac{-11}{2} & -3 & \frac{3}{2} \end{bmatrix}$$

We now form a  $3 \times 3$  matrix using  $2 \times 2$  pivotal determinants. This will yield the first column to be all zeros. Interchanging two columns of the resulting matrix is also shown below:

$$\begin{aligned} \mathbf{X}_2 &= 6 \cdot \frac{-37}{6} \cdot 1 \begin{bmatrix} 0 & \frac{170}{37} & \frac{-85}{37} \\ 0 & \frac{4}{37} & \frac{-2}{37} \\ 0 & \frac{98}{37} & \frac{-49}{37} \end{bmatrix} = 6 \cdot \frac{-37}{6} \cdot 1 \begin{bmatrix} \frac{170}{37} & \frac{-85}{37} & 0 \\ \frac{4}{37} & \frac{-2}{37} & 0 \\ \frac{98}{37} & \frac{-49}{37} & 0 \end{bmatrix} \\ &= 6 \cdot \frac{-37}{6} \cdot \frac{170}{37} \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ \frac{4}{37} & \frac{-2}{37} & 0 \\ \frac{98}{37} & \frac{-49}{37} & 0 \end{bmatrix} \end{aligned}$$

We now form a  $2 \times 2$  matrix

$$\mathbf{X}_3 = 6 \cdot \frac{-37}{6} \cdot \frac{170}{37} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = -170 \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

After three cycles we have a premature termination with a  $2 \times 2$  zero matrix. Hence the rank of this matrix is 3 and the maximum determinant is  $6 \cdot \frac{-37}{6} \cdot \frac{170}{37} = -170$ .

### **Inverse of a Matrix**

Since division by a matrix is not defined, we define the inverse of a nonsingular matrix as the matrix  $\mathbf{X}^{-1}$  such that  $\mathbf{X}\mathbf{X}^{-1} = \mathbf{X}^{-1}\mathbf{X} = \mathbf{I}$ , an identity matrix. Thus inverse matrices commute with their direct counterparts. The cofactor expansion of the determinant of the matrix  $\mathbf{X}$  from Eq. (16.1.5) is

$$\det \mathbf{X} = \sum_{i=1}^n x_{ij} \Delta_{ij} = x_{1j} \Delta_{1j} + x_{2j} \Delta_{2j} + \cdots + x_{nj} \Delta_{nj} \quad (16.1.5)$$

where the expansion is carried about the  $n$  elements in any column  $j$ . Equation (16.1.5) can be rewritten as

$$\det \mathbf{X} = \sum_{i=1}^n x_{ij} \Delta_{ik} = \begin{cases} \det X, & k=j \\ 0, & k \neq j \end{cases} \quad (16.1.13)$$

In Eq. (16.1.13), when  $k = j$ , we obtain  $\det \mathbf{X}$  by Eq. (16.1.5), but when  $k \neq j$ , then we have cofactor expansion about the  $n$  elements in the  $j$ th column but with cofactors obtained from the  $k$ th column. This is equivalent to finding the determinant by replacing the  $k$ th column with the  $j$ th column in the matrix  $\mathbf{X}$ . Thus  $\mathbf{X}$  will contain two identical columns and hence from property 3 in the previous section,  $\det \mathbf{X} = 0$ .

We divide Eq. (16.1.13) by  $\det \mathbf{X}$  and obtain

$$\sum_{i=1}^n \left( \frac{\Delta_{ik}}{\det \mathbf{X}} \right) x_{ij} = \delta_{kj}, \quad \begin{cases} 1 \leq k \leq n \\ 1 \leq j \leq n \end{cases} \quad (16.1.14)$$

where  $\delta_{kj}$  is the Krönecker delta, which equals 1 if  $k = j$  and 0 if  $k \neq j$ . Equation (16.1.14) is a scalar statement of the matrix equation  $\mathbf{X}^{-1}\mathbf{X} = \mathbf{I}$ , and the inverse is given by

$$\mathbf{X}^{-1} = \{\alpha_{ki}\} = \left\{ \frac{\Delta_{ik}}{\det \mathbf{X}} \right\} x_{ij} = \frac{1}{\det \mathbf{X}} \begin{bmatrix} \Delta_{11} & \Delta_{21} & \cdots & \Delta_{n1} \\ \Delta_{12} & \Delta_{22} & \cdots & \Delta_{n2} \\ \vdots & \vdots & & \vdots \\ \Delta_{1n} & \Delta_{2n} & \cdots & \Delta_{nn} \end{bmatrix} \quad (16.1.15)$$

Note that the cofactors are in reverse order of the sequence of elements of the matrix  $\mathbf{X}$ .

The matrix  $\{\Delta_{ji}\}$  of the cofactors is sometimes called the *adjoint matrix*. Finding the inverse using Eq. (16.1.15) is very inefficient and is of only theoretical interest since it gives an explicit expression for the inverse.

### Solution of Linear Equations

We can use matrix analysis to solve linear equations of the form  $\mathbf{Ax} = \mathbf{y}$  where  $\mathbf{x}$  and  $\mathbf{y}$  are  $n$ -vectors:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad (16.1.16)$$

We will assume that  $\mathbf{A}$  is a nonsingular  $n \times n$  matrix and  $\mathbf{x}$  and  $\mathbf{y}$  are  $n \times 1$  vectors. We want to solve for  $\mathbf{x}$  in terms of  $\mathbf{y}$ . Since  $\mathbf{A}$  is invertible, we can write the solution as

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} \quad (16.1.17)$$

This equation can be expanded as

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} \Delta_{11} & \Delta_{21} & \cdots & \Delta_{n1} \\ \Delta_{12} & \Delta_{22} & \cdots & \Delta_{n2} \\ \vdots & \vdots & & \vdots \\ \Delta_{1n} & \Delta_{2n} & \cdots & \Delta_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \frac{1}{\det \mathbf{A}} \sum_{j=1}^n \begin{bmatrix} \Delta_{j1} \\ \Delta_{j2} \\ \vdots \\ \Delta_{jn} \end{bmatrix} y_j \quad (16.1.18)$$

Note again that the order of the cofactors is transposed from that of the original matrix  $\mathbf{A}$ , that is, if  $a_{ij}$  is an element in the  $i$ th row and  $j$ th column, then  $\Delta_{ji}$  is in the  $j$ th row and  $i$ th column.

**Example 16.1.7** We shall now solve the following set of linear equations for  $x_1$ ,  $x_2$ ,  $x_3$ ,  $x_4$ :

$$\begin{aligned} x_1 - 3x_3 - 2x_4 &= -4 \\ -\frac{5}{2}x_1 + \frac{1}{2}x_2 + \frac{15}{2}x_3 + 6x_4 &= \frac{23}{2} \\ 2x_1 - 7x_3 - 6x_4 &= -11 \\ -\frac{7}{2}x_1 + \frac{1}{2}x_2 + \frac{25}{2}x_3 + 10x_4 &= \frac{39}{2} \end{aligned}$$

or in matrix form

$$\begin{bmatrix} 1 & 0 & -3 & -2 \\ -5 & 1 & 15 & 6 \\ \frac{-5}{2} & \frac{1}{2} & \frac{15}{2} & 6 \\ 2 & 0 & -7 & -6 \\ \frac{-7}{2} & \frac{1}{2} & \frac{25}{2} & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -4 \\ \frac{23}{2} \\ \frac{2}{2} \\ -11 \\ \frac{39}{2} \end{bmatrix}$$

The determinant of the matrix  $\mathbf{A}$  is found as follows:

$$|\mathbf{A}| = \begin{bmatrix} 1 & 0 & -3 & -2 \\ -5 & 1 & 15 & 6 \\ \frac{-5}{2} & \frac{1}{2} & \frac{15}{2} & 6 \\ 2 & 0 & -7 & -6 \\ \frac{-7}{2} & \frac{1}{2} & \frac{25}{2} & 10 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 & 15 & 6 \\ 0 & -7 & -6 \\ \frac{1}{2} & \frac{25}{2} & 10 \end{bmatrix} - 3 \cdot \begin{bmatrix} -5 & 1 & 6 \\ 2 & 0 & -6 \\ \frac{-7}{2} & \frac{1}{2} & 10 \end{bmatrix} \\ + 2 \cdot \begin{bmatrix} -5 & 1 & 15 \\ 2 & 0 & -7 \\ \frac{-7}{2} & \frac{1}{2} & \frac{25}{2} \end{bmatrix}$$

Or  $|\mathbf{A}| = 1 + 3 - 3 = 1$ . The inverse of matrix  $\mathbf{A}$  is

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & 0 & -3 & -2 \\ -5 & 1 & 15 & 6 \\ \frac{-5}{2} & \frac{1}{2} & \frac{15}{2} & 6 \\ 2 & 0 & -7 & -6 \\ \frac{-7}{2} & \frac{1}{2} & \frac{25}{2} & 10 \end{bmatrix}^{-1} = \frac{1}{1} \begin{bmatrix} 1 & -2 & 1 & 2 \\ 2 & 1 & 2 & 1 \\ -1 & -1 & 1 & 1 \\ \frac{3}{2} & \frac{1}{2} & -1 & \frac{-1}{2} \end{bmatrix}$$

The solution to the original set of equations is given by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 1 & 2 \\ 2 & 1 & 2 & 1 \\ -1 & -1 & 1 & 1 \\ \frac{3}{2} & \frac{1}{2} & -1 & \frac{-1}{2} \end{bmatrix} \begin{bmatrix} -4 \\ \frac{23}{2} \\ \frac{2}{2} \\ -11 \\ \frac{39}{2} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

### Properties of Inverses (X, Y, Z Same Order and Nonsingular)

1.  $\mathbf{X} \mathbf{X}^{-1} = \mathbf{X}^{-1} \mathbf{X}$  matrix commutes with its inverse
2.  $\mathbf{X}^m \mathbf{X}^n = \mathbf{X}^{m+n}$
3.  $(\mathbf{XY})^{-1} = \mathbf{Y}^{-1} \mathbf{X}^{-1}$
4.  $\mathbf{XY} = \mathbf{XZ} \Rightarrow \mathbf{Y} = \mathbf{Z}$
5.  $(\mathbf{X}^T)^{-1} = (\mathbf{X}^{-1})^T$

6.  $\det \mathbf{X}^{-1} = \frac{1}{\det \mathbf{X}}$
7.  $\mathbf{X}^{-1} = \frac{(-1)^{i+j} M_{ij}}{\det \mathbf{X}}$ , where  $M_{ij}$  is the minor of  $\mathbf{X}$  (16.1.19)

### Orthogonal Matrices

The matrix  $\mathbf{X}$  is orthonormal if  $\mathbf{X}^T \mathbf{X} = \mathbf{I}$ . Since  $\mathbf{X}^{-1} \mathbf{X}$  is also equal to the identity matrix  $\mathbf{I}$ , we have

$$\mathbf{X}^T \mathbf{X} = \mathbf{I} = \mathbf{X}^{-1} \mathbf{X} \quad \text{or} \quad \mathbf{X}^T = \mathbf{X}^{-1} \quad (16.1.20)$$

It immediately follows from Eq. (16.1.20) that the determinant of an orthogonal matrix is equal to  $\pm 1$ . Note that the converse is not true. In any orthogonal matrix  $\mathbf{X}$ , real or complex, individual row or column vectors  $\mathbf{x}$  are orthonormal:

$$\mathbf{x}_i^H \mathbf{x}_j = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases} \quad (16.1.21)$$

where  $\mathbf{H}$  stands for *conjugate transpose*.

An example of a  $4 \times 4$  orthogonal matrix used in discrete Fourier transform analysis is shown below:

$$\mathbf{W} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & e^{-j(2\pi/4)} & e^{-j(4\pi/4)} & e^{-j(6\pi/4)} \\ 1 & e^{-j(4\pi/4)} & e^{-j(8\pi/4)} & e^{-j(12\pi/4)} \\ 1 & e^{-j(6\pi/4)} & e^{-j(12\pi/4)} & e^{-j(18\pi/4)} \end{bmatrix} \quad (16.1.22)$$

It can be verified that the rows and columns of  $\mathbf{W}$  are orthonormal and  $\det \mathbf{W} = 1$ .

## 16.2 EIGENVALUES AND EIGENVECTORS OF MATRICES

### Eigenvalues

The concept of eigenvalues and eigenvectors plays an important role. We motivate this concept by investigating solutions to a linear system of equations given by

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{I}\mathbf{x} = \lambda \mathbf{x}$$

where  $\mathbf{A}$  is an  $n \times n$  matrix and  $\mathbf{x}$  is an  $n \times 1$  vector,  $\lambda$  is a scalar constant, and  $\mathbf{I}$  is an  $n \times n$  identity matrix. Simplification of the equation above yields

$$(\lambda \mathbf{I} - \mathbf{A})\mathbf{x} = 0 \quad (16.2.1)$$

A trivial solution to Eq. (16.2.1) is  $\mathbf{x} = 0$ . Nontrivial solutions can exist only if

$$\det(\lambda \mathbf{I} - \mathbf{A}) = |\lambda \mathbf{I} - \mathbf{A}| = 0 \quad (16.2.2)$$

Equation (16.2.2) is called the *characteristic equation* of the matrix  $\mathbf{A}$ . Expansion of the determinant in Eq. (16.2.2) yields a polynomial in  $\lambda$  of degree  $n$  called the *characteristic polynomial* of the matrix  $\mathbf{A}$ , given by

$$\lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_0 = 0 \quad (16.2.3)$$

The characteristic equation has a fundamental significance in linear systems. The  $n$  roots of the characteristic equation are called the *eigenvalues* of the matrix  $\mathbf{A}$ , and in linear systems they are also known as the “poles” of the transfer function. The eigenvalues can be real, repeated, or complex. From the well-known result for the roots of polynomials, we also have the sum of the eigenvalues as  $-a_{n-1}$ , and the product of the eigenvalues, which is also the determinant of the matrix  $\mathbf{A}$ , is given by  $(-1)^n a_0$ .

**Example 16.2.1** The eigenvalues of the following nonsingular matrix have to be determined:

$$\mathbf{A} = \begin{bmatrix} -1.5625 & -4.5 & 2.0625 & 5.25 & -0.75 \\ 0.4375 & -2.5 & -1.9375 & 6.25 & 1.25 \\ 0.4375 & -2.5 & -1.9375 & 7.25 & 1.25 \\ -0.0625 & -3.5 & -0.4375 & 6.25 & 0.25 \\ 0.5 & 0 & -1.5 & 3 & 1 \end{bmatrix}$$

The characteristic equation of  $\mathbf{A}$  is  $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$  is given by

$$|\lambda\mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda + 1.5625 & 4.5 & -2.0625 & -5.25 & 0.75 \\ -0.4375 & \lambda + 2.5 & 1.9375 & -6.25 & -1.25 \\ -0.4375 & 2.5 & \lambda + 1.9375 & -7.25 & -1.25 \\ 0.0625 & 3.5 & 0.4375 & \lambda - 6.25 & -0.25 \\ -0.5 & 0 & 1.5 & -3 & \lambda - 1 \end{vmatrix} \\ = \lambda^5 - 1.25\lambda^4 - 2.375\lambda^3 + \lambda^2 + 1.375\lambda + 0.25 = 0$$

The solutions of the fifth-order polynomial of this equation are the eigenvalues of  $\mathbf{A}$ , given by

$$\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = -1, \lambda_4 = -0.5, \lambda_5 = -0.25$$

The product of the eigenvalues is the determinant given by  $(-1)^5 \cdot (a_0) = -0.25$  and the sum of the eigenvalues is  $a_4 = -1.25$ .

We can also find the eigenvalues of a square matrix that is singular. In this case the number of nonzero eigenvalues represents the rank of the matrix and the number of zero eigenvalues, the degree of singularity.

### Cayley–Hamilton Theorem

The Cayley–Hamilton theorem states that every square matrix satisfies its own characteristic equation. In Eq. (16.2.3), if we substitute the matrix  $\mathbf{A}$  for  $\lambda$ , we obtain the Cayley–Hamilton equation

$$\mathbf{A}^n + a_{n-1}\mathbf{A}^{n-1} + a_{n-2}\mathbf{A}^{n-2} + \cdots + a_0\mathbf{A}^0 = \mathbf{0} \quad (16.2.4)$$

where  $\mathbf{0}$  represents a zero matrix.

**Example 16.2.2** Given the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 6 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}$$

the characteristic equation is  $\lambda^3 + 6\lambda^2 + 11\lambda + 6 = 0$ . According to the Cayley–Hamilton theorem, we should have

$$\begin{aligned} & \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}^3 + 6 \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}^2 + 11 \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}^1 \\ & + 6 \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}^0 = \begin{bmatrix} -6 & -11 & -6 \\ 36 & 60 & 25 \\ -150 & -239 & -90 \end{bmatrix} + \begin{bmatrix} 0 & 0 & -6 \\ -36 & -66 & -36 \\ 216 & 360 & 150 \end{bmatrix} \\ & + \begin{bmatrix} 0 & 11 & 0 \\ 0 & 0 & 11 \\ -66 & -121 & -66 \end{bmatrix} + \begin{bmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{bmatrix} \end{aligned}$$

or

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

showing the validity of the theorem.

### Definiteness of Matrices

A matrix  $\mathbf{A}$  is *positive definite* if all its eigenvalues are either positive or have positive real parts. For example, the matrix  $\mathbf{A}$  in Example 16.2.2 is positive definite, whereas the matrix  $\mathbf{A}$  in Example 16.2.1 is nondefinite. Similarly, matrices having eigenvalues that are negative or have negative real parts are known as *negative definite*. If some of the eigenvalues are zero, then the matrix is called *positive semidefinite* or *negative semidefinite* accordingly.

Since the determinants are  $(-1)^n$  times the product of eigenvalues, definiteness can also be defined by the sign of the determinants of all principal minors of a matrix. *Principal minors* are those minors obtained by successively deleting the rows and columns of a matrix starting from the first row and first column. If all the determinants of principal minors are positive, then the matrix is positive-definite: if they are negative then they are negative definite.

**Example 16.2.3** We will check the definiteness of the following matrix by finding the eigenvalues:

$$\mathbf{A} = \begin{bmatrix} 1 & 7 & -7 & -5 & 6 \\ 3 & 9 & -11 & -4 & 8 \\ 3 & 9 & -11 & -3 & 8 \\ 2.5 & 8 & -9.5 & -4 & 7 \\ 0.5 & 0 & -1.5 & 3 & 1 \end{bmatrix}$$

As in the previous example, setting  $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$  gives

$$|\lambda\mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda - 1 & -7 & 7 & 5 & -6 \\ -3 & \lambda - 9 & 11 & 4 & -8 \\ -3 & -9 & \lambda + 11 & 3 & -8 \\ -2.5 & -8 & 9.5 & \lambda + 4 & -7 \\ -0.5 & 0 & 1.5 & -3 & \lambda - 1 \end{vmatrix} = \lambda^2(\lambda^3 + 4\lambda^2 + \lambda - 6) = 0$$

Thus the rank of the matrix  $\mathbf{A}$  is 3 since there are two zero eigenvalues. The eigenvalues are given by  $\lambda_1 = -3, \lambda_2 = -2, \lambda_3 = 1, \lambda_4 = 0, \lambda_5 = 0$ . Here the matrix  $\mathbf{A}$  is nondefinite since it has positive, negative, and zero eigenvalues.

### Eigenvectors

*Eigenvectors* are solutions to Eq. (16.2.1) for any given eigenvalue. We shall assume that all eigenvalues are distinct and there are no repeated eigenvalues. If we substitute any eigenvalue  $\lambda_i$  in  $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$ , then, by the very definition, we have  $\det(\lambda_i\mathbf{I} - \mathbf{A}) = 0$  and hence solutions to Eq. (16.2.1) are possible, and these solutions are the unnormalized eigenvectors  $\psi_i$ . However, these solutions are unique only upto a multiplicative constant. To make the solutions unique we can normalize these eigenvectors and the normalized eigenvectors are denoted by  $\phi_i = \psi_i / |\psi_i|$ .

**Example 16.2.4** Find the eigenvectors of the nonsingular matrix of Example 16.2.1, repeated below:

$$\mathbf{A} = \begin{bmatrix} -1.5625 & -4.5 & 2.0625 & 5.25 & -0.75 \\ 0.4375 & -2.5 & -1.9375 & 6.25 & 1.25 \\ 0.4375 & -2.5 & -1.9375 & 7.25 & 1.25 \\ -0.0625 & -3.5 & -0.4375 & 6.25 & 0.25 \\ 0.5 & 0 & -1.5 & 3 & 1 \end{bmatrix}$$

Substituting  $\lambda_1 = 2$  in this equation, we can solve for the unnormalized eigenvectors:

$$(\lambda_1\mathbf{I} - \mathbf{A})\Psi_1 = \begin{bmatrix} 3.5625 & 4.5 & -2.0625 & -5.25 & 0.75 \\ -0.4375 & 4.5 & 1.9375 & -6.25 & -1.25 \\ -0.4375 & 2.5 & 3.9375 & -7.25 & -1.25 \\ 0.0625 & 3.5 & 0.4375 & -4.25 & -0.25 \\ -0.5 & 0 & 1.5 & -3 & 1 \end{bmatrix} \begin{bmatrix} \psi_{11} \\ \psi_{12} \\ \psi_{13} \\ \psi_{14} \\ \psi_{15} \end{bmatrix} = 0$$

As mentioned earlier, the equation above can be solved for only four of the unknowns in terms of the fifth. Let us assume that  $\psi_{11} = 1$  and solve for  $\psi_{12}, \psi_{13}, \psi_{14}, \psi_{15}$  in terms of  $\psi_{11}$ . The preceding equation can be rewritten by substituting  $\psi_{11} = 1$ :

$$\begin{bmatrix} 4.5 & 1.9375 & -6.25 & -1.25 \\ 2.5 & 3.9375 & -7.25 & -1.25 \\ 3.5 & 0.4375 & -4.25 & -0.25 \\ 0 & 1.5 & -3 & 1 \end{bmatrix} \begin{bmatrix} \psi_{12} \\ \psi_{13} \\ \psi_{14} \\ \psi_{15} \end{bmatrix} = \begin{bmatrix} 0.4375 \\ 0.4375 \\ -0.0625 \\ 0.5 \end{bmatrix}$$

Solving for  $\psi_{12}, \psi_{13}, \psi_{14}, \psi_{15}$  and adjoining  $\psi_{11} = 1$ , we obtain the unnormalized eigenvector corresponding to the eigenvalue  $\lambda_1 = 2$ :

$$\begin{bmatrix} \psi_{12} \\ \psi_{13} \\ \psi_{14} \\ \psi_{15} \end{bmatrix} = \begin{bmatrix} 1.333333 \\ 1.952381 \\ 1.238095 \\ 1.285714 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \psi_{11} \\ \psi_{12} \\ \psi_{13} \\ \psi_{14} \\ \psi_{15} \end{bmatrix} = \begin{bmatrix} 1 \\ 1.333333 \\ 1.952381 \\ 1.238095 \\ 1.285714 \end{bmatrix}$$

Normalization of this equation yields the eigenvector  $\Phi_1$ :

$$\Phi_1 = \frac{\Psi_1}{|\Psi_1|} = \frac{1}{3.126581} \begin{bmatrix} 1 \\ 1.333333 \\ 1.952381 \\ 1.238095 \\ 1.285714 \end{bmatrix} = \begin{bmatrix} 0.319838 \\ 0.426451 \\ 0.624446 \\ 0.395990 \\ 0.411220 \end{bmatrix}$$

$\lambda_1=2$

In a similar manner, the other eigenvectors corresponding to  $\lambda_2 = 1, \lambda_3 = -1, \lambda_4 = -0.5$ , and  $\lambda_5 = -0.25$  are shown below:

$$\Phi_2 = \begin{bmatrix} 0.275010 \\ 0.366679 \\ 0.641681 \\ 0.275010 \\ 0.550019 \end{bmatrix}, \quad \Phi_3 = \begin{bmatrix} -0.774597 \\ 0.516398 \\ 0.258199 \\ 0.258199 \\ 0 \end{bmatrix}, \quad \Phi_4 = \begin{bmatrix} 0.679189 \\ -0.549819 \\ 0.032342 \\ -0.291081 \\ 0.388108 \end{bmatrix},$$

$\lambda_2=1 \qquad \qquad \qquad \lambda_3=-1 \qquad \qquad \qquad \lambda_4=-0.5$

$$\Phi_5 = \begin{bmatrix} -0.679104 \\ 0.325871 \\ -0.340796 \\ 0.166667 \\ -0.531373 \end{bmatrix}$$

$\lambda_5=-0.25$

### Modal Matrix

A matrix  $\Phi$  consisting of the linearly independent eigenvectors of a matrix  $\mathbf{A}$  is called the *modal matrix* of  $\mathbf{A}$ . In the previous example the modal matrix of  $\mathbf{A}$  was

$$\Phi = \begin{bmatrix} 0.319838 & 0.319838 & -0.774597 & 0.679189 & -0.679104 \\ 0.426451 & 0.426451 & 0.516398 & -0.549819 & 0.325871 \\ 0.624446 & 0.624446 & 0.258199 & 0.032342 & -0.340796 \\ 0.395990 & 0.395990 & 0.258199 & -0.291081 & 0.166667 \\ 0.411220 & 0.411220 & 0 & 0.388108 & -0.531373 \end{bmatrix}$$

The determinant of  $\Phi = -0.00141$ .

### Diagonalization of Matrices

An  $n \times n$  square matrix  $\mathbf{A}$  is *diagonalizable* if there exists another nonsingular  $n \times n$  matrix  $\mathbf{T}$  such that  $\mathbf{T}^{-1}\mathbf{AT} = \mathbf{D}$ , where  $\mathbf{D}$  is a diagonal matrix. The conditions of diagonalizability reduce to the existence of  $n$  linearly independent eigenvectors of  $\mathbf{A}$  or the modal matrix  $\Phi$  of  $\mathbf{A}$  is invertible, in which case the  $j$ th diagonal element of  $\mathbf{D}$  is the  $j$ th eigenvalue of  $\mathbf{A}$ . In this case the modal matrix  $\Phi$  is the desired matrix  $\mathbf{T}$ . The modal matrix  $\Phi$  is not the only matrix that will diagonalize  $\mathbf{A}$ . There are other matrices that may also diagonalize  $\mathbf{A}$ , but all the diagonal elements obtained from the modal matrix only are the eigenvalues of  $\mathbf{A}$ .

As a consequence of the preceding discussion, we can write the matrix equivalent of the eigenequation (16.2.1) as follows:

$$\mathbf{A}\Phi = \Phi\Lambda \quad \text{or} \quad \Phi^{-1}\mathbf{A}\Phi = \Lambda \quad (16.2.5)$$

**Example 16.2.5** Diagonalize the following matrix:

$$\mathbf{A} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ -4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 \\ 5 & 1 & 1 & 0 \end{bmatrix}$$

The eigenvalues of  $\mathbf{A}$  are obtained by solving the determinant equation  $|\lambda\mathbf{I} - \mathbf{A}| = 0$ , or

$$|\lambda\mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda & 1 & 0 & 0 \\ 4 & \lambda & 0 & 0 \\ 0 & -2 & \lambda & -1 \\ -5 & -1 & -1 & \lambda \end{vmatrix} = \lambda^4 - 5\lambda^2 + 4 = 0$$

The characteristic equation can be factored as  $(\lambda - 2)(\lambda + 2)(\lambda - 1)(\lambda + 1) = 0$ , yielding the eigenvalues as  $\lambda_1 = 2$ ,  $\lambda_2 = -2$ ,  $\lambda_3 = 1$ ,  $\lambda_4 = -1$ . We find the eigenvectors corresponding to these eigenvalues. The eigenvectors following the Example 16.2.4 are

$$\Phi_1 = \begin{bmatrix} \frac{-3}{\sqrt{74}} \\ \frac{6}{\sqrt{74}} \\ \frac{5}{\sqrt{74}} \\ \frac{-2}{\sqrt{74}} \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} \frac{-3}{\sqrt{146}} \\ \frac{-6}{\sqrt{146}} \\ \frac{1}{\sqrt{146}} \\ \frac{10}{\sqrt{146}} \end{bmatrix}, \quad \Phi_3 = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \Phi_4 = \begin{bmatrix} 0 \\ 0 \\ \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

or

$$\Phi_1 = \begin{bmatrix} -0.348743 \\ 0.697486 \\ 0.581238 \\ -0.232495 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} -0.248282 \\ -0.496564 \\ 0.082761 \\ 0.827606 \end{bmatrix}, \quad \Phi_3 = \begin{bmatrix} 0 \\ 0 \\ 0.707107 \\ 0.707107 \end{bmatrix},$$

$\lambda_1=2 \qquad \qquad \lambda_2=-2 \qquad \qquad \lambda_3=1$

$$\Phi_4 = \begin{bmatrix} 0 \\ 0 \\ -0.707107 \\ 0.707107 \end{bmatrix}$$

$\lambda_4=-1$

The modal matrix  $\Phi$  is given by

$$\Phi = [\Phi_1 \ \Phi_2 \ \Phi_3 \ \Phi_4] = \begin{bmatrix} \frac{-3}{\sqrt{74}} & \frac{-3}{\sqrt{146}} & 0 & 0 \\ \frac{6}{\sqrt{74}} & \frac{-6}{\sqrt{146}} & 0 & 0 \\ \frac{5}{\sqrt{74}} & \frac{1}{\sqrt{146}} & 1 & -1 \\ \frac{-2}{\sqrt{74}} & \frac{10}{\sqrt{146}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{\sqrt{74}}{\sqrt{146}} & \frac{\sqrt{146}}{\sqrt{2}} & \frac{\sqrt{2}}{\sqrt{2}} & \frac{\sqrt{2}}{\sqrt{2}} \end{bmatrix}$$

and

$$\Phi^{-1}A\Phi = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

### Symmetric Matrices

Real symmetric matrices are those matrices that satisfy the condition

$$A^T = A \quad (16.2.6)$$

Complex symmetric matrices satisfy the Hermitian symmetry condition

$$A^H = (A^*)^T = A \quad (16.2.7)$$

where the symbol  $H$  defined earlier stands for the *conjugate transpose*.

Symmetric matrices play a very important role in probability; for example, the covariance matrices are all real symmetric matrices. They have some important properties that come in handy in simplifying algebraic manipulations. Some of the more important properties of symmetric matrices are

1. All eigenvalues are real.
2. The eigenvectors corresponding to distinct eigenvalues are orthogonal. Hence the eigenvectors can form the basis vectors of an  $n$ -dimensional coordinate system.
3. The modal matrix  $\Phi$  is an orthogonal matrix, and hence  $\Phi^{-1} = \Phi^T$ . The diagonalization of a symmetric matrix can be accomplished by the transformation  $\Phi^T A \Phi = \Lambda$ .

**Example 16.2.6** The eigenvectors and eigenvalues of the symmetric matrix given below are to be determined and their properties studied:

$$A = \begin{bmatrix} 1 & -2 & 1 & -1 \\ -2 & 1 & 1 & -1 \\ 1 & 1 & 2 & -2 \\ -1 & -1 & -2 & 0 \end{bmatrix}$$

The characteristic equation is  $\det[\lambda I - A] = 0$ :

$$|\lambda I - A| = \begin{vmatrix} \lambda - 1 & 2 & -1 & 1 \\ 2 & \lambda - 1 & -1 & 1 \\ -1 & -1 & \lambda - 2 & 2 \\ 1 & 1 & 2 & \lambda \end{vmatrix} = \lambda^4 - 4\lambda^3 - 7\lambda^2 + 22\lambda + 24 = 0$$

The eigenvalues are  $\lambda_1 = 3$ ,  $\lambda_2 = -2$ ,  $\lambda_3 = -1$ ,  $\lambda_4 = 4$ , obtained by factoring the characteristic equation. The four normalized eigenvectors are obtained as outlined in the previous examples and they are given by

$$\Phi_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \Phi_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \quad \Phi_3 = \frac{1}{\sqrt{10}} \begin{bmatrix} -1 \\ -1 \\ 2 \\ 2 \end{bmatrix}, \quad \Phi_4 = \frac{1}{\sqrt{15}} \begin{bmatrix} 1 \\ 1 \\ 3 \\ -2 \end{bmatrix}$$

$\lambda_1=3 \quad \lambda_2=-2 \quad \lambda_3=-1 \quad \lambda_4=4$

The modal matrix  $\Phi$  is

$$\Phi = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{3}} \frac{1}{\sqrt{10}} \frac{1}{\sqrt{15}} \begin{bmatrix} -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 0 & 0 & 2 & 3 \\ 0 & 1 & 2 & -2 \end{bmatrix}, \quad |\Phi| = \frac{30}{30} = 1$$

It can be verified that

$$\Phi_i \Phi_j = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases}$$

showing that the modal matrix is orthogonal and  $|\Phi| = 1$ .

The original matrix  $\mathbf{A}$  can be diagonalized by the modal matrix as shown below:

$$\begin{aligned} \Phi^T \mathbf{A} \Phi &= \frac{1}{30} \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ -1 & -1 & 2 & 2 \\ 1 & 1 & -3 & -2 \end{bmatrix} \begin{bmatrix} 1 & -2 & 1 & -1 \\ -2 & 1 & 1 & -1 \\ 1 & 1 & 2 & -2 \\ -1 & -1 & -2 & 0 \end{bmatrix} \\ &\times \begin{bmatrix} -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 0 & 0 & 2 & 3 \\ 0 & 1 & 2 & -2 \end{bmatrix} \frac{1}{30} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} \end{aligned}$$

### Similarity Transformations

In the previous section we obtained the diagonal matrix  $\Lambda$  by transforming the matrix  $\mathbf{A}$  by the modal matrix  $\Phi$ . The resulting diagonal matrix  $\Lambda$  has all the fundamental properties of the original matrix  $\mathbf{A}$ . For example these two matrices have the same determinants and the same traces. Such matrices are called *similar matrices*, and the transformation is known as a *similarity transformation*. Any nonsingular matrix  $\mathbf{A}$  can be transformed into another similar matrix  $\mathbf{B}$  by the following similarity transformation

$$\mathbf{T}^{-1} \mathbf{A} \mathbf{T} = \mathbf{B}$$

where  $\mathbf{T}$  is another nonsingular transforming matrix.

**Example 16.2.7** With

$$\mathbf{T} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 1 & 0 & 1 & 1 \\ -1 & 1 & -1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} -1 & -2 & 1 & -2 \\ 2 & -1 & -2 & 1 \\ -1 & 1 & -1 & 0 \\ 2 & 1 & -1 & 2 \end{bmatrix}$$

we have

$$\mathbf{T}^{-1}\mathbf{AT} = \mathbf{B} = \begin{bmatrix} -7 & 0 & -7 & -5 \\ 2 & 2 & 2 & 3 \\ 7 & 4 & 6 & 7 \\ 2 & -3 & 1 & -2 \end{bmatrix}$$

In this example the determinants of  $\mathbf{A}$  and  $\mathbf{B}$  are 1 and the traces of  $\mathbf{A}$  and  $\mathbf{B}$  are  $-1$ .

## 16.3 VECTOR AND MATRIX DIFFERENTIATION

The definitions for differentiation with respect to a real vector and a complex vector are slightly different. In the case of differentiation of a complex vector, the question of whether the derivative is a total or partial also makes a difference. We shall first discuss differentiation with respect to real vectors.

### ***Differentiation with Respect to Real Vectors***

Let  $\mathbf{x}$  be an  $n$ -vector variable defined by  $\mathbf{x} = \{x_1, x_2, \dots, x_n\}^T$ . The total and partial derivative operators with respect to  $\mathbf{x}$  are defined by

$$\frac{d}{d\mathbf{x}} = \begin{bmatrix} \frac{d}{dx_1} \\ \frac{d}{dx_2} \\ \vdots \\ \frac{d}{dx_n} \end{bmatrix}, \quad \frac{\partial}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} \quad (16.3.1)$$

Just as the  $(d/dx)x = 1$ , we want the derivative with respect to the vector  $\mathbf{x}$  to be equal to the identity matrix  $\mathbf{I}$ , taking care that the dimensional compatibility conditions are satisfied. Thus,  $(d/d\mathbf{x})\mathbf{x}$  cannot be defined since the compatibility conditions are not met. We can, however, define  $(d/d\mathbf{x})\mathbf{x}^T = \mathbf{I}$  as shown below:

$$\frac{d}{d\mathbf{x}} \mathbf{x}^T = \begin{bmatrix} \frac{d}{dx_1} \\ \frac{d}{dx_2} \\ \vdots \\ \frac{d}{dx_n} \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \mathbf{I} \quad (16.3.2)$$

In an analogous fashion we can define the following operation:

$$\frac{d}{d\mathbf{x}} \mathbf{x}^T \mathbf{x} = \begin{bmatrix} \frac{d}{dx_1} \\ \frac{d}{dx_2} \\ \vdots \\ \frac{d}{dx_n} \end{bmatrix} \begin{bmatrix} x_1^2 + x_2^2 + \cdots + x_n^2 \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \\ 2x_n \end{bmatrix} = 2\mathbf{x} \quad (16.3.3)$$

We can also write the following differentiation formulas as follows

$$\begin{aligned}\frac{d}{d\mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x} &= (\mathbf{A} + \mathbf{A}^T) \mathbf{x} \\ \frac{d}{d\mathbf{x}} \mathbf{y}^T \mathbf{A} \mathbf{x} &= \frac{d}{d\mathbf{x}} \mathbf{x}^T \mathbf{A}^T \mathbf{y} = \mathbf{A}^T \mathbf{y}\end{aligned}\quad (16.3.4)$$

Note that in Eq. (16.3.4)  $\mathbf{x}^T \mathbf{A} \mathbf{x}$  is a scalar, known as the *quadratic form*, and  $\mathbf{y}^T \mathbf{A} \mathbf{x}$  is also a scalar, known as the *bilinear form*. A scalar is its own transpose, that is,  $\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T \mathbf{A}^T \mathbf{x}$  and  $\mathbf{y}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T \mathbf{A}^T \mathbf{y}$ .

Corresponding to Eq. (16.3.2), we can also define the following operations:

$$\begin{aligned}\left(\frac{d}{d\mathbf{x}}\right)^T \mathbf{x} &= \left[\frac{d}{dx_1} \quad \frac{d}{dx_2} \quad \cdots \quad \frac{d}{dx_n}\right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = n \\ \left(\frac{d}{d\mathbf{x}}\right)^T \mathbf{x} \mathbf{x}^T &= \left[\frac{d}{dx_1} \quad \frac{d}{dx_2} \quad \cdots \quad \frac{d}{dx_n}\right] \\ &\times \begin{bmatrix} x_1^2 & x_1 x_2 & \cdots & x_1 x_n \\ x_2 x_1 & x_2^2 & \cdots & x_2 x_n \\ \vdots & \vdots & & \vdots \\ x_n x_1 & x_n x_2 & \cdots & x_n^2 \end{bmatrix} = (n+1) \mathbf{x}^T \\ \left(\frac{d}{d\mathbf{x}}\right)^T \mathbf{x} \mathbf{y}^T &= \left[\frac{d}{dx_1} \quad \frac{d}{dx_2} \quad \cdots \quad \frac{d}{dx_n}\right] \\ &\times \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & & \vdots \\ x_n y_1 & x_n y_2 & \cdots & x_n y_n \end{bmatrix} = n \mathbf{y}^T\end{aligned}\quad (16.3.5)$$

**Example 16.3.1** Differentiate  $f(\mathbf{x})$  with respect to the vector  $\mathbf{x}$ :

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1^2 + 6x_1x_2 + 5x_2^2$$

Differentiating  $f(\mathbf{x})$  with respect to  $x_1$  and  $x_2$ , we obtain

$$\left. \begin{aligned}\frac{d}{dx_1} (3x_1^2 + 6x_1x_2 + 5x_2^2) &= 6x_1 + 6x_2 \\ \frac{d}{dx_2} (3x_1^2 + 6x_1x_2 + 5x_2^2) &= 6x_1 + 10x_2\end{aligned}\right\} = \begin{bmatrix} 6 & 6 \\ 6 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Using the formula  $[d(\mathbf{x}^T \mathbf{A} \mathbf{x})/d\mathbf{x}] = [\mathbf{A} + \mathbf{A}^T]\mathbf{x}$ , we obtain

$$\frac{d}{d\mathbf{x}} f(\mathbf{x}) = [\mathbf{A} + \mathbf{A}^T]\mathbf{x} = \left\{ \begin{bmatrix} 3 & 2 \\ 4 & 5 \end{bmatrix} + \begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix} \right\} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ 6 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

**Example 16.3.2** We differentiate  $f(\mathbf{x}, \mathbf{y})$  with respect to the vector  $\mathbf{x}$  in the following example:

$$f(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T \mathbf{A} \mathbf{x} = \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1y_1 + 4x_1y_2 + 2x_2y_1 + 5x_2y_2$$

Using the formula  $[d(\mathbf{y}^T \mathbf{A} \mathbf{x})/d\mathbf{x}] = \mathbf{A}^T \mathbf{y}$ , we can write

$$\mathbf{A}^T \mathbf{y} = \begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 3y_1 + 4y_2 \\ 2y_1 + 5y_2 \end{bmatrix}$$

### Complex Differentiation

**Total Derivatives** Differentiation with respect to a complex variable  $z = x + jy$  is a bit more involved. If we have a function of a complex variable

$$w = f(z) = u(x, y) + jv(x, y) \quad (16.3.6)$$

then the total derivative of  $f(z)$  with respect to  $z$  is defined by

$$\frac{d}{dz} w = \lim_{\Delta z \rightarrow 0} \frac{\Delta w}{\Delta z} = \lim_{\Delta z \rightarrow 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} \quad (16.3.7)$$

If the derivative  $df(z)/dz$  is to exist, then the limit must exist independent of the way in which  $\Delta z$  approaches zero. This condition imposes restrictions on the existence of the total derivative. With  $\Delta w = \Delta u + j\Delta v$  and  $\Delta z = \Delta x + j\Delta y$ , we have the following from Eq. (16.3.7):

$$\lim_{\Delta z \rightarrow 0} \Delta z = \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \frac{\Delta u + j\Delta v}{\Delta x + j\Delta y} \quad (16.3.8)$$

If we first let  $\Delta y \rightarrow 0$  and then let  $\Delta x \rightarrow 0$  in Eq. (16.3.8), we have

$$\frac{df(z)}{dz} = \frac{\partial u}{\partial x} + j \frac{\partial v}{\partial x} \quad (16.3.9)$$

Similarly, if we first let  $\Delta x \rightarrow 0$  and then let  $\Delta y \rightarrow 0$ , we obtain

$$\frac{df(z)}{dz} = \frac{\partial v}{\partial y} - j \frac{\partial u}{\partial y} \quad (16.3.10)$$

If the derivative  $df(z)/dz$  is to be unique then Eqs. (16.3.9) and (16.3.10) must match. Hence

$$\begin{aligned} \frac{\partial u}{\partial x} &= \frac{\partial v}{\partial y} \\ \frac{\partial v}{\partial x} &= -\frac{\partial u}{\partial y} \end{aligned} \quad (16.3.11)$$

These conditions are known as the *Cauchy–Riemann* (C–R) conditions. If C–R conditions are satisfied, then the derivative of  $f(z)$  is given by either Eq. (16.3.9) or Eq. (16.3.10).

A complex function  $f(z)$  is said to be analytic in the *neighborhood* of  $z = z_0$ , if C–R conditions are satisfied in that neighborhood. In addition, if it is also analytic at  $z = z_0$ , then the point  $z_0$  is a *regular point*. However, if it is not analytic at  $z = z_0$ , then  $z_0$  is called a *singular point*.

**Example 16.3.3** We will find the derivative of  $f(z) = z^2 = (x + jy)^2 = x^2 - y^2 + 2jxy$ . We will determine whether  $f(z)$  satisfies the Cauchy–Riemann conditions of Eq. (16.3.9).

$$z^2 = (x + jy)^2 = x^2 - y^2 + 2jxy$$

$$u(x, y) = x^2 - y^2 \quad : \quad v(x, y) = 2xy$$

$$\frac{\partial u}{\partial x} = 2x = \frac{\partial v}{\partial y} = 2x$$

$$\text{and hence} \quad \frac{\partial v}{\partial x} = 2y = -\frac{\partial u}{\partial y} = 2y$$

The C–R conditions are satisfied and  $df(z)/dz$  is given by

$$\frac{d}{dz} z^2 = \frac{\partial u}{\partial x} + j \frac{\partial v}{\partial x} = 2(x + jy) = 2z$$

**Example 16.3.4** On the other hand, we will find the derivative of  $f(z) = z^{*2} = (x - jy)^2 = x^2 - y^2 - 2jxy$ . We will determine whether  $f(z)$  satisfies the Cauchy–Riemann conditions of Eq. (16.3.9):

$$z^{*2} = (x - jy)^2 = x^2 - y^2 - 2jxy$$

$$u(x, y) = x^2 - y^2 \quad : \quad v(x, y) = -2xy$$

$$\frac{\partial u}{\partial x} = 2x \neq \frac{\partial v}{\partial y} = -2x$$

$$\frac{\partial v}{\partial x} = -2y \neq -\frac{\partial u}{\partial y} = 2y$$

Since the C–R conditions are not satisfied at points other than at  $z = 0$ ,  $z^{*2}$  does not have a total derivative  $dz^{*2}/dz$  at points other than at  $z = 0$ .

**Partial Derivatives** Many functions occurring in engineering practice are not analytic, and we are interested in optimization problems involving partial derivatives with respect to a complex variable  $z$ . With  $z = x + jy$ , we shall define the partial derivative operator  $\partial/\partial z$  as

$$\frac{\partial}{\partial z} = \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} \quad (16.3.12)$$

From the definition given in Eq. (16.3.12), we have the following:

$$\begin{aligned}\frac{\partial}{\partial z} z &= \left( \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} \right) (x + jy) = 1 - 1 = 0 \\ \frac{\partial}{\partial z} z^* &= \left( \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} \right) (x - jy) = 1 + 1 = 2\end{aligned}\quad (16.3.13)$$

Note that from Example 16.3.3 the total derivative  $(d/dz)z^*$  does not exist since  $z^*$  is not analytic except at  $z = 0$ .

From Eq. (16.3.12) we can also write

$$\begin{aligned}\frac{\partial}{\partial z} (z^* z) &= \left( \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} \right) (x^2 + y^2) \\ &= 2(x + jy) = 2z \\ \frac{\partial}{\partial z} z^2 &= \left( \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} \right) (x^2 + 2jxy - y^2) \\ &= 2(x + jy) - 2(x + jy) = 0 \\ \frac{\partial}{\partial z} (z^* w) &= \left( \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} \right) (xu + yv + j(xv - yu)) \\ &= 2(u + jv) = 2w \\ \frac{\partial}{\partial z} (zw) &= \left( \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} \right) (xu - yv + j(xv + yu)) \\ &= (u + jv) - (u + jv) = 0\end{aligned}\quad (16.3.14)$$

Note that if we used the chain rule of differentiation in Eq. (16.3.14) and apply Eq. (16.3.13), we obtain the same results. This fact gives credibility to the definition of the partial derivative  $\partial/\partial z$ . Note also that if  $z$  and  $w$  were real, then we would get  $(\partial/\partial z)z^2 = 2z$  and  $(\partial/\partial z)zw = w$ . These differences also manifest themselves in defining vector differentiation with respect to a complex variable.

**Vector Partial Differentiation** Analogous to Eq. (16.3.12), partial derivatives with respect to a vector complex variable can be defined. With  $\mathbf{z} = [x_1 + jy_1, x_2 + jy_2, \dots, x_n + jy_n]^T$ , partial derivatives  $\partial/\partial \mathbf{z}$  and  $\partial/\partial \mathbf{z}^*$  are defined as follows:

$$\mathbf{z} = \begin{bmatrix} x_1 + jy_1 \\ x_2 + jy_2 \\ \vdots \\ x_n + jy_n \end{bmatrix}; \quad \frac{\partial}{\partial \mathbf{z}} = \begin{bmatrix} \frac{\partial}{\partial x_1} + j \frac{\partial}{\partial y_1} \\ \frac{\partial}{\partial x_2} + j \frac{\partial}{\partial y_2} \\ \vdots \\ \frac{\partial}{\partial x_n} + j \frac{\partial}{\partial y_n} \end{bmatrix}; \quad \frac{\partial}{\partial \mathbf{z}^*} = \begin{bmatrix} \frac{\partial}{\partial x_1} - j \frac{\partial}{\partial y_1} \\ \frac{\partial}{\partial x_2} - j \frac{\partial}{\partial y_2} \\ \vdots \\ \frac{\partial}{\partial x_n} - j \frac{\partial}{\partial y_n} \end{bmatrix} \quad (16.3.15)$$

With this definition of  $\partial/\partial\mathbf{z}$  and  $\partial/\partial\mathbf{z}^*$ , we can define the following operations:

$$\begin{aligned}\frac{\partial}{\partial\mathbf{z}}\mathbf{z}^{\mathbf{H}} &= \begin{bmatrix} \frac{\partial}{\partial x_1} + j\frac{\partial}{\partial y_1} \\ \frac{\partial}{\partial x_2} + j\frac{\partial}{\partial y_2} \\ \vdots \\ \frac{\partial}{\partial x_n} + j\frac{\partial}{\partial y_n} \end{bmatrix} \begin{bmatrix} x_1 - jy_1 & x_2 - jy_2 & \cdots & x_n - jy_n \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} = 2\mathbf{I}_n\end{aligned}\quad (16.3.16)$$

where  $\mathbf{z}^{\mathbf{H}}$  is the conjugate transpose of  $\mathbf{z}$  and  $\mathbf{I}_n$  is an  $n$ th-order identity matrix. In a similar manner, we obtain

$$\begin{aligned}\frac{\partial}{\partial\mathbf{z}^*}\mathbf{z}^{\mathbf{T}} &= \begin{bmatrix} \frac{\partial}{\partial x_1} - j\frac{\partial}{\partial y_1} \\ \frac{\partial}{\partial x_2} - j\frac{\partial}{\partial y_2} \\ \vdots \\ \frac{\partial}{\partial x_n} - j\frac{\partial}{\partial y_n} \end{bmatrix} \begin{bmatrix} x_1 + jy_1 & x_2 + jy_2 & \cdots & x_n + jy_n \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} = 2\mathbf{I}_n\end{aligned}\quad (16.3.17)$$

In addition, we also have

$$\begin{aligned}\frac{\partial}{\partial\mathbf{z}}\mathbf{z}^{\mathbf{T}} &= \begin{bmatrix} \frac{\partial}{\partial x_1} + j\frac{\partial}{\partial y_1} \\ \frac{\partial}{\partial x_2} + j\frac{\partial}{\partial y_2} \\ \vdots \\ \frac{\partial}{\partial x_n} + j\frac{\partial}{\partial y_n} \end{bmatrix} \begin{bmatrix} x_1 + jy_1 & x_2 + jy_2 & \cdots & x_n + jy_n \end{bmatrix} \\ &= \mathbf{0}_n = \frac{\partial}{\partial\mathbf{z}^*}\mathbf{z}^{\mathbf{H}}\end{aligned}\quad (16.3.18)$$

where  $\mathbf{0}_n$  is an  $n$ th-order zero matrix.

Equations (16.3.16)–(16.3.18) are very useful in complex vector minimization of scalar quadratic error performance criteria, as we will show in the following examples.

**Example 16.3.5** The terms  $\mathbf{z}$  and  $\mathbf{w}$  are two complex 2-vectors as shown below. It is desired to minimize the complex matrix scalar criteria  $\varepsilon_1 = \mathbf{z}^{\mathbf{H}}\mathbf{w}$  and  $\varepsilon_2 = \mathbf{w}^{\mathbf{H}}\mathbf{z}$  with

respect to the complex vector  $\mathbf{z}$ . With

$$\mathbf{z} = \begin{bmatrix} a + jb \\ c + jd \end{bmatrix}; \quad \mathbf{w} = \begin{bmatrix} e + jf \\ g + jh \end{bmatrix}$$

and the corresponding error criteria

$$\varepsilon_1 = \mathbf{z}^H \mathbf{w} = ae + bf + cg + dh + j(af - be + ch - dg)$$

$$\varepsilon_2 = \mathbf{w}^H \mathbf{z} = ae + bf + cg + dh - j(af - be + ch - dg)$$

we can differentiate  $\varepsilon_1$  and  $\varepsilon_2$  partially with respect to  $\mathbf{z}$ , and using Eqs. (16.3.16) and (16.3.18), we obtain

$$\begin{aligned} \frac{\partial}{\partial \mathbf{z}} \varepsilon_1 &= \left( \frac{\partial}{\partial \mathbf{z}} \mathbf{z}^H \right) \mathbf{w} = \begin{bmatrix} e + e + j(f + f) \\ g + g + j(h + h) \end{bmatrix} = 2\mathbf{w} \\ \frac{\partial}{\partial \mathbf{z}} \varepsilon_2 &= \frac{\partial}{\partial \mathbf{z}} \mathbf{w}^H \mathbf{z} = \left( \frac{\partial}{\partial \mathbf{z}} \mathbf{z}^T \right) \mathbf{w}^* = \begin{bmatrix} e - e + j(f - f) \\ g - g + j(h - h) \end{bmatrix} = 0 \end{aligned}$$

The same result can be obtained by the chain rule of differentiation and Eqs. (16.3.16)–(16.3.18).

**Example 16.3.6** We take an example of minimizing a scalar quadratic error surface with respect to a complex weight vector  $\mathbf{a}_p$  called *adaptive coefficients* and finding the optimum value of  $\mathbf{a}_p$ . The quadratic error surface is given by

$$E_p = \sigma^2 + \mathbf{a}_p^H \mathbf{R}_p \mathbf{a}_p + \mathbf{r}_p^H \mathbf{a}_p + \mathbf{a}_p^H \mathbf{r}_p \quad (16.3.19)$$

We differentiate with respect to  $\mathbf{a}_p$  and set the result to zero and solve for  $\mathbf{a}_{po}$  using Eqs. (16.3.16)–(16.3.18):

$$\begin{aligned} \frac{\partial}{\partial \mathbf{a}_p} E_p &= \frac{\partial}{\partial \mathbf{a}_p} [\sigma^2 + \mathbf{a}_p^H \mathbf{R}_p \mathbf{a}_p + \mathbf{r}_p^H \mathbf{a}_p + \mathbf{a}_p^H \mathbf{r}_p] = 0 \\ &= \frac{\partial}{\partial \mathbf{a}_p} \sigma^2 + \frac{\partial}{\partial \mathbf{a}_p} \mathbf{a}_p^H \mathbf{R}_p \mathbf{a}_p + \frac{\partial}{\partial \mathbf{a}_p} \mathbf{r}_p^H \mathbf{a}_p + \frac{\partial}{\partial \mathbf{a}_p} \mathbf{a}_p^H \mathbf{r}_p = 0 \\ &= 0 + 2\mathbf{R}_p \mathbf{a}_p + 2\mathbf{r}_p = 0, \end{aligned} \quad (16.3.20)$$

or

$$\mathbf{a}_{po} = \mathbf{R}_p^{-1} \mathbf{r}_p$$

The solution  $\mathbf{a}_{po} = \mathbf{R}_p^{-1} \mathbf{r}_p$  is called the *Wiener solution* for the adaptive coefficients  $\mathbf{a}_p$ .

#### Summary of Some Vector Partial Derivatives of Real and Complex Quadratic Forms

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}} J(\mathbf{x}) &= \frac{\partial}{\partial \mathbf{x}} [\mathbf{x}^T \mathbf{A} \mathbf{x}] = \left[ \frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \right] \mathbf{A} \mathbf{x} + \left[ \frac{\partial}{\partial \mathbf{x}} \mathbf{x} \right] \mathbf{A}^T \mathbf{x} : \mathbf{x} \text{ real} \\ \frac{\partial}{\partial \mathbf{x}} J(\mathbf{x}) &= \frac{\partial}{\partial \mathbf{x}} [\mathbf{x}^H \mathbf{A} \mathbf{x}] = \left[ \frac{\partial}{\partial \mathbf{x}} \mathbf{x}^H \right] \mathbf{A} \mathbf{x} + \left[ \frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \right] \mathbf{A}^T \mathbf{x}^* = 2\mathbf{A} \mathbf{x} + 0 : \mathbf{x} \text{ complex} \\ \frac{\partial}{\partial \mathbf{x}} J(\mathbf{x}, \mathbf{y}) &= \frac{\partial}{\partial \mathbf{x}} [\mathbf{x}^T \mathbf{A} \mathbf{y}] = \left[ \frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \right] \mathbf{A} \mathbf{y} = \mathbf{A} \mathbf{y} : \mathbf{x} \text{ real} \\ \frac{\partial}{\partial \mathbf{x}} J(\mathbf{x}, \mathbf{y}) &= \frac{\partial}{\partial \mathbf{x}} [\mathbf{x}^H \mathbf{A} \mathbf{y}] = \left[ \frac{\partial}{\partial \mathbf{x}} \mathbf{x}^H \right] \mathbf{A} \mathbf{y} = 2\mathbf{A} \mathbf{y} : \mathbf{x} \text{ complex} \end{aligned} \quad (16.3.21)$$

A final point to note in taking the vector partial derivatives is that if  $\mathbf{x}$  is real, then  $(\partial/\partial\mathbf{x})\mathbf{x}^T = \mathbf{I}$ , and if  $\mathbf{x}$  is complex, then  $(\partial/\partial\mathbf{x})\mathbf{x}^H = 2\mathbf{I}$  and  $(\partial/\partial\mathbf{x})\mathbf{x}^T = \mathbf{0}$ .

### **Differentiation with Respect to Real Matrices**

In some minimization problems there is a need to differentiate a scalar function of a real matrix  $\mathbf{A}$  with respect to the real matrix  $\mathbf{A}$ . We will use the following result in Section 22.3. If  $\mathbf{B}$  is any square matrix and  $\mathbf{A}$  is any other matrix, then

$$\frac{\partial}{\partial\mathbf{A}} \text{Tr}[\mathbf{ABA}^T] = \mathbf{A}[\mathbf{B} + \mathbf{B}^T] \quad (16.3.22)$$

If  $\mathbf{B}$  is symmetric, then

$$\frac{\partial}{\partial\mathbf{A}} \text{Tr}[\mathbf{ABA}^T] = 2\mathbf{AB} \quad (16.3.23)$$

**Example 16.3.7** We are given the matrices  $\mathbf{X}$  and  $\mathbf{B}$  as

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

We will show that Eq. (16.3.22) is true. The trace of  $\mathbf{X}\mathbf{BX}^T$  is given by

$$\begin{aligned} \text{Tr}[\mathbf{X}\mathbf{BX}^T] &= x_{11}^2 b_{11} + x_{22}^2 b_{22} + x_{12}^2 b_{11} + x_{21}^2 b_{22} \\ &\quad + x_{11}x_{21}b_{12} + x_{11}x_{21}b_{21} + x_{22}x_{12}b_{12} + x_{22}x_{12}b_{21} \end{aligned}$$

Taking partial derivatives with respect to  $x_{11}$ ,  $x_{22}$ ,  $x_{12}$ , and  $x_{21}$ , we obtain

$$\frac{\partial}{\partial x_{11}} \text{Tr}[\mathbf{X}\mathbf{BX}^T] = 2x_{11}b_{11} + x_{21}(b_{21} + b_{12})$$

$$\frac{\partial}{\partial x_{22}} \text{Tr}[\mathbf{X}\mathbf{BX}^T] = 2x_{22}b_{22} + x_{12}(b_{21} + b_{12})$$

$$\frac{\partial}{\partial x_{12}} \text{Tr}[\mathbf{X}\mathbf{BX}^T] = 2x_{12}b_{11} + x_{22}(b_{21} + b_{12})$$

$$\begin{aligned} \frac{\partial}{\partial x_{21}} \text{Tr}[\mathbf{X}\mathbf{BX}^T] &= 2x_{21}b_{22} + x_{11}(b_{21} + b_{12}) \\ &= \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \left\{ \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{bmatrix} \right\} = \mathbf{X}\{\mathbf{B} + \mathbf{B}^T\} \end{aligned}$$

## **16.4 BLOCK MATRICES**

In some applications we will encounter block matrices whose elements are also matrices. An  $(n \times m) \times (n \times m)$  matrix  $\mathbf{Z}$  can be formulated as a block matrix consisting of

individual blocks **A**, **B**, **C**, **D**, as shown below:

$$\begin{aligned}
 \mathbf{Z} &= \left[ \begin{array}{cccccc}
 x_{11} & \cdots & x_{1n} & x_{1, n+1} & \cdots & x_{1, n+m} \\
 & \mathbf{A} & & & & \mathbf{B} \\
 \hline
 x_{n1} & \cdots & x_{n,n} & x_{n, n+1} & \cdots & x_{n, n+m} \\
 \mathbf{C} & & & & & \mathbf{D} \\
 \hline
 x_{n+1,1} & \cdots & x_{n+1,n} & x_{n+1, n+1} & \cdots & x_{n+1, n+m} \\
 x_{n+m,1} & \cdots & x_{n+m,n} & x_{n+m, n+1} & \cdots & x_{n+m, n+m}
 \end{array} \right] \\
 &= \left[ \begin{array}{c|c}
 \mathbf{A} & \mathbf{B} \\
 \hline
 \begin{matrix} n \times n \\ \mathbf{C} \\ m \times n \end{matrix} & \begin{matrix} n \times m \\ \mathbf{D} \\ m \times m \end{matrix}
 \end{array} \right] \tag{16.4.1}
 \end{aligned}$$

We can enumerate some properties of block matrices that may be useful in later chapters:

1. *Multiplication.* The compatibility of multiplying blocks must be satisfied as shown:

$$\begin{aligned}
 &\left[ \begin{array}{c|c}
 \mathbf{A} & \mathbf{B} \\
 \hline
 \begin{matrix} n \times n \\ \mathbf{C} \\ m \times n \end{matrix} & \begin{matrix} n \times m \\ \mathbf{D} \\ m \times m \end{matrix}
 \end{array} \right] \left[ \begin{array}{c|c}
 \mathbf{E} & \mathbf{F} \\
 \hline
 \begin{matrix} n \times n \\ \mathbf{G} \\ m \times n \end{matrix} & \begin{matrix} n \times m \\ \mathbf{H} \\ m \times m \end{matrix}
 \end{array} \right] \\
 &= \left[ \begin{array}{c|c}
 \mathbf{AE} + \mathbf{BG} & \mathbf{AF} + \mathbf{BH} \\
 \hline
 \begin{matrix} n \times n \\ \mathbf{CE} + \mathbf{DG} \\ m \times n \end{matrix} & \begin{matrix} n \times m \\ \mathbf{CF} + \mathbf{DH} \\ m \times m \end{matrix}
 \end{array} \right] \tag{16.4.2}
 \end{aligned}$$

2. *Determinant.* If **A** is nonsingular and **D** is square, then

$$\det \left[ \begin{array}{c|c}
 \mathbf{A} & \mathbf{B} \\
 \hline
 \mathbf{C} & \mathbf{D}
 \end{array} \right] = \det \mathbf{A} \cdot \det(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}) \tag{16.4.3}$$

3. *Inverse.*

- a. *Matrix Inversion Lemma.* If **A** and **D** are nonsingular  $n \times n$  and  $m \times m$  matrices respectively, then,

$$(\mathbf{A} + \mathbf{BD}^{-1}\mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} + \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} \tag{16.4.4}$$

- b. If inverses for **A** and **D** exist then,

$$\begin{aligned}
 \left[ \begin{array}{cc}
 \mathbf{A} & \mathbf{0} \\
 \mathbf{C} & \mathbf{D}
 \end{array} \right]^{-1} &= \left[ \begin{array}{cc}
 \mathbf{A}^{-1} & \mathbf{0} \\
 -\mathbf{D}^{-1}\mathbf{CA}^{-1} & \mathbf{D}^{-1}
 \end{array} \right]; \\
 \left[ \begin{array}{cc}
 \mathbf{A} & \mathbf{B} \\
 \mathbf{0} & \mathbf{D}
 \end{array} \right]^{-1} &= \left[ \begin{array}{cc}
 \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{BD}^{-1} \\
 \mathbf{0} & \mathbf{D}^{-1}
 \end{array} \right]
 \end{aligned} \tag{16.4.5}$$

c. If inverses for  $\mathbf{A}$  and  $\mathbf{D}$  exist then,

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \Delta_{\mathbf{A}}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\Delta_{\mathbf{D}}^{-1} \\ -\Delta_{\mathbf{D}}^{-1}\mathbf{C}\mathbf{A}^{-1} & \Delta_{\mathbf{D}}^{-1} \end{bmatrix} \quad (16.4.6)$$

where  $\Delta_{\mathbf{A}}$  and  $\Delta_{\mathbf{D}}$  are the Schur components of the matrices  $\mathbf{A}$  and  $\mathbf{D}$  defined by

$$\begin{aligned} \Delta_{\mathbf{A}} &= \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} \\ \Delta_{\mathbf{D}} &= \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{aligned} \quad (16.4.7)$$

d. Using property 3a, the inverses of the Schur component  $\Delta_{\mathbf{A}}$  and  $\Delta_{\mathbf{D}}$  are given by

$$\begin{aligned} \Delta_{\mathbf{A}}^{-1} &= \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\Delta_{\mathbf{D}}^{-1}\mathbf{C}\mathbf{A}^{-1} \\ \Delta_{\mathbf{D}}^{-1} &= \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\Delta_{\mathbf{A}}^{-1}\mathbf{B}\mathbf{D}^{-1} \end{aligned} \quad (16.4.8)$$

We can substitute for either  $\Delta_{\mathbf{A}}^{-1}$  or  $\Delta_{\mathbf{D}}^{-1}$  from Eq. (16.4.8) and solve Eq. (16.4.6).

# Random Vectors and Mean-Square Estimation

## 17.1 DISTRIBUTIONS AND DENSITIES

### *Joint Distributions and Densities*

We have come across sequences of random variables in Chapter 14, and these can be represented by a random  $n$ -vectors. We can define two random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  as

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}; \quad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} \quad (17.1.1)$$

The joint distribution of the random vector  $\mathbf{X}$  is given by

$$F_{\mathbf{X}}(\mathbf{x}) = P[\mathbf{X} \leq \mathbf{x}] = P[X_1 \leq x_1, \dots, X_n \leq x_n] \quad (17.1.2)$$

with  $F_{\mathbf{X}}(\infty) = \mathbf{1}$  and  $F_{\mathbf{X}}(-\infty) = \mathbf{0}$ . The corresponding joint density function  $f_{\mathbf{X}}(\mathbf{x})$  can be defined as

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \lim_{\Delta \mathbf{X} \rightarrow 0} \frac{P(\mathbf{x} < \mathbf{X} \leq \mathbf{x} + \Delta \mathbf{x})}{\Delta \mathbf{x}} \\ &= \lim_{\substack{\Delta x_1 \rightarrow 0 \\ \vdots \\ \Delta x_n \rightarrow 0}} \frac{P(x_1 < X_1 \leq x_1 + \Delta x_1, \dots, x_n < X_n \leq x_n + \Delta x_n)}{\Delta x_1 \cdots \Delta x_n} \end{aligned} \quad (17.1.3)$$

which is the  $n$ th partial derivative of the distribution function  $F_{\mathbf{X}}(\mathbf{x})$ :

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^n F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \cdots \partial x_n} \quad (17.1.4)$$

Given the density function  $f_{\mathbf{X}}(\mathbf{x})$ , the distribution function  $F_{\mathbf{X}}(\mathbf{x})$  can be obtained by an  $n$ -fold integration:

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{\mathbf{X}}(\mathbf{x}') dx'_1 \cdots dx'_n \quad (17.1.5)$$

**Joint Distribution of Two Vectors  $\mathbf{X}$  and  $\mathbf{Y}$**  Given an  $n$ -vector  $\mathbf{X} = [X_1 \ \cdots \ X_n]^T$  and an  $m$ -vector  $\mathbf{Y} = [Y_1 \ \cdots \ Y_m]^T$ , the joint distribution of  $\mathbf{X}$  and  $\mathbf{Y}$  is given by

$$F_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = P(\mathbf{X} \leq \mathbf{x}, \mathbf{Y} \leq \mathbf{y}) \quad (17.1.6)$$

and the joint density, by

$$\begin{aligned} f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) &= \lim_{\substack{\Delta \mathbf{x} \rightarrow 0 \\ \Delta \mathbf{y} \rightarrow 0}} \frac{P(\mathbf{x} < \mathbf{X} \leq \mathbf{x} + \Delta \mathbf{x}, \mathbf{y} < \mathbf{Y} \leq \mathbf{y} + \Delta \mathbf{y})}{\Delta \mathbf{x} \Delta \mathbf{y}} \\ &= \lim_{\substack{\Delta x_1 \rightarrow 0 \\ \vdots \\ \Delta x_n \rightarrow 0 \\ \Delta y_1 \rightarrow 0 \\ \vdots \\ \Delta y_m \rightarrow 0}} \frac{P\left(\begin{array}{l} x_1 < X_1 \leq x_1 + \Delta x_1, \dots, x_n < X_n \leq x_n + \Delta x_n \\ y_1 < Y_1 \leq y_1 + \Delta y_1, \dots, y_m < Y_m \leq y_m + \Delta y_m \end{array}\right)}{\Delta x_1 \cdots \Delta x_n \Delta y_1 \cdots \Delta y_m} \end{aligned}$$

or

$$f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = \frac{\partial^n \partial^m F_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})}{\partial x_1 \cdots \partial x_n \partial y_1 \cdots \partial y_m} \quad (17.1.7)$$

The distribution function  $F_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$  can also be obtained by integrating Eq. (17.1.7):

$$F_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = \int_{-\infty}^{\mathbf{x}} \int_{-\infty}^{\mathbf{y}} f_{\mathbf{XY}}(\mathbf{x}', \mathbf{y}') d\mathbf{y}' d\mathbf{x}' \quad (17.1.8)$$

Since Eq. (17.1.8) is a distribution function, we infer the following relationships:

$$\begin{aligned} F_{\mathbf{XY}}(\mathbf{x}, -\infty) &= F_{\mathbf{XY}}(-\infty, \mathbf{y}) = \mathbf{0} \\ F_{\mathbf{XY}}(\mathbf{x}, +\infty) &= F_{\mathbf{X}}(\mathbf{x}), \text{ marginal distribution of } \mathbf{X} \\ F_{\mathbf{XY}}(+\infty, \mathbf{y}) &= F_{\mathbf{Y}}(\mathbf{y}), \text{ marginal distribution of } \mathbf{Y} \\ F_{\mathbf{XY}}(+\infty, +\infty) &= \mathbf{1} \end{aligned} \quad (17.1.9)$$

The marginal density functions are obtained from Eq. (17.1.7) by integration as follows:

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) dy_1 \cdots dy_m \\ f_{\mathbf{Y}}(\mathbf{y}) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) dx_1 \cdots dx_n \end{aligned} \quad (17.1.10)$$

**Example 17.1.1** The joint density function of a random 2-vector  $\mathbf{X} = [X_1 \ X_2]^T$  and a random 2-vector  $\mathbf{Y} = [Y_1 \ Y_2]^T$  is given by

$$f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = \frac{1}{80} \left[ (x_1 + 2x_2)(4y_1 + y_2) \right], \quad \begin{cases} 0 < x_1 \leq 1, & 0 < x_2 \leq 2 \\ 1 < y_1 \leq 2, & 1 < y_2 \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

We have to determine (1) marginal density  $f_{\mathbf{X}}(\mathbf{x})$ , (2) marginal density  $f_{\mathbf{Y}}(\mathbf{y})$ , (3) marginal distribution  $F_{\mathbf{X}}(\mathbf{x})$ , (4) marginal distribution  $F_{\mathbf{Y}}(\mathbf{y})$ , (5) the distribution  $F_{\mathbf{X}}(x_1)$ , and (6) the distribution  $F_{\mathbf{X}}(x_2)$ .

1. The marginal density  $f_{\mathbf{X}}(\mathbf{x})$  is obtained by integrating  $f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$  over all  $\mathbf{y}$ :

$$f_{\mathbf{X}}(\mathbf{x}) = \int_1^3 \int_1^2 \frac{1}{80} [(x_1 + 2x_2)(4y_1 + y_2)] dy_1 dy_2 = \frac{1}{5} (x_1 + 2x_2)$$

2. The marginal density  $f_{\mathbf{Y}}(\mathbf{y})$  is obtained by integrating  $f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$  over all  $\mathbf{x}$ :

$$f_{\mathbf{Y}}(\mathbf{y}) = \int_0^2 \int_0^1 \frac{1}{80} [(x_1 + 2x_2)(4y_1 + y_2)] dx_1 dx_2 = \frac{1}{16} (4y_1 + y_2)$$

3. The marginal distribution  $F_{\mathbf{X}}(\mathbf{x})$  is obtained by integrating  $f_{\mathbf{X}}(\mathbf{x})$  as shown below:

$$F_{\mathbf{X}}(\mathbf{x}) = \int_0^{x_2} \int_0^{x_1} \frac{1}{5} (\xi_1 + 2\xi_2) d\xi_1 d\xi_2 = \frac{1}{10} (x_1^2 x_2 + 2x_1 x_2^2)$$

4. The marginal distribution  $F_{\mathbf{Y}}(\mathbf{y})$  is obtained by integrating  $f_{\mathbf{Y}}(\mathbf{y})$  as shown below:

$$\begin{aligned} F_{\mathbf{Y}}(\mathbf{y}) &= \int_1^{y_2} \int_1^{y_1} \frac{1}{16} (4\eta_1 + \eta_2) d\eta_1 d\eta_2 \\ &= \frac{1}{32} [4(y_2 - 1)y_1^2 + (y_2^2 - 1)y_1 - y_2^2 - 4y_2 + 5] \end{aligned}$$

5. The distribution  $F_{\mathbf{X}}(x_1)$  is obtained by substituting the end value  $x_2 = 2$  in  $F_{\mathbf{X}}(\mathbf{x})$ :

$$F_{\mathbf{X}}(x_1) = \frac{1}{10} (x_1^2 x_2 + 2x_1 x_2^2) \Big|_{x_2=2} = \frac{1}{5} (x_1^2 + 4x_1)$$

6. The distribution  $F_{\mathbf{X}}(x_2)$  is obtained by substituting the end value  $x_1 = 1$  in  $F_{\mathbf{X}}(\mathbf{x})$ :

$$F_{\mathbf{X}}(x_2) = \frac{1}{10} (x_1^2 x_2 + 2x_1 x_2^2) \Big|_{x_1=1} = \frac{1}{10} (2x_2^2 + x_2)$$

The random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  are independent since  $f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) \cdot f_{\mathbf{Y}}(\mathbf{y})$ , but the random variables  $X_1$  and  $X_2$  are not independent since  $F_{\mathbf{X}}(x_1, x_2) \neq F_{\mathbf{X}}(x_1) \cdot F_{\mathbf{X}}(x_2)$ .

### Conditional Distributions and Densities

The conditional distribution function  $F_{\mathbf{X}|A}(\mathbf{x}|A)$  given that an event  $A$  has occurred is defined by

$$F_{\mathbf{X}|A}(\mathbf{x}|A) = P(\mathbf{X} \leq \mathbf{x} | A) = \frac{P(\mathbf{X} \leq \mathbf{x}, A)}{P(A)}, \quad P(A) \neq 0 \quad (17.1.11)$$

Analogous to the one-dimensional case, we can also define a total probability. If there are disjoint events  $\{A_i, i = 1, \dots, n\}$  such that

$$\bigcup_{i=0}^n A_i = S, \text{ and } \bigcap_{i \neq j} A_i A_j = \emptyset$$

then the total probability is given by

$$F_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x}) = \sum_{i=1}^n P(\mathbf{X} \leq \mathbf{x} | A_i) P(A_i) = \sum_{i=1}^n F_{\mathbf{X}|A_i}(\mathbf{x} | A_i) P(A_i) \quad (17.1.12)$$

The density functions corresponding to Eqs. (17.1.11) and (17.1.12) are given by

$$f_{\mathbf{X}|A}(\mathbf{x} | A) = \frac{\partial^n F_{\mathbf{X}|A}(\mathbf{x} | A)}{\partial x_1 \cdots \partial x_n}, P(A) \neq 0 \quad (17.1.13)$$

$$f_{\mathbf{X}}(\mathbf{x}) = \sum_{i=1}^n f_{\mathbf{X}|A_i}(\mathbf{x} | A_i) P(A_i) \quad (17.1.14)$$

If there are events  $A$  and  $B$  defined by  $A: \mathbf{x} < \mathbf{X} \leq \mathbf{x} + \Delta\mathbf{x}$  and  $B: \mathbf{y} < \mathbf{Y} \leq \mathbf{y} + \Delta\mathbf{y}$ , then the probability of  $B$  occurring conditioned on  $A$  is given by

$$\begin{aligned} P(B | A) &= P(\mathbf{y} < \mathbf{Y} \leq \mathbf{y} + \Delta\mathbf{y} | \mathbf{x} < \mathbf{X} \leq \mathbf{x} + \Delta\mathbf{x}) = f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{X}) \Delta\mathbf{y} \\ &= \frac{P(\mathbf{y} < \mathbf{Y} \leq \mathbf{y} + \Delta\mathbf{y}, \mathbf{x} < \mathbf{X} \leq \mathbf{x} + \Delta\mathbf{x})}{P(\mathbf{x} < \mathbf{X} \leq \mathbf{x} + \Delta\mathbf{x})} \\ &= \frac{f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) \Delta\mathbf{x} \Delta\mathbf{y}}{f_{\mathbf{X}}(\mathbf{x}) \Delta\mathbf{x}} = \frac{f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) \Delta\mathbf{y}}{f_{\mathbf{X}}(\mathbf{x})} \end{aligned} \quad (17.1.15)$$

From Eq. (17.1.15) we can conclude that the conditional density  $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{X})$  is given by

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{X}) = \frac{f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})} \quad (17.1.16)$$

The corresponding expression for the conditional distribution  $F_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y} | \mathbf{X} \leq \mathbf{x})$  is

$$F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{X} \leq \mathbf{x}) = \frac{P(\mathbf{Y} \leq \mathbf{y}, \mathbf{X} \leq \mathbf{x})}{P(\mathbf{X} \leq \mathbf{x})} = \frac{F_{\mathbf{YX}}(\mathbf{y}, \mathbf{x})}{F_{\mathbf{X}}(\mathbf{x})} \quad (17.1.17)$$

**Example 17.1.2** In Example 17.1.1 we established that the random variables  $\mathbf{X}$  and  $\mathbf{Y}$  are independent and the random variables  $X_1$  and  $X_2$  are not independent. We will now find the conditional density  $f_{X_2|X_1}(x_2 | X_1 = x_1)$ . By Eq. (17.1.16), we have

$$\begin{aligned} f_{X_2|X_1}(x_2 | X_1 = x_1) &= \frac{f_{X_2X_1}(x_2, x_1)}{f_{X_1}(x_1)} \\ &= \frac{\frac{(x_1 + 2x_2)}{5}}{\frac{2(x_1 + 2)}{5}} = \frac{x_1 + 2x_2}{2(x_1 + 2)} \end{aligned}$$

and the corresponding conditional distribution  $F_{X_2|X_1}(x_2 | X_1 = x_1)$  is obtained by integrating the conditional density and is given by

$$F_{X_2|X_1}(x_2 | X_1 = x_1) = \int_0^{x_2} \frac{x_1 + 2\xi_2}{2(x_1 + 2)} d\xi_2 = \frac{x_2(x_1 + x_2)}{2(x_1 + 2)}$$

Note that when the end value of  $x_2 = 2$  is substituted in the distribution shown above, we obtain  $F_{X_2|X_1}(2|X_1 = x_1) = 1$ , as we should.

We can also find the conditional distribution  $F_{X_2|X_1}(x_2|X_1 \leq x_1)$  as follows:

$$\begin{aligned} F_{X_2|X_1}(x_2|X_1 \leq x_1) &= \frac{P(X_2 \leq x_2, X_1 \leq x_1)}{P(X_1 \leq x_1)} = \frac{F_{X_2X_1}(x_2, x_1)}{F_{X_1}(x_1)} \\ &= \frac{\frac{1}{10}(x_1^2 x_2 + 2x_1 x_2^2)}{\frac{1}{5}(x_1^2 + 4x_1)} = \frac{x_2(x_1 + 2x_2)}{2(x_1 + 4)} \end{aligned}$$

Here also if the end value of  $x_2 = 2$  is substituted in the preceding distribution, we obtain  $F_{X_2|X_1}(2|X_1 \leq x_1) = 1$ , as we should.

By differentiating the preceding equation with respect to  $x_2$ , we can find the corresponding conditional density function  $f_{X_2|X_1}(x_2|X_1 \leq x_1)$ .

$$f_{X_2|X_1}(x_2|X_1 \leq x_1) = \frac{d}{dx_2} \left[ \frac{x_2(x_1 + 2x_2)}{2(x_1 + 4)} \right] = \frac{x_1 + 4x_2}{2(x_1 + 4)}$$

The difference between these two conditional distributions is to be particularly noted. In the first case the conditioning event is  $\{X_1 = x_1\}$ , and in the second case it is  $\{X_1 \leq x_1\}$ .

### Points of Clarity

In Eq. (8.2.7) we noted some points of clarity regarding conditional densities and distributions. We will elaborate on this point. The joint density function  $f_{XY}(x,y)$  is defined in Eq. (9.2.3) but can also be defined as an infinitesimal probability

$$f_{XY}(x,y)\Delta x \Delta y = P(x < X \leq x + \Delta x, y < Y \leq y + \Delta y) \quad (17.1.18)$$

and the corresponding distribution function  $F_{XY}(x,y)$ , by

$$F_{XY}(x,y) = P(X \leq x, Y \leq y) \quad (9.2.2)$$

which can also be obtained by integrating  $f_{XY}(x,y)$  over  $x$  and  $y$ :

$$F_{XY}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(\xi,\eta) d\eta d\xi \quad (17.1.19)$$

The function  $f_{XY}(x,y)$  can also be obtained by differentiating  $F_{XY}(x,y)$  as given in Eq. (9.2.4):

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y} \quad (9.2.4)$$

There is a need for defining another density–distribution function  $f_{XY}(X \leq x, Y \leq y)$  that can be used in formulating directly conditional density problems involving the conditioning event such as  $\{X \leq x\}$ :

$$f_{XY}(X \leq x, Y \leq y) = \lim_{\Delta y \rightarrow 0} \frac{P(X \leq x, Y \leq y + \Delta y)}{\Delta y} \quad (17.1.20)$$

Equation (17.1.20) can be obtained by integrating  $f_{XY}(x,y)$  over  $x$  as shown below:

$$f_{XY}(X \leq x, y) = \int_{-\infty}^x f_{XY}(\xi, y) d\xi \quad (17.1.21)$$

We can integrate Eq. (17.1.21) over  $y$  and obtain  $F_{XY}(x,y)$ :

$$F_{XY}(x,y) = \int_{-\infty}^y f_{XY}(X \leq x, \eta) d\eta \quad (17.1.22)$$

The conditional density  $f_{Y|X}(y | X \leq x)$  can be determined as follows:

$$\begin{aligned} f_{Y|X}(y | X \leq x) \Delta y &= P(y < Y \leq y + \Delta y | X \leq x) \\ &= \frac{P(y < Y \leq y + \Delta y, X \leq x)}{P(X \leq x)} \\ &= \frac{f_{YX}(y, X \leq x) \Delta y}{F_X(x)} \end{aligned}$$

Hence

$$f_{Y|X}(y | X \leq x) = \frac{f_{YX}(y, X \leq x)}{F_X(x)} \quad (17.1.23)$$

**Example 17.1.3** We will continue the previous example (Example 17.1.2) and find the conditional density  $f_{X_2|X_1}(x_2 | X_1 \leq x_1)$ . From Eq. (17.1.23) we have

$$f_{X_2|X_1}(x_2 | X_1 \leq x_1) = \frac{f_{X_2X_1}(x_2, X_1 \leq x_1)}{F_{X_1}(x_1)}$$

From Eq. (17.1.21) we have

$$f_{X_2X_1}(x_2, X_1 \leq x_1) = \int_0^{x_1} \frac{\xi_1 + 2x_2}{5} d\xi_1 = \frac{x_1(x_1 + 4x_2)}{10} \quad \text{and} \quad F_{X_1}(x_1) = \frac{x_1^2 + 4x_1}{5}$$

Hence

$$f_{X_2|X_1}(x_2 | X_1 \leq x_1) = \frac{\frac{x_1(x_1 + 4x_2)}{10}}{\frac{x_1^2 + 4x_1}{5}} = \frac{x_1 + 4x_2}{2(x_1 + 4)}$$

By integrating  $f_{X_2|X_1}(x_2 | X_1 \leq x_1)$  with respect to  $x_2$ , we obtain the conditional distribution

$$F_{X_2|X_1}(x_2 | X_1 \leq x_1) = \frac{x_2(x_1 + 2x_2)}{2(x_1 + 4)}$$

### Bayes' Theorem

Similar to the one-dimensional case, we can formulate a vector form of Bayes' theorem given that the event  $A_i$  has occurred. We want to find the a posteriori probability  $P(A_i | \mathbf{X} \leq \mathbf{x})$  given that the event  $\{\mathbf{X} \leq \mathbf{x}\}$  has occurred. This probability is given by

$$P(A_i | \mathbf{X} \leq \mathbf{x}) = \frac{P(\mathbf{X} \leq \mathbf{x}, A_i)}{P(\mathbf{X} \leq \mathbf{x})} = \frac{P(\mathbf{X} \leq \mathbf{x} | A_i)P(A_i)}{P(\mathbf{X} \leq \mathbf{x})}$$

Substituting for  $P(\mathbf{X} \leq \mathbf{x})$  from Eq. (17.1.12), we obtain the vector form of Bayes' theorem:

$$P(A_i | \mathbf{X} \leq \mathbf{x}) = \frac{P(\mathbf{X} \leq \mathbf{x} | A_i)P(A_i)}{\sum_{k=1}^n P(\mathbf{X} \leq \mathbf{x} | A_k)P(A_k)} \quad (17.1.24)$$

We can also formulate Bayes' rule for random vector densities. If  $f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$  represents the joint density associated with random vectors  $\mathbf{X}$  and  $\mathbf{Y}$ , then, from Eq. (17.1.16), we obtain

$$f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{X} = \mathbf{x})f_{\mathbf{X}}(\mathbf{x})$$

and Bayes' theorem can be formulated for  $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} | \mathbf{Y} = \mathbf{y})$  as follows:

$$\begin{aligned} f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} | \mathbf{Y} = \mathbf{y}) &= \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{X} = \mathbf{x})f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y})} \\ &= \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{X} = \mathbf{x})f_{\mathbf{X}}(\mathbf{x})}{\int_{-\infty}^{\infty} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{X} = \mathbf{x}')f_{\mathbf{X}}(\mathbf{x}')d\mathbf{x}'} \end{aligned} \quad (17.1.25)$$

In many engineering estimation problems the random vector  $\mathbf{X}$  cannot be observed but the random vector  $\mathbf{Y}$  connected to  $\mathbf{X}$  can be observed. In this connection  $f_{\mathbf{X}}(\mathbf{x})$  is called the *a priori density* and  $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} | \mathbf{Y} = \mathbf{y})$  is called the *a posteriori density* after observing  $\{\mathbf{Y} = \mathbf{y}\}$ .

**Example 17.1.4** The density  $f_{\mathbf{X}}(\mathbf{x})$  of a two-dimensional random vector  $\mathbf{X}$  is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{4x_1 + 7x_2}{18}, \quad 0 < x_1 \leq 1, 0 < x_2 \leq 2$$

The conditional density  $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x})$  of another two-dimensional random vector  $\mathbf{Y}$  is

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}) = \frac{4x_1y_1 + 2x_1y_2 + 5x_2y_1 + 5x_2y_2}{5(4x_1 + 7x_2)}, \quad \begin{cases} 0 < x_1 \leq 1, 0 < x_2 \leq 2 \\ 1 < y_1 \leq 2, 1 < y_2 \leq 3 \end{cases}$$

Given this information, we have to find the a posteriori density  $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} | \mathbf{y})$  using Bayes' theorem of Eq. (17.1.24). The joint density function  $f_{\mathbf{YX}}(\mathbf{x}, \mathbf{y})$  is given by

$$\begin{aligned} f_{\mathbf{YX}}(\mathbf{y}, \mathbf{x}) &= f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x})f_{\mathbf{X}}(\mathbf{x}) \\ &= \frac{4x_1y_1 + 2x_1y_2 + 5x_2y_1 + 5x_2y_2}{5(4x_1 + 7x_2)} \frac{4x_1 + 7x_2}{18} \\ &= \frac{1}{90}(4x_1y_1 + 2x_1y_2 + 5x_2y_1 + 5x_2y_2), \quad \begin{cases} 0 < x_1 \leq 1, 0 < x_2 \leq 2 \\ 1 < y_1 \leq 2, 1 < y_2 \leq 3 \end{cases} \end{aligned}$$

The marginal density function  $f_{\mathbf{Y}}(\mathbf{y})$  is obtained by integrating the preceding equation with respect to  $\mathbf{x}$ :

$$f_{\mathbf{Y}}(\mathbf{y}) = \int_0^1 \int_0^2 \frac{1}{90} \left( 4x_1y_1 + 2x_1y_2 + 5x_2y_1 + 5x_2y_2 \right) dx_2 dx_1 = \frac{7y_1 + 6y_2}{45}$$

The a posteriori density can be obtained from Eq. (17.1.25) as

$$\begin{aligned}
 f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) &= \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y})} \\
 &= \frac{\frac{4x_1y_1 + 2x_1y_2 + 5x_2y_1 + 5x_2y_2}{5(4x_1 + 7x_2)} \frac{4x_1 + 7x_2}{18}}{\frac{7y_1 + 6y_2}{45}} \\
 &= \frac{4x_1y_1 + 2x_1y_2 + 5x_2y_1 + 5x_2y_2}{2(7y_1 + 6y_2)}, \quad \begin{cases} 0 < x_1 < 1, 0 < x_2 \leq 2 \\ 1 < y_1 \leq 2, 1 < y_2 \leq 3 \end{cases}
 \end{aligned}$$

which, when integrated over all  $\mathbf{x}$ , indeed gives the result 1.

**Example 17.1.5** The density  $f_{\mathbf{X}}(\mathbf{x})$  of a two-dimensional vector  $\mathbf{X}$  is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \alpha_1 \delta(x_1 - \mu_1, x_2 - \mu_2) + \alpha_2 \delta(x_1 - \mu_1, x_2 - \mu_2)$$

where  $\alpha_1 + \alpha_2 = 1$ ,  $\delta(x_1, x_2)$  is a two-dimensional Dirac delta function and  $\mu_1, \mu_2$  are constants. The conditional density of a two-dimensional random vector  $\mathbf{Y}$  correlated with  $\mathbf{X}$  is

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi} \left\{ e^{-[(y_1 - x_1)^2 + (y_2 - x_2)^2]} \right\}$$

From Eq. (17.1.6) the joint density function  $f_{\mathbf{YX}}(\mathbf{y}, \mathbf{x})$  is given by

$$\begin{aligned}
 f_{\mathbf{YX}}(\mathbf{y}, \mathbf{x}) &= f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})f_{\mathbf{X}}(\mathbf{x}) \\
 &= \frac{1}{\pi} \left\{ e^{-[(y_1 - x_1)^2 + (y_2 - x_2)^2]} \right\} \{ \alpha_1 \delta(x_1 - \mu_1, x_2 - \mu_2) + \alpha_2 \delta(x_1 - \mu_1, x_2 - \mu_2) \} \\
 &= \frac{1}{\pi} \left\{ e^{-[(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2]} \right\} \{ \alpha_1 \delta(x_1 - \mu_1, x_2 - \mu_2) + \alpha_2 \delta(x_1 - \mu_1, x_2 - \mu_2) \}
 \end{aligned}$$

The marginal density function  $f_{\mathbf{Y}}(\mathbf{y})$  is obtained by integrating this equation with respect to  $\mathbf{x}$ :

$$\begin{aligned}
 f_{\mathbf{Y}}(\mathbf{y}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\pi} \left\{ e^{-[(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2]} \right\} \left\{ \alpha_1 \delta(x_1 - \mu_1, x_2 - \mu_2) + \alpha_2 \delta(x_1 - \mu_1, x_2 - \mu_2) \right\} dx_1 dx_2 \\
 &= \frac{1}{\pi} \left\{ \alpha_1 e^{-[(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2]} + \alpha_2 e^{-[(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2]} \right\}
 \end{aligned}$$

The a posteriori density can be obtained from Eq. (17.1.25) as

$$\begin{aligned}
 f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) &= \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y})} \\
 &= \frac{\left\{ e^{-[(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2]} \right\} \{ \alpha_1 \delta(x_1 - \mu_1, x_2 - \mu_2) + \alpha_2 \delta(x_1 - \mu_1, x_2 - \mu_2) \}}{\alpha_1 e^{-[(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2]} + \alpha_2 e^{-[(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2]}} \\
 &= \beta_1(\mathbf{y}) \delta(x_1 - \mu_1, x_2 - \mu_2) + \beta_2(\mathbf{y}) \delta(x_1 - \mu_1, x_2 - \mu_2)
 \end{aligned}$$

where

$$\beta_i(\mathbf{y}) = \frac{\alpha_i e^{-[(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2]}}{\alpha_1 e^{-[(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2]} + \alpha_2 e^{-[(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2]}}, \quad i = 1, 2$$

## 17.2 MOMENTS OF RANDOM VECTORS

### Expectation Vector

Analogous to the single-dimensional case, the *expected value* of a random  $n$ -vector  $\mathbf{X}$  is another  $n$ -vector  $\mu_{\mathbf{X}}$  defined by

$$E[\mathbf{X}] = \mu_{\mathbf{X}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{x} f_{\mathbf{X}}(\mathbf{x}) dx_1 \cdots dx_n \quad (17.2.1)$$

and

$$\begin{aligned} E[X_i] = \mu_{X_i} &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i f_{\mathbf{X}}(\mathbf{x}) dx_1 \cdots dx_i \cdots dx_n \\ &= \int_{-\infty}^{\infty} x_i f_{X_i}(x_i) dx_i \end{aligned} \quad (17.2.2)$$

since

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n = 1$$

Hence the mean value  $n$ -vector  $\mu_{\mathbf{X}}$  can be written as

$$\mu_{\mathbf{X}} = [\mu_{X_1} \ \cdots \ \mu_{X_i} \ \cdots \ \mu_{X_n}]^T \quad (17.2.3)$$

**Example 17.2.1** The term  $\mathbf{X}$  is a 3-vector whose pdf  $f_{\mathbf{X}}(\mathbf{x})$  is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{bmatrix} 2e^{-2x_1} \\ 3e^{-3x_2} \\ 4e^{-4x_3} \end{bmatrix}, \quad x_1, x_2, x_3 > 0$$

The mean value vector  $\mu_{\mathbf{X}}$  is:

$$\mu_{\mathbf{X}} = \begin{bmatrix} E(X_1) \\ E(X_2) \\ E(X_3) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{4} \end{bmatrix}$$

### Correlation Matrix

The *correlation matrix* is a square matrix  $\mathbf{R}_{\mathbf{X}}$  representing the set of all second moments of the components of a real random vector  $\mathbf{X}$  of the form

$$E[X_i X_j] = \int_{-\infty}^{\infty} x_i x_j f_{X_i X_j}(x_i, x_j) dx_i dx_j, \quad i \neq j, \quad i, j = 1, \dots, n \quad (17.2.4)$$

and is defined by

$$E[\mathbf{X} \mathbf{X}^T] = \mathbf{R}_{\mathbf{X}} = \begin{bmatrix} E[X_1^2] & E[X_1 X_2] & \cdots & E[X_1 X_n] \\ E[X_2 X_1] & E[X_2^2] & \cdots & E[X_2 X_n] \\ \vdots & \vdots & & \vdots \\ E[X_n X_1] & E[X_n X_2] & \cdots & E[X_n^2] \end{bmatrix} \quad (17.2.5)$$

The quantity  $E[\mathbf{XX}^T]$  is defined as the *outer product* of the vector  $\mathbf{X}$ , which is equal to the correlation matrix  $\mathbf{R}_X$ . The *inner product*  $\mathbf{R}_X$  is defined by  $E[\mathbf{X}^T \mathbf{X}] = \sum_{i=1}^n E[X_i^2]$ , which is equal to the trace of  $\mathbf{R}_X$  or  $E[\mathbf{X}^T \mathbf{X}] = \text{Tr}[\mathbf{R}_X]$ .

The expected value of a scalar function  $g(\mathbf{X})$  of a vector random variable  $\mathbf{X}$  can be defined as follows:

$$E[g(\mathbf{X})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\mathbf{X}) f_{\mathbf{X}}(\mathbf{x}) dx_1 \cdots dx_n \quad (17.2.6)$$

**Example 17.2.2** The correlation matrix of Example 17.2.1 can be obtained from Eq. 17.2.4 assuming that  $X_1, X_2, X_3$  are independent and hence the joint densities  $f_{X_i X_j}(x_i, x_j)$  are given by the product of individual densities as shown below:

$$f_{X_i X_j}(x_i, x_j) = f_{X_i}(x_i) f_{X_j}(x_j), \quad i \neq j, \quad i, j = 1, 2, 3$$

or

$$f_{X_1 X_2}(x_1, x_2) = 6e^{-(2x_1 + 3x_2)} u(x_1) u(x_2)$$

$$f_{X_2 X_3}(x_2, x_3) = 12e^{-(3x_2 + 4x_3)} u(x_2) u(x_3)$$

$$f_{X_1 X_3}(x_1, x_3) = 8e^{-(2x_1 + 4x_3)} u(x_1) u(x_3)$$

Hence  $\mathbf{R}_X$  is given by

$$\mathbf{R}_X = \begin{bmatrix} E[X_1^2] & E[X_1 X_2] & E[X_1 X_3] \\ E[X_2 X_1] & E[X_2^2] & E[X_2 X_3] \\ E[X_3 X_1] & E[X_3 X_2] & E[X_3^2] \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{8} \\ \frac{1}{6} & \frac{2}{9} & \frac{1}{12} \\ \frac{1}{8} & \frac{1}{12} & \frac{1}{8} \end{bmatrix}$$

The trace of the matrix  $\text{Tr}(\mathbf{R}_X) = \frac{1}{2} + \frac{2}{9} + \frac{1}{8} = \frac{61}{72}$ . We note that the correlation matrix is symmetric.

### Covariance Matrix

The *covariance matrix*  $\mathbf{C}_X$  associated with a real random vector  $\mathbf{X}$  is the expected value of the outer product of the vector  $(\mathbf{X} - \mu_X)$  and is given by

$$\mathbf{C}_X = E[\mathbf{X} - \mu_X][\mathbf{X} - \mu_X]^T = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} & \cdots & \sigma_{X_1 X_j} & \cdots & \sigma_{X_1 X_n} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2 X_j} & \cdots & \sigma_{X_2 X_n} \\ \vdots & \vdots & & \vdots & & \vdots \\ \sigma_{X_i X_1} & \sigma_{X_i X_2} & \cdots & \sigma_{X_i X_j} & \cdots & \sigma_{X_i X_n} \\ \vdots & \vdots & & \vdots & & \vdots \\ \sigma_{X_n X_1} & \sigma_{X_n X_2} & \cdots & \sigma_{X_n X_j} & \cdots & \sigma_{X_n}^2 \end{bmatrix} \quad (17.2.7)$$

where  $\sigma_{X_i X_j}$  is the covariance between the random variables  $X_i$  and  $X_j$  given by

$$\sigma_{X_i X_j} = E(X_i - \mu_{X_i})(X_j - \mu_{X_j}) \quad (17.2.8)$$

Equation (17.2.7) can be expanded to express the covariance matrix  $\mathbf{C}_X$  in terms of the correlation matrix  $\mathbf{R}_X$ :

$$\begin{aligned} \mathbf{C}_X &= E[\mathbf{XX}^T] - E[\mathbf{X}]\mu_X^T - \mu_X E[\mathbf{X}^T] + \mu_X \mu_X^T \\ &= E[\mathbf{XX}^T] - \mu_X \mu_X^T = \mathbf{R}_X - \mu_X \mu_X^T \end{aligned} \quad (17.2.9)$$

**Example 17.2.3** We will find the covariance matrix for Examples 17.2.1 and 17.2.2. The correlation matrix  $\mathbf{R}_X$  has already been found. The matrix  $\mu_X \mu_X^T$  is given by

$$\begin{aligned}\mu_X \mu_X^T &= \begin{bmatrix} E(X_1) \\ E(X_2) \\ E(X_3) \end{bmatrix} \begin{bmatrix} E(X_1) & E(X_2) & E(X_3) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{4} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{16} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{12} \\ \frac{1}{4} & \frac{1}{12} & \frac{1}{16} \end{bmatrix}\end{aligned}$$

From Eq. (17.2.9) the covariance matrix can be given by

$$\mathbf{C}_X = \mathbf{R}_X - \mu_X \mu_X^T = \begin{bmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{8} \\ \frac{1}{6} & \frac{1}{9} & \frac{1}{12} \\ \frac{1}{8} & \frac{1}{12} & \frac{1}{8} \end{bmatrix} - \begin{bmatrix} \frac{1}{4} & \frac{1}{6} & \frac{1}{8} \\ \frac{1}{6} & \frac{1}{9} & \frac{1}{12} \\ \frac{1}{8} & \frac{1}{12} & \frac{1}{16} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{9} & 0 \\ 0 & 0 & \frac{1}{16} \end{bmatrix}$$

Since the cross-terms are zero, the random variables  $X_1, X_2$ , and  $X_3$  are pairwise uncorrelated.

### Joint Moments of Two Random Vectors

Given a random  $n$ -vector  $\mathbf{X}$  and  $m$ -vector  $\mathbf{Y}$ , with joint density  $f_{XY}(\mathbf{x}, \mathbf{y})$ , the expectation of a scalar function  $g(\mathbf{X}, \mathbf{Y})$  of the random vectors  $\mathbf{X}, \mathbf{Y}$  is defined by

$$\begin{aligned}E[g(\mathbf{X}, \mathbf{Y})] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\mathbf{x}, \mathbf{y}) f_{XY}(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\mathbf{x}, \mathbf{y}) f_{XY}(\mathbf{x}, \mathbf{y}) dx_1 \cdots dx_n dy_1 \cdots dy_m\end{aligned}\quad (17.2.10)$$

The *cross-correlation matrix*  $\mathbf{R}_{XY}$  represents the set of all joint moments of the components of real random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  of the form

$$E[X_i Y_j] = \int_{-\infty}^{\infty} x_i y_j f_{X_i Y_j}(x_i, y_j) dx_i dy_j, \quad i = 1, \dots, n, j = 1, \dots, m \quad (17.2.11)$$

and is defined by the outer product  $E[\mathbf{XY}^T]$  as follows:

$$E[\mathbf{XY}^T] = \mathbf{R}_{XY} = \begin{bmatrix} E[X_1 Y_1] & E[X_1 Y_2] & \cdots & E[X_1 Y_m] \\ E[X_2 Y_1] & E[X_2 Y_2] & \cdots & E[X_2 Y_m] \\ \vdots & \vdots & & \vdots \\ E[X_n Y_1] & E[X_n Y_2] & \cdots & E[X_n Y_m] \end{bmatrix} \quad (17.2.12)$$

It is not a square matrix unless  $m = n$ . The inner product  $E[\mathbf{X}^T \mathbf{Y}]$  is not defined unless  $m = n$ , in which case  $E[\mathbf{X}^T \mathbf{Y}] = \sum_{i=1}^n E[X_i Y_i]$ .

### Cross-Covariance Matrix

The cross-covariance matrix  $\mathbf{C}_{\mathbf{XY}}$  associated with random  $n$ -vector  $\mathbf{X}$  and a random  $m$ -vector  $\mathbf{Y}$  is expected value of the outer product  $(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}})^T$  and is given by

$$\mathbf{C}_{\mathbf{XY}} = E[\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}][\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}}]^T = \begin{bmatrix} \sigma_{X_1 Y_1} & \sigma_{X_1 Y_2} & \cdots & \sigma_{X_1 Y_j} & \cdots & \sigma_{X_1 Y_m} \\ \sigma_{X_2 Y_1} & \sigma_{X_2 Y_2} & \cdots & \sigma_{X_2 Y_j} & \cdots & \sigma_{X_2 Y_m} \\ \vdots & \vdots & & \vdots & & \vdots \\ \sigma_{X_i Y_1} & \sigma_{X_i Y_2} & \cdots & \sigma_{X_i Y_j} & \cdots & \sigma_{X_i Y_m} \\ \vdots & \vdots & & \vdots & & \vdots \\ \sigma_{X_n Y_1} & \sigma_{X_n Y_2} & \cdots & \sigma_{X_n Y_j} & \cdots & \sigma_{X_n Y_m} \end{bmatrix} \quad (17.2.13)$$

Again  $\mathbf{C}_{\mathbf{XY}}$  is not a square matrix unless  $m = n$ . If  $m = n$  then  $\text{Tr}[\mathbf{C}_{\mathbf{XY}}] = \sum_{i=1}^n E[X_i Y_i]$ .

Similar to Eq. (17.2.9), Eq. (17.2.13) can be expanded to express the cross-covariance matrix  $\mathbf{C}_{\mathbf{XY}}$  in terms of the cross-correlation matrix  $\mathbf{R}_{\mathbf{XY}}$ :

$$\begin{aligned} \mathbf{C}_{\mathbf{XY}} &= E[\mathbf{XY}^T] - E[\mathbf{X}]\boldsymbol{\mu}_{\mathbf{Y}}^T - \boldsymbol{\mu}_{\mathbf{X}}E[\mathbf{Y}^T] + \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\mu}_{\mathbf{Y}}^T = E[\mathbf{XY}^T] - \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\mu}_{\mathbf{Y}}^T \\ &= \mathbf{R}_{\mathbf{XY}} - \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\mu}_{\mathbf{Y}}^T \end{aligned} \quad (17.2.14)$$

**Example 17.2.4** We will now find the cross-covariance matrix of the two-dimensional vectors  $\mathbf{X}$  and  $\mathbf{Y}$  given the joint density

$$f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = \frac{1}{90}(4x_1y_1 + 2x_1y_2 + 5x_2y_1 + 5x_2y_2), \begin{cases} 0 < x_1 \leq 1, 0 < x_2 \leq 2 \\ 1 < y_1 \leq 2, 1 < y_2 \leq 3 \end{cases}$$

We will first find the four densities,  $f_{X_1}(x_1)$ ,  $f_{X_2}(x_2)$ ,  $f_{Y_1}(y_1)$ ,  $f_{Y_2}(y_2)$  as follows:

$$\begin{aligned} f_{X_1}(x_1) &= \int_0^2 \int_1^2 \int_1^3 \frac{1}{90}(4x_1y_1 + 2x_1y_2 + 5x_2y_1 + 5x_2y_2) dy_2 dy_1 dx_2 = \frac{4x_1 + 7}{9} \\ f_{X_2}(x_2) &= \int_0^1 \int_1^2 \int_1^3 \frac{1}{90}(4x_1y_1 + 2x_1y_2 + 5x_2y_1 + 5x_2y_2) dy_2 dy_1 dx_1 = \frac{7x_2 + 2}{18} \\ f_{Y_1}(y_1) &= \int_0^1 \int_0^2 \int_1^3 \frac{1}{90}(4x_1y_1 + 2x_1y_2 + 5x_2y_1 + 5x_2y_2) dy_2 dx_2 dx_1 = \frac{14y_1 + 24}{45} \\ f_{Y_2}(y_2) &= \int_0^1 \int_0^2 \int_1^2 \frac{1}{90}(4x_1y_1 + 2x_1y_2 + 5x_2y_1 + 5x_2y_2) dy_1 dx_2 dx_1 = \frac{4y_1 + 7}{30} \end{aligned}$$

The mean vectors  $\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\mu}_{\mathbf{Y}}$  and the outer product  $\boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\mu}_{\mathbf{Y}}^T$  can be obtained from the densities as follows:

$$\boldsymbol{\mu}_{\mathbf{X}} = \begin{bmatrix} \frac{29}{54} \\ \frac{5}{27} \end{bmatrix}, \quad \boldsymbol{\mu}_{\mathbf{Y}} = \begin{bmatrix} \frac{134}{45} \\ \frac{119}{180} \end{bmatrix}, \quad \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\mu}_{\mathbf{Y}}^T = \begin{bmatrix} \frac{1943}{1215} & \frac{3451}{9720} \\ \frac{134}{243} & \frac{119}{972} \end{bmatrix}$$

To find the cross-correlation matrix  $\mathbf{R}_{\mathbf{XY}}$ , we need the joint densities  $f_{X_1 Y_1}(x_1, y_1)$ ,  $f_{X_1 Y_2}(x_1, y_2)$ ,  $f_{X_2 Y_1}(x_2, y_1)$ ,  $f_{X_2 Y_2}(x_2, y_2)$ . They can be determined from  $f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$  and are shown below:

$$\begin{aligned} f_{X_1 Y_1}(x_1, y_1) &= \frac{8x_1 y_1 + 8x_1 + 10y_1 + 20}{45} \\ f_{X_1 Y_2}(x_1, y_2) &= \frac{4x_1 y_2 + 12x_1 + 10y_2 + 15}{90} \\ f_{X_2 Y_1}(x_2, y_1) &= \frac{5x_2 y_1 + 10x_2 + 2y_1 + 2}{45} \\ f_{X_2 Y_2}(x_2, y_2) &= \frac{10x_2 y_2 + 15x_2 + 2y_2 + 6}{180} \end{aligned}$$

We can now find the various joint expectations as follows:

$$\begin{aligned} E[X_1 Y_1] &= \frac{332}{405} & E[X_1 Y_2] &= \frac{454}{405}, \quad \text{hence} \quad \mathbf{R}_{\mathbf{XY}} = \begin{bmatrix} \frac{332}{405} & \frac{454}{405} \\ \frac{778}{405} & \frac{1066}{405} \end{bmatrix} \\ E[X_2 Y_1] &= \frac{778}{405} & E[X_2 Y_2] &= \frac{1066}{405} \end{aligned}$$

From Eq. (17.2.14) we can evaluate  $\mathbf{C}_{\mathbf{XY}} = \mathbf{R}_{\mathbf{XY}} - \boldsymbol{\mu}_{\mathbf{X}} \boldsymbol{\mu}_{\mathbf{Y}}^T$ :

$$\begin{aligned} \mathbf{C}_{\mathbf{XY}} &= \begin{bmatrix} \frac{332}{405} & \frac{454}{405} \\ \frac{778}{405} & \frac{1066}{405} \end{bmatrix} - \begin{bmatrix} \frac{1943}{1215} & \frac{3451}{9720} \\ \frac{134}{243} & \frac{119}{972} \end{bmatrix} = \begin{bmatrix} \frac{-163,963}{108,135} & \frac{1489}{1944} \\ \frac{1664}{1215} & \frac{12,197}{4860} \end{bmatrix} \\ &= \begin{bmatrix} -1.516 & 0.766 \\ 1.370 & 2.510 \end{bmatrix} \end{aligned}$$

### Definitions

1. Two random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  are *independent* if  $f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}}(\mathbf{y})$ .
2. Two random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  are *uncorrelated* if  $\mathbf{R}_{\mathbf{XY}} = \boldsymbol{\mu}_{\mathbf{X}} \boldsymbol{\mu}_{\mathbf{Y}}^T$ , or  $\mathbf{C}_{\mathbf{XY}} = \mathbf{0}$ .
3. Two random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  are *orthogonal* if  $\mathbf{R}_{\mathbf{XY}} = \mathbf{0}$ .

The definitions and properties of correlation and covariance matrices for two real random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  are shown in Table 17.2.1.

## 17.3 VECTOR GAUSSIAN RANDOM VARIABLES

Gaussian random variables play an important role in signal processing and communication problems. For a real  $n$ -vector random variable, the multivariate Gaussian density takes the form

$$f_{\mathbf{X}}(\mathbf{X}) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}_{\mathbf{X}}|^{1/2}} e^{-(1/2)(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T \mathbf{C}_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})} \quad (17.3.1)$$

TABLE 17.2.1

Property	Correlation functions	Covariance functions
<i>Autocorrelation and Autocovariance</i>		
Definition	$\mathbf{R}_X = E[\mathbf{XX}^T]$	$\mathbf{C}_X = E[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{X} - \boldsymbol{\mu}_X)^T]$
Symmetry	$\mathbf{R}_X = \mathbf{R}_X^T$	$\mathbf{C}_X = \mathbf{C}_X^T$
Definiteness	Positive definite	Positive definite
Eigenvalues	Greater than 0	Greater than 0
Eigenvectors	Orthogonal	Orthogonal
Principal diagonals	Greater than 0	Greater than 0
Relationship	$\mathbf{R}_X = \mathbf{C}_X + \boldsymbol{\mu}_X \boldsymbol{\mu}_X^T$	$\mathbf{C}_X = \mathbf{R}_X - \boldsymbol{\mu}_X \boldsymbol{\mu}_X^T$
<i>Cross-Correlation and Cross-Covariance</i>		
Definition	$\mathbf{R}_{XY} = E[\mathbf{XY}^T]$	$\mathbf{C}_{XY} = E[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{Y} - \boldsymbol{\mu}_Y)^T]$
Symmetry	$\mathbf{R}_{XY} = \mathbf{R}_{YX}^T$	$\mathbf{C}_{XY} = \mathbf{C}_{YX}^T$
Definiteness	Nondefinite	Nondefinite
Relationship	$\mathbf{R}_{XY} = \mathbf{C}_{XY} + \boldsymbol{\mu}_X \boldsymbol{\mu}_Y^T$	$\mathbf{C}_{XY} = \mathbf{R}_{XY} - \boldsymbol{\mu}_X \boldsymbol{\mu}_Y^T$
Uncorrelated	$\mathbf{R}_{XY} = \boldsymbol{\mu}_X \boldsymbol{\mu}_Y^T$	$\mathbf{C}_{XY} = \mathbf{0} = \mathbf{C}_{YX}$
Orthogonal	$\mathbf{R}_{XY} = \mathbf{0} = \mathbf{R}_{YX}$	
Sum of $\mathbf{X}$ and $\mathbf{Y}$	$\mathbf{R}_{X+Y} = \mathbf{R}_X + \mathbf{R}_{XY} + \mathbf{R}_{YX} + \mathbf{R}_Y$ $\mathbf{R}_{X+Y} = \mathbf{R}_X + \mathbf{R}_Y$ if $\mathbf{X}, \mathbf{Y}$ are orthogonal	$\mathbf{C}_{X+Y} = \mathbf{C}_X + \mathbf{C}_{XY} + \mathbf{C}_{YX} + \mathbf{C}_Y$ $\mathbf{C}_{X+Y} = \mathbf{C}_X + \mathbf{C}_Y$ if $\mathbf{X}, \mathbf{Y}$ are uncorrelated

The joint density of two random variables  $X$  and  $Y$  is given by a bivariate Gaussian as in Eq. (10.5.1)

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[ \left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 \right] \right\} \quad (10.5.1)$$

where  $\rho$  is the correlation coefficient between random variables  $X$  and  $Y$  given by  $\rho = \sigma_{xy}/(\sigma_x\sigma_y)$ . This can also be expressed in the matrix form of Eq. (17.3.1) with  $n = 2$

$$f_{XY}(x,y) = \frac{1}{2\pi\sqrt{|\mathbf{C}_X|}} \exp\left(-\frac{1}{2} \left\{ [x - \mu_x \ y - \mu_y] \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}^{-1} \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix} \right\}\right) \quad (17.3.2)$$

where the covariance matrix  $\mathbf{C}_X$ , its inverse  $\mathbf{C}_X^{-1}$ , and the mean vector  $\boldsymbol{\mu}$  are given by

$$\mathbf{C}_X = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}, \quad \mathbf{C}_X^{-1} = \begin{bmatrix} \frac{1}{\sigma_X^2(1-\rho^2)} & \frac{-\rho}{\sigma_X\sigma_Y(1-\rho^2)} \\ \frac{-\rho}{\sigma_X\sigma_Y(1-\rho^2)} & \frac{1}{\sigma_Y^2(1-\rho^2)} \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} \quad (17.3.3)$$

In a similar fashion, when  $n = 1$ ,  $\mathbf{C}_X = \sigma_X^2$ ,  $\mathbf{C}_X^{-1} = 1/\sigma_X^2$ ,  $\rho = 0$ , and the density for a single-dimensional random variable  $X$  from Eq. (17.3.1) is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-(1/2)[(x-\mu_X)/\sigma_X]^2} \quad (17.3.4)$$

a result that has already been stated in Eq. (6.4.1).

If random  $n$ -vector  $\mathbf{X}$  and  $m$ -vector  $\mathbf{Y}$  are Gaussian-distributed, then the joint density  $f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$  can be obtained by forming an  $(n+m)$  vector

$$\mathbf{Z} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \quad (17.3.5)$$

with the corresponding mean vector and the covariance matrix given by

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{Z}} &= E\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{X}} \\ \boldsymbol{\mu}_{\mathbf{Y}} \end{bmatrix} \\ \mathbf{C}_{\mathbf{Z}} &= E\begin{bmatrix} \mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}} \\ \mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}} \end{bmatrix}[(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})^T(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}})^T] = \begin{bmatrix} \mathbf{C}_X & \mathbf{C}_{XY} \\ \mathbf{C}_{YX} & \mathbf{C}_Y \end{bmatrix} \end{aligned} \quad (17.3.6)$$

Since  $\mathbf{C}_{\mathbf{Z}}$  is block symmetric, we note that  $\mathbf{C}_{YX}^T = \mathbf{C}_{XY}$ .

The joint density  $f_{\mathbf{Z}}(\mathbf{z}) = f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$  can be obtained from Eq. (17.3.1) as

$$f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^{(n+m)/2} |\mathbf{C}_{\mathbf{Z}}|^{1/2}} e^{-(1/2)(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{Z}})^T \mathbf{C}_{\mathbf{Z}}^{-1} (\mathbf{z} - \boldsymbol{\mu}_{\mathbf{Z}})} \quad (17.3.7)$$

where the determinant  $|\mathbf{C}_{\mathbf{Z}}|$  is obtained from Eq. (16.4.3)

$$|\mathbf{C}_{\mathbf{Z}}| = |\mathbf{C}_X| \cdot |\mathbf{C}_Y - \mathbf{C}_{YX} \mathbf{C}_X^{-1} \mathbf{C}_{XY}| \quad (17.3.8)$$

and the inverse  $\mathbf{C}_{\mathbf{Z}}^{-1}$ , from Eq. (16.4.6)

$$\begin{aligned} \mathbf{C}_{\mathbf{Z}}^{-1} &= \begin{bmatrix} \mathbf{C}_X & \mathbf{C}_{XY} \\ \mathbf{C}_{YX} & \mathbf{C}_Y \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{C}_X^{-1} + \mathbf{C}_X^{-1} \mathbf{C}_{XY} \Delta_{\mathbf{C}_Y}^{-1} \mathbf{C}_{YX} \mathbf{C}_X^{-1} & -\mathbf{C}_X^{-1} \mathbf{C}_{XY} \Delta_{\mathbf{C}_Y}^{-1} \\ -\Delta_{\mathbf{C}_Y}^{-1} \mathbf{C}_{YX} \mathbf{C}_X^{-1} & \Delta_{\mathbf{C}_Y}^{-1} \end{bmatrix} \end{aligned} \quad (17.3.9)$$

where the inverse of the Schur component  $\Delta_{\mathbf{C}_Y}^{-1}$  of  $\mathbf{C}_Y$  is obtained from Eq. (16.4.7):

$$\Delta_{\mathbf{C}_Y}^{-1} = (\mathbf{C}_Y - \mathbf{C}_{YX} \mathbf{C}_X^{-1} \mathbf{C}_{XY})^{-1}$$

It can be seen from Eq. (17.3.7) that when the densities  $f_{\mathbf{X}}(\mathbf{x})$  and  $f_{\mathbf{Y}}(\mathbf{y})$  are jointly Gaussian, the density  $f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$  is also jointly Gaussian.

The conditional density  $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$  is obtained from  $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) / f_{\mathbf{X}}(\mathbf{x})$ . Performing the indicated division and after some involved matrix manipulation, we obtain

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{m/2} |\mathbf{C}_{\mathbf{Y}|\mathbf{X}}|^{1/2}} e^{-(1/2)(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}})^T \mathbf{C}_{\mathbf{Y}|\mathbf{X}}^{-1} (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}})} \quad (17.3.10)$$

where  $\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} = \boldsymbol{\mu}_{\mathbf{Y}} + \mathbf{C}_{\mathbf{YX}} \mathbf{C}_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})$  and  $\mathbf{C}_{\mathbf{Y}|\mathbf{X}} = \mathbf{C}_{\mathbf{Y}} - \mathbf{C}_{\mathbf{YX}} \mathbf{C}_{\mathbf{X}}^{-1} \mathbf{C}_{\mathbf{XY}}$ .

We note that the conditional density is also Gaussian.

**Example 17.3.1** Two-dimensional zero mean Gaussian random vectors  $\mathbf{X}$  and  $\mathbf{Y}$  have covariance matrices equal to

$$\mathbf{C}_{\mathbf{X}} = \begin{bmatrix} 1 & 0 \\ 0 & 9 \end{bmatrix} \quad \text{and} \quad \mathbf{C}_{\mathbf{Y}} = \begin{bmatrix} 4 & 0 \\ 0 & 16 \end{bmatrix}$$

The cross-covariance matrix between  $\mathbf{X}$  and  $\mathbf{Y}$  is given by

$$\mathbf{C}_{\mathbf{XY}} = \begin{bmatrix} C_{X_1 Y_1} & C_{X_1 Y_2} \\ C_{X_2 Y_1} & C_{X_2 Y_2} \end{bmatrix} = \begin{bmatrix} 1 & \frac{6}{5} \\ \frac{18}{5} & \frac{12}{5} \end{bmatrix} = \mathbf{C}_{\mathbf{YX}}^T$$

We have to find the joint density  $f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$  and the conditional density  $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$  using Eqs. (17.3.7) and (17.3.10).

### **Joint Density**

From the data given, we can form the matrix

$$\mathbf{C}_{\mathbf{Z}} = \begin{bmatrix} \mathbf{C}_{\mathbf{X}} & \mathbf{C}_{\mathbf{XY}} \\ \mathbf{C}_{\mathbf{YX}} & \mathbf{C}_{\mathbf{Y}} \end{bmatrix} \quad \text{with} \quad \mathbf{C}_{\mathbf{X}} = \begin{bmatrix} 1 & 0 \\ 0 & 9 \end{bmatrix}, \quad \mathbf{C}_{\mathbf{Y}} = \begin{bmatrix} 4 & 0 \\ 0 & 16 \end{bmatrix}$$

and the cross-covariance matrices  $\mathbf{C}_{\mathbf{XY}}$  and  $\mathbf{C}_{\mathbf{YX}}$ , given by

$$\mathbf{C}_{\mathbf{XY}} = \begin{bmatrix} 1 & \frac{6}{5} \\ \frac{18}{5} & \frac{12}{5} \end{bmatrix} \quad \text{and} \quad \mathbf{C}_{\mathbf{YX}} = \mathbf{C}_{\mathbf{XY}}^T = \begin{bmatrix} 1 & \frac{18}{5} \\ \frac{6}{5} & \frac{12}{5} \end{bmatrix}$$

The mean vectors  $\boldsymbol{\mu}_{\mathbf{X}}$  and  $\boldsymbol{\mu}_{\mathbf{Y}}$  are zero vectors. From Eq. (17.3.8),  $|\mathbf{C}_{\mathbf{Z}}|$  and  $\sqrt{|\mathbf{C}_{\mathbf{Z}}|}$  are

$$|\mathbf{C}_{\mathbf{Z}}| = \frac{95904}{625} = 153.446 \quad \text{and} \quad \sqrt{|\mathbf{C}_{\mathbf{Z}}|} = \frac{36}{25} \sqrt{74} = 12.387$$

The inverse matrix  $\mathbf{C}_Z^{-1}$  is obtained from Eq. (17.3.9):

$$\begin{aligned}
 \mathbf{C}_Z^{-1} &= \left[ \begin{array}{c|c} \mathbf{C}_X^{-1} + \mathbf{C}_X^{-1} \mathbf{C}_{XY} \Delta_{CY}^{-1} \mathbf{C}_{YX} \mathbf{C}_X^{-1} & -\mathbf{C}_X^{-1} \mathbf{C}_{XY} \Delta_{CY}^{-1} \\ \hline -\Delta_{CY}^{-1} \mathbf{C}_{YX} \mathbf{C}_X^{-1} & \Delta_{CY}^{-1} \end{array} \right] \\
 &= \left[ \begin{array}{c|c} \begin{array}{cc} \frac{250}{111} & \frac{50}{111} \\ \hline \frac{50}{111} & \frac{275}{999} \end{array} & \begin{array}{cc} \frac{-215}{222} & \frac{-35}{148} \\ \hline \frac{-40}{111} & \frac{-25}{333} \end{array} \\ \hline \begin{array}{cc} \frac{-215}{222} & \frac{-40}{111} \\ \hline \frac{725}{888} & \frac{75}{592} \end{array} & \begin{array}{cc} \frac{75}{592} & \frac{325}{3552} \\ \hline \frac{148}{333} & \frac{3552}{3552} \end{array} \end{array} \right] \\
 &= \left[ \begin{array}{cccc} 2.252 & 0.45 & -0.968 & -0.236 \\ 0.45 & 0.275 & -0.36 & -0.075 \\ -0.968 & -0.36 & 0.816 & 0.127 \\ -0.236 & -0.075 & 0.127 & 0.091 \end{array} \right]
 \end{aligned}$$

With this information, the joint density  $f_{XY}(\mathbf{x}, \mathbf{y})$  is given by

$$\begin{aligned}
 f_{XY}(\mathbf{x}, \mathbf{y}) &= \frac{1}{(2\pi)^2 \sqrt{153.446}} \\
 &\times \exp \left( -\frac{1}{2} \left\{ [\mathbf{x} \mid \mathbf{y}] \left[ \begin{array}{c|c} 2.252 & 0.45 \\ \hline 0.45 & 0.275 \end{array} \right. \right. \right. \\
 &\quad \left. \left. \left. \left. \begin{array}{cc} -0.968 & -0.236 \\ -0.36 & -0.075 \end{array} \right] \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \right\} \right)
 \end{aligned}$$

or

$$\begin{aligned}
 f_{XY}(\mathbf{x}, \mathbf{y}) &= \frac{1}{(2\pi)^2 \sqrt{153.446}} \\
 &\times \exp \left[ \begin{array}{c} -0.408y_1^2 + (-0.127y_2 + 0.968x_1 + 0.36x_2)y_1 \\ -0.045y_2^2 + (0.236x_1 + 0.075x_2)y_2 - 1.126x_1^2 \\ -0.45x_1x_2 - 0.1375x_2^2 \end{array} \right]
 \end{aligned}$$

*Conditional Density.* We will now calculate the conditional density  $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$  from Eq. (17.3.10). The conditional expectation vector  $\mu_{\mathbf{Y}|\mathbf{X}}$  is given by

$$\begin{aligned}\mu_{\mathbf{Y}|\mathbf{X}} &= \mu_{\mathbf{Y}} + \mathbf{C}_{\mathbf{XY}}^T \mathbf{C}_{\mathbf{X}}^{-1} (\mathbf{x} - \mu_{\mathbf{X}}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &+ \begin{bmatrix} 1 & \frac{18}{5} \\ \frac{6}{5} & \frac{12}{5} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{9} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + \frac{2x_2}{5} \\ \frac{6x_1}{5} + \frac{4x_2}{15} \end{bmatrix}\end{aligned}$$

The conditional covariance matrix  $\mathbf{C}_{\mathbf{Y}|\mathbf{X}}$  is obtained as follows:

$$\begin{aligned}\mathbf{C}_{\mathbf{Y}|\mathbf{X}} &= \mathbf{C}_{\mathbf{Y}} - \mathbf{C}_{\mathbf{XY}}^T \mathbf{C}_{\mathbf{X}}^{-1} \mathbf{C}_{\mathbf{XY}} = \begin{bmatrix} 4 & 0 \\ 0 & 16 \end{bmatrix} - \begin{bmatrix} 1 & \frac{18}{5} \\ \frac{6}{5} & \frac{12}{5} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{9} \end{bmatrix} \begin{bmatrix} 1 & \frac{6}{5} \\ \frac{18}{5} & \frac{12}{5} \end{bmatrix} \\ &= \begin{bmatrix} \frac{39}{25} & \frac{-54}{25} \\ \frac{-54}{25} & \frac{348}{25} \end{bmatrix}\end{aligned}$$

The determinant  $|\mathbf{C}_{\mathbf{Y}|\mathbf{X}}| = \frac{10656}{625} = 17.05$  and  $\sqrt{|\mathbf{C}_{\mathbf{Y}|\mathbf{X}}|} = 4.129$  and

$$\mathbf{C}_{\mathbf{Y}|\mathbf{X}}^{-1} = \begin{bmatrix} \frac{39}{25} & \frac{-54}{25} \\ \frac{-54}{25} & \frac{348}{25} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{725}{888} & \frac{75}{592} \\ \frac{75}{592} & \frac{325}{3552} \end{bmatrix} = \begin{bmatrix} 0.816 & 0.127 \\ 0.127 & 0.091 \end{bmatrix}$$

Substituting these values in Eq. (17.3.10), we obtain

$$\begin{aligned}f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) &= \frac{1}{2\pi\sqrt{17.05}} \exp\left(-\frac{1}{2} \left\{ \begin{bmatrix} y_1 - x_1 - \frac{2x_2}{5} & y_1 - \frac{6x_1}{5} - \frac{4x_2}{15} \\ 0.816 & 0.127 \\ 0.127 & 0.091 \end{bmatrix} \begin{bmatrix} y_1 - x_1 - \frac{2x_2}{5} \\ y_1 - \frac{6x_1}{5} - \frac{4x_2}{15} \end{bmatrix} \right\} \right) \\ &\times \begin{bmatrix} 0.816 & 0.127 \\ 0.127 & 0.091 \end{bmatrix} \begin{bmatrix} y_1 - x_1 - \frac{2x_2}{5} \\ y_1 - \frac{6x_1}{5} - \frac{4x_2}{15} \end{bmatrix}\end{aligned}$$

Or expanding this equation results in the conditional density  $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$  as follows:

$$\begin{aligned}f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) &= \frac{1}{2\pi\sqrt{17.05}} \\ &\times \exp \left[ \begin{aligned} &-0.408y_1^2 + (-0.127y_2 + 0.968x_1 + 0.36x_2)y_1 - 0.045y_2^2 \\ &+(0.236x_1 + 0.075x_2)y_2 - 0.626x_1^2 - 0.45x_1x_2 - 0.082x_2^2 \end{aligned} \right]\end{aligned}$$

### Linear Transformations

In Section 13.5 we discussed the general case of the transformation  $\mathbf{Z} = \mathbf{g}(\mathbf{X})$ , where  $\mathbf{Z}$  and  $\mathbf{X}$  are random  $n$ -vectors and  $\mathbf{g}$  is a function  $n$ -vector. We will now discuss a linear

transformation of the form

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{AX} \quad (17.3.11)$$

where  $\mathbf{X}$  is a Gaussian random  $n$ -vector and  $\mathbf{A}$  is a square nonsingular  $n$ -matrix  $\{a_{ij}\}$ . The density function for  $\mathbf{X}$  is given in Eq. (17.3.1):

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}_X|^{1/2}} e^{-(1/2)(\mathbf{x} - \boldsymbol{\mu}_X)^T \mathbf{C}_X^{-1} (\mathbf{x} - \boldsymbol{\mu}_X)} \quad (17.3.1)$$

Using the transformation techniques developed in Chapter 13, the solution for  $\mathbf{X}$  is given by  $\mathbf{X} = \mathbf{A}^{-1}\mathbf{Y}$ . The Jacobian determinant  $\|J\| = |\mathbf{A}|$ . The density function  $f_{\mathbf{Y}}(\mathbf{y})$  is given by

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \frac{1}{(2\pi)^{n/2} |\mathbf{A}| |\mathbf{C}_X|^{1/2}} e^{-(1/2)(\mathbf{A}^{-1}\mathbf{y} - \boldsymbol{\mu}_X)^T \mathbf{C}_X^{-1} (\mathbf{A}^{-1}\mathbf{y} - \boldsymbol{\mu}_X)} \\ &= \frac{1}{(2\pi)^{n/2} [|\mathbf{A}| |\mathbf{C}_X| |\mathbf{A}^T|]^{1/2}} e^{-(1/2)(\mathbf{y} - \mathbf{A}\boldsymbol{\mu}_X)^T (\mathbf{A}^{-T} \mathbf{C}_X^{-1} \mathbf{A}^{-1}) (\mathbf{y} - \mathbf{A}\boldsymbol{\mu}_X)} \\ &= \frac{1}{(2\pi)^{n/2} [|\mathbf{A}| |\mathbf{C}_X| |\mathbf{A}^T|]^{1/2}} e^{-(1/2)(\mathbf{y} - \mathbf{A}\boldsymbol{\mu}_X)^T (\mathbf{A} \mathbf{C}_X \mathbf{A}^T)^{-1} (\mathbf{y} - \mathbf{A}\boldsymbol{\mu}_X)} \end{aligned} \quad (17.3.12)$$

Taking the expectation and outer product of Eq. (17.3.11), we obtain

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{Y}} &= \mathbf{A}\boldsymbol{\mu}_{\mathbf{X}} \\ E[\mathbf{Y}\mathbf{Y}^T] &= \mathbf{C}_Y = E[\mathbf{A}\mathbf{X}\mathbf{X}^T \mathbf{A}^T] = \mathbf{A}E[\mathbf{X}\mathbf{X}^T]\mathbf{A}^T = \mathbf{A}\mathbf{C}_X \mathbf{A}^T \end{aligned} \quad (17.3.13)$$

Substituting Eqs. (17.3.13) into Eq. (17.3.12), we can write  $f_{\mathbf{Y}}(\mathbf{y})$  as

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{n/2} [|\mathbf{C}_Y|]^{1/2}} e^{-(1/2)(\mathbf{y} - \boldsymbol{\mu}_Y)^T \mathbf{C}_Y^{-1} (\mathbf{y} - \boldsymbol{\mu}_Y)} \quad (17.3.14)$$

**Example 17.3.2** Starting with the same example (Example 13.5.1), we will solve for  $f_{\mathbf{Y}}(\mathbf{y})$  using matrix methods. A zero mean Gaussian random 3-vector  $\mathbf{X}$  has the density function  $f_{\mathbf{X}}(\mathbf{x})$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}} e^{-(1/2)\mathbf{x}^T \mathbf{C}_X^{-1} \mathbf{x}}$$

where

$$\mathbf{C}_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad |\mathbf{C}_X| = 1, \quad \mathbf{C}_X^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The transformation  $\mathbf{Y} = \mathbf{AX}$  is given by

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 1 \\ 5 & 3 & 3 \\ 3 & 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{AX}$$

Taking expectation on both sides of the preceding equation we have  $\mu_{\mathbf{Y}} = \mathbf{0}$ , since  $\mu_{\mathbf{X}} = \mathbf{0}$ . The covariance matrix  $\mathbf{C}_{\mathbf{Y}}$  is given by

$$\mathbf{C}_{\mathbf{Y}} = \mathbf{AC}_{\mathbf{X}}\mathbf{A}^T = \begin{bmatrix} 2 & 2 & 1 \\ 5 & 3 & 3 \\ 3 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 5 & 3 \\ 2 & 3 & 2 \\ 1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 9 & 19 & 12 \\ 19 & 43 & 27 \\ 12 & 27 & 17 \end{bmatrix}$$

The inverse of  $\mathbf{C}_{\mathbf{Y}}$  is

$$\mathbf{C}_{\mathbf{Y}}^{-1} = \begin{bmatrix} 2 & 1 & -3 \\ 1 & 9 & -15 \\ 13 & -15 & 26 \end{bmatrix}$$

and the determinant of  $\mathbf{C}_{\mathbf{Y}} = 1$ . Hence  $f_{\mathbf{Y}}(\mathbf{y})$  is obtained from Eq. (17.3.14)

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{3/2} \cdot 1} \exp \left\{ -\frac{1}{2} \mathbf{y}^T \begin{bmatrix} 2 & 1 & -3 \\ 1 & 9 & -15 \\ 13 & -15 & 26 \end{bmatrix} \mathbf{y} \right\}$$

or

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{3/2}} e^{-(1/2)[2y_1^2 + 9y_2^2 + 26y_3^2 + 2y_1y_2 - 30y_2y_3 - 6y_3y_1]}$$

which is exactly the same as obtained in Example 13.5.1.

## 17.4 DIAGONALIZATION OF COVARIANCE MATRICES

### *Diagonalization Principles*

In Section 16.2 we discussed diagonalization of a nonsingular square matrix by another transformation matrix of eigenvectors called the *modal matrix*. Since covariance matrices are symmetric and positive definite, the modal matrix is orthogonal. This property makes diagonalization of covariance matrices somewhat simpler. Diagonalization is needed to make a problem simpler, and correlated Gaussian random variables become independent.

Given any positive definite covariance  $n$ -matrix  $\mathbf{C}$  with distinct eigenvalues, we find the eigenvalue  $n$ -matrix  $\mathbf{\Lambda}$  by solving the characteristic equation  $|\lambda\mathbf{I} - \mathbf{C}| = 0$  as outlined in Section 16.2.

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \quad (17.4.1)$$

We find the eigenvectors  $\{\Phi_i, i = 1, \dots, n\}$  corresponding to each of the eigenvalues  $\{\lambda_i, i = 1, \dots, n\}$  and form the modal matrix  $\mathbf{M}$  of eigenvectors:

$$\mathbf{M} = [\Phi_1 \ \Phi_2 \ \dots \ \Phi_n] = \begin{bmatrix} \phi_{11} & \phi_{21} & \dots & \phi_{n1} \\ \phi_{12} & \phi_{22} & \dots & \phi_{n2} \\ \vdots & \vdots & & \vdots \\ \phi_{1n} & \phi_{2n} & \dots & \phi_{nn} \end{bmatrix} \quad (17.4.2)$$

Since  $\mathbf{C}$  is symmetric,  $\mathbf{M}$  will be orthogonal:

$$\Phi_i^T \Phi_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad \text{and} \quad \mathbf{M}^T = \mathbf{M}^{-1} \quad (17.4.3)$$

With these definitions, the following eigenequation can be formed:

$$\mathbf{CM} = \mathbf{M}\Lambda \quad \text{or} \quad \mathbf{M}^{-1}\mathbf{CM} = \Lambda \quad \text{and} \quad \mathbf{M}^T\mathbf{CM} = \Lambda \quad (17.4.4)$$

The problem now is to find a transformation  $\mathbf{T}$  between two random  $n$ -vectors  $\mathbf{X}$  and  $\mathbf{Y}$  such that the covariance matrix  $\mathbf{C}_X$  is diagonalized to the covariance matrix  $\mathbf{C}_Y$ . We take the expected value and the covariance of  $\mathbf{Y}$  in the transformation

$$\mathbf{Y} = \mathbf{TX} \quad (17.4.5)$$

$$\boldsymbol{\mu}_Y = \mathbf{T}\boldsymbol{\mu}_X \quad (17.4.6)$$

Thus

$$E[(\mathbf{Y} - \boldsymbol{\mu}_Y)(\mathbf{Y} - \boldsymbol{\mu}_Y)^T] = E[\mathbf{T}(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{X} - \boldsymbol{\mu}_X)^T \mathbf{T}^T], \quad \text{or} \quad \mathbf{C}_Y = \mathbf{TC}_X\mathbf{T}^T \quad (17.4.7)$$

Or, identifying terms in Eqs. (17.4.4) and (17.4.7), we obtain

$$\mathbf{C}_Y \sim \Lambda, \quad \mathbf{C}_X \sim \mathbf{C}, \quad \mathbf{T} \sim \mathbf{M}^T \quad (17.4.8)$$

with the required transformation matrix  $\mathbf{T} \sim \mathbf{M}^T$ .

### Diagonalizing to an Identity Matrix

We can carry the diagonalization process one step further and find a transformation matrix  $\mathbf{D}$  such that  $\mathbf{C}_Y$  is an identity matrix. Since  $\Lambda$  is a diagonal matrix, we can take the positive square root and write  $\Lambda = (\sqrt{\Lambda})^T \mathbf{I}_n (\sqrt{\Lambda})$ , where  $\mathbf{I}_n$  is an identity  $n$ -matrix. Incorporating this result in Eq. (17.4.4) and noting that  $(\sqrt{\Lambda})^T = (\sqrt{\Lambda})$ , we obtain

$$\mathbf{M}^T \mathbf{CM} = (\sqrt{\Lambda}) \mathbf{I}_n (\sqrt{\Lambda}) \quad \text{or} \quad (\sqrt{\Lambda})^{-1} \mathbf{M}^T \mathbf{CM} (\sqrt{\Lambda})^{-1} = \mathbf{I}_n \quad (17.4.9)$$

Thus the required transformation matrix  $\mathbf{D}$  in the transformation  $\mathbf{D}^T \mathbf{CD} = \mathbf{I}_n$  is

$$\mathbf{D} = \mathbf{M} (\sqrt{\Lambda})^{-1} \quad (17.4.10)$$

The transformation  $\mathbf{T}$  is orthogonal, whereas  $\mathbf{D}$  is not.

**Example 17.4.1** The covariance matrix  $\mathbf{C}_X$  of a zero mean Gaussian random 3-vector is given by

$$\mathbf{C}_X = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{bmatrix}$$

We have to determine

1. The joint density  $f_{\mathbf{X}}(\mathbf{x})$
2. The correlation coefficient matrix  $\rho$ , given by

$$\rho = \begin{bmatrix} 1 & \rho_{X_1X_2} & \rho_{X_1X_3} \\ \rho_{X_2X_1} & 1 & \rho_{X_2X_3} \\ \rho_{X_3X_1} & \rho_{X_3X_2} & 1 \end{bmatrix}$$

3. The modal matrix  $\mathbf{M}$  and show that it is orthogonal
4. The transformation matrix  $\mathbf{T}$  in  $\mathbf{Y} = \mathbf{TX}$  that will make  $\mathbf{C}_Y$  diagonal
5. The transformation matrix  $\mathbf{D}$  that will make  $\mathbf{C}_X$  an identity matrix
6. The joint density  $f_{\mathbf{Y}}(\mathbf{y})$

1. With  $|\mathbf{C}_X| = 12$  and

$$\mathbf{C}_X^{-1} = \begin{bmatrix} \frac{5}{12} & -\frac{1}{4} & \frac{1}{12} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{4} \\ \frac{1}{12} & -\frac{1}{4} & \frac{5}{12} \end{bmatrix}$$

in the equation

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{3/2} \sqrt{|\mathbf{C}_X|}} e^{-(1/2)\mathbf{x}^T \mathbf{C}_X^{-1} \mathbf{x}}$$

the joint density is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{3/2} \sqrt{12}} \exp \left\{ -\frac{1}{2} \left[ \frac{5x_1^2}{12} - \left( \frac{-x_3}{6} + \frac{x_2}{2} \right) x_1 - \frac{x_2 x_3}{2} + \frac{3x_2^2}{4} + \frac{5x_3^2}{12} \right] \right\}$$

2. From the given covariance matrix  $\mathbf{C}_X$ , we have

$$\sigma_{X_1}^2 = 3, \quad \sigma_{X_2}^2 = 2, \quad \sigma_{X_3}^2 = 3 \quad \text{and} \quad \sigma_{X_1X_2} = 1, \quad \sigma_{X_1X_3} = 0, \quad \sigma_{X_2X_3} = 1$$

Hence the correlation matrix  $\rho$  is given by

$$\rho = \begin{bmatrix} 1 & \frac{1}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{6}} & 1 & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{6}} & 1 \end{bmatrix}$$

3. From the determinant of the matrix

$$|\lambda \mathbf{I} - \mathbf{C}_X| = \left| \begin{bmatrix} \lambda - 3 & -1 & 0 \\ -1 & \lambda - 2 & -1 \\ 0 & -1 & \lambda - 3 \end{bmatrix} \right| = 0$$

we obtain the characteristic equation,  $\lambda^3 - 8\lambda^2 + 19\lambda - 12 = (\lambda - 1)(\lambda - 3)(\lambda - 4) = 0$ , and the eigenvalues are  $\lambda_1 = 1$ ,  $\lambda_2 = 3$ ,  $\lambda_3 = 4$ . The eigenvectors corresponding to these eigenvalues are obtained by the methods described in Example 16.2.5 and are shown below with the modal matrix  $\mathbf{M}$ :

$$\Phi_1 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \Phi_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix},$$

$$\mathbf{M} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

By direct multiplication we can see that  $\Phi_1^T \Phi_2 = 0$ ,  $\Phi_1^T \Phi_3 = 0$ ,  $\Phi_2^T \Phi_3 = 0$ , and  $\Phi_1^T \Phi_1 = \Phi_2^T \Phi_2 = \Phi_3^T \Phi_3 = 1$ , thus showing orthogonality. Further,  $|\mathbf{M}| = 1$ .

4. From Eq. (14.7.8) the transformation matrix  $\mathbf{T}$  is

$$\mathbf{T} = \mathbf{M}^T = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \quad \text{and} \quad \mathbf{T} \mathbf{C}_X \mathbf{T}^T = \mathbf{C}_Y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

5. From Eq. (17.4.10) the transformation matrix  $\mathbf{D}$  that takes  $\mathbf{C}_X$  to an identity matrix  $\mathbf{I}$  is given by

$$\mathbf{D} = \mathbf{M} \left( \sqrt{\Lambda} \right)^{-1} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{4}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{2\sqrt{3}} \\ -\frac{\sqrt{2}}{\sqrt{3}} & 0 & \frac{1}{2\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & \frac{1}{2\sqrt{3}} \end{bmatrix}$$

6. Since  $\mathbf{C}_Y$  is a diagonal matrix, the random variables  $Y_1, Y_2, Y_3$  are independent, Gaussian, and zero mean. Hence the joint density is

$$f_Y(\mathbf{y}) = \frac{1}{(2\pi)^{3/2} \sqrt{12}} e^{-(1/2)[(y_1^2/1)+(y_2^2/3)+(y_3^2/4)]}$$

## 17.5 SIMULTANEOUS DIAGONALIZATION OF COVARIANCE MATRICES\*

We are given two positive definite covariance matrices  $\mathbf{A}$  and  $\mathbf{B}$  with distinct eigenvalues. In pattern recognition problems it will be very useful to diagonalize both matrices by means of the same transformation matrix  $\mathbf{T}$ . The transforming matrix  $\mathbf{T}$  will take  $\mathbf{A}$  to an identity matrix while at the same time diagonalizing  $\mathbf{B}$ . We shall now analyze the procedure that will accomplish this. The two covariance matrices  $\mathbf{A}$  and  $\mathbf{B}$  with distinct eigenvalues are

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} \quad (17.5.1)$$

We have seen in Section 16.2 how the modal matrix of eigenvectors  $\Phi = \{\Phi_1, \Phi_2, \dots, \Phi_n\}$  of  $\mathbf{A}$  will transform  $\mathbf{A}$  to a diagonal matrix of its eigenvalues  $\{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n\}$  as shown:

$$\Phi^T \mathbf{A} \Phi = \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \quad (17.5.2)$$

From Eq. (17.4.9) we can write

$$\Phi^T \mathbf{A} \Phi = (\sqrt{\Lambda}) \mathbf{I}_n (\sqrt{\Lambda}) \quad \text{or} \quad (\sqrt{\Lambda})^{-1} \Phi^T \mathbf{A} \Phi (\sqrt{\Lambda})^{-1} = \mathbf{I}_n \quad (17.5.3)$$

By defining  $\mathbf{U} = \Phi (\sqrt{\Lambda})^{-1}$ , we can write Eq. (17.5.3) as follows:

$$(\sqrt{\Lambda})^{-1} \Phi^T \mathbf{A} \Phi (\sqrt{\Lambda})^{-1} = \mathbf{U}^T \mathbf{A} \mathbf{U} = \mathbf{I}_n \quad (17.5.4)$$

This process is known as “whitening” the covariance matrix  $\mathbf{A}$ . Since  $\mathbf{B}$  is a covariance matrix, we can show that  $\mathbf{C} = \mathbf{U}^T \mathbf{B} \mathbf{U}$  is also a symmetric matrix as follows:

$$\begin{aligned} \mathbf{C}^T &= \left[ (\sqrt{\Lambda})^{-1} \Phi^T \mathbf{B} \Phi (\sqrt{\Lambda})^{-1} \right]^T = \left[ (\sqrt{\Lambda})^{-1} \Phi^T \mathbf{B}^T \Phi (\sqrt{\Lambda})^{-1} \right] \\ &= (\sqrt{\Lambda})^{-1} \Phi^T \mathbf{B} \Phi (\sqrt{\Lambda})^{-1} = \mathbf{C} \end{aligned} \quad (17.5.5)$$

The symmetric matrix  $\mathbf{C} = \mathbf{U}^T \mathbf{B} \mathbf{U}$  can therefore be diagonalized by its orthogonal modal matrix  $\mathbf{M}$ , that is,  $\mathbf{M}^T \mathbf{C} \mathbf{M} = \mathbf{M}^T \mathbf{U}^T \mathbf{B} \mathbf{U} \mathbf{M} = \mathbf{D} = \text{diag}\{d_1, d_2, \dots, d_n\}$ . The

\*This section is adapted, with permission, from Chapter 31 of the book *Linear Systems Properties—a Quick Reference*, by Venkatarama Krishnan, published by CRC Press, 1988.

transformation  $\mathbf{U}\mathbf{M}$  has diagonalized  $\mathbf{B}$ , and now we will show that  $\mathbf{U}\mathbf{M}$  applied to  $\mathbf{A}$  retains the whitening transformation. Using Eq. (17.5.4), we can write

$$\begin{aligned}\mathbf{M}^T \mathbf{U}^T \mathbf{A} \mathbf{U} \mathbf{M} &= \mathbf{M}^T \left( \sqrt{\mathbf{\Lambda}} \right)^{-1} \mathbf{\Phi}^T \mathbf{A} \mathbf{\Phi} \left( \sqrt{\mathbf{\Lambda}} \right)^{-1} \mathbf{M} = \mathbf{M}^T \mathbf{I}_n \mathbf{M} = \mathbf{M}^T \mathbf{M} \\ &= \mathbf{I}_n\end{aligned}\quad (17.5.6)$$

since  $\mathbf{M}$  is an orthogonal matrix  $\mathbf{M}^T \mathbf{M} = \mathbf{I}_n$ .

Thus, the required transformation matrix  $\mathbf{T}$  that takes  $\mathbf{A}$  to an identity matrix and  $\mathbf{B}$  to a diagonal matrix is given by

$$\mathbf{T} = \mathbf{\Phi} \left( \sqrt{\mathbf{\Lambda}} \right)^{-1} \mathbf{M} = \mathbf{U} \mathbf{M} \quad (17.5.7)$$

With this transformation, we have

$$\mathbf{T}^T \mathbf{A} \mathbf{T} = \mathbf{I}_n, \quad \mathbf{T}^T \mathbf{B} \mathbf{T} = \mathbf{D} \quad (17.5.8)$$

For computational purposes the procedure described above can be simplified as follows. From Eq. (17.5.8) we can write

$$\mathbf{A}\mathbf{T} = \mathbf{T}^{-T} \quad \text{and} \quad \mathbf{B}\mathbf{T} = \mathbf{T}^{-T} \mathbf{D} \quad (17.5.9)$$

Hence  $\mathbf{B}\mathbf{T} = \mathbf{A}\mathbf{T}\mathbf{D}$ , and we have the generalized eigenequation

$$\mathbf{A}^{-1} \mathbf{B}\mathbf{T} = \mathbf{T}\mathbf{D} \quad (17.5.10)$$

where the matrix  $\mathbf{T}$  serves as the modal matrix for the product matrix  $\mathbf{A}^{-1}\mathbf{B}$  with  $\mathbf{D}$  as its diagonal eigenvalue matrix. From Eq. (17.5.10) we can now formulate a procedure for finding the transformation matrix  $\mathbf{T}$  that diagonalizes simultaneously the covariance matrices  $\mathbf{A}$  and  $\mathbf{B}$ .

### Summary of Procedure for Simultaneous Diagonalization

1. Calculate the  $n$  distinct eigenvalues  $\{d_1, d_2, \dots, d_n\}$  of the matrix  $\mathbf{R} = \mathbf{A}^{-1}\mathbf{B}$  from  $\det(\lambda\mathbf{I} - \mathbf{R}) = 0$ .
2. Calculate the eigenvectors  $\tau_i$  of  $\mathbf{R}$  from the eigenequation  $(\lambda\mathbf{I} - \mathbf{R})\tau_i = 0$ .
3. Form the modal matrix  $\mathbf{N} = \{\tau_1, \tau_2, \dots, \tau_n\}$ .
4. Form the matrix  $\mathbf{N}^T \mathbf{A} \mathbf{N} = \mathbf{W}$  (normalization with respect to  $\mathbf{A}$ ).
5. Since  $\mathbf{W}$  is diagonal with positive values, step 4 can be rewritten, as  $(\sqrt{\mathbf{W}})^{-1} \mathbf{N}^T \mathbf{A} \mathbf{N} (\sqrt{\mathbf{W}})^{-1} = \mathbf{I}$  and the desired transformation is  $\mathbf{T} = \mathbf{N}(\sqrt{\mathbf{W}})^{-1}$ .

**Example 17.5.1** Two covariance matrices  $\mathbf{A}$  and  $\mathbf{B}$  are given by

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 5 & 2 \\ 2 & 2 & 6 \end{pmatrix}$$

There is no need to find the eigenvalues of  $\mathbf{A}$ , and for comparison, the eigenvalues of  $\mathbf{B}$  are

$$\mathbf{\Lambda}_B = \begin{bmatrix} 2.37 & 0 & 0 \\ 0 & 3.52 & 0 \\ 0 & 0 & 9.11 \end{bmatrix}$$

1. The matrix  $\mathbf{R} = \mathbf{A}^{-1}\mathbf{B}$  is given by

$$\mathbf{R} = \mathbf{A}^{-1}\mathbf{B} = \begin{bmatrix} 2 & 0.176 & 0.235 \\ 0 & 1.588 & 0.118 \\ 0 & 0.059 & 1.412 \end{bmatrix}$$

2. The characteristic equation of  $\mathbf{R}$  is  $(\lambda - 2)(\lambda^2 - 3\lambda + 2.235) = 0$ , and the eigenvalues are

$$\mathbf{D} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1.621 & 0 \\ 0 & 0 & 1.379 \end{bmatrix}$$

3. The modal matrix  $\mathbf{N}$  of eigenvectors of  $\mathbf{R}$  is

$$\mathbf{N} = \begin{bmatrix} 1 & -0.525 & -0.188 \\ 0 & 0.82 & -0.481 \\ 0 & 0.23 & 0.856 \end{bmatrix}$$

4. The matrix  $\mathbf{W} = \mathbf{N}^T \mathbf{A} \mathbf{N}$  is given by

$$\mathbf{W} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2.053 & 0 \\ 0 & 0 & 2.733 \end{bmatrix} \quad \text{and} \quad \sqrt{\mathbf{W}} = \begin{bmatrix} 1.414 & 0 & 0 \\ 0 & 1.433 & 0 \\ 0 & 0 & 1.653 \end{bmatrix}$$

5. The desired transformation  $\mathbf{T} = \mathbf{N}(\sqrt{\mathbf{W}})^{-1}$  is given by

$$\mathbf{T} = \mathbf{N}(\sqrt{\mathbf{W}})^{-1} = \begin{bmatrix} 0.707 & -0.366 & -0.114 \\ 0 & 0.572 & -0.291 \\ 0 & 0.161 & 0.518 \end{bmatrix}$$

### Comments

1. The matrix  $\mathbf{T}$  is not orthogonal since

$$\begin{aligned} \mathbf{T}^T &= \begin{bmatrix} 0.707 & 0 & 0 \\ -0.366 & 0.572 & -0.616 \\ -0.114 & -0.291 & 0.518 \end{bmatrix} \neq \mathbf{T}^{-1} \\ &= \begin{bmatrix} 1.414 & 0.707 & 0.707 \\ 0 & 1.51 & 0.848 \\ 0 & 0.161 & 0.518 \end{bmatrix} \quad \text{and} \quad |\mathbf{T}| = 0.243 \end{aligned}$$

2. We can check the diagonalization process:

$$\mathbf{T}^T \mathbf{A} \mathbf{T} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{T}^T \mathbf{B} \mathbf{T} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1.621 & 0 \\ 0 & 0 & 1.379 \end{bmatrix}$$

whereas  $\mathbf{\Lambda}_B = \begin{bmatrix} 2.37 & 0 & 0 \\ 0 & 3.52 & 0 \\ 0 & 0 & 9.11 \end{bmatrix}$

3. The diagonal elements in the transformation of  $\mathbf{B}$  are the eigenvalues of  $\mathbf{R}$  and not the eigenvalues  $\mathbf{\Lambda}_B$  of  $\mathbf{B}$ .

## 17.6 LINEAR ESTIMATION OF VECTOR VARIABLES

We will take up more detailed estimation problems in the next chapter. We are interested in estimating the vector  $\mathbf{X}$  given the equation

$$\mathbf{Y}_{m \times 1} = \mathbf{H}_{m \times n} \mathbf{X}_{n \times 1} + \mathbf{N}_{m \times 1} \quad (17.6.1)$$

where  $\mathbf{Y}$  is an  $m \times 1$  observation vector,  $\mathbf{H}$  is an  $m \times n$  system vector,  $\mathbf{X}$  is an  $n \times 1$  vector to be estimated, and  $\mathbf{N}$  is an  $m \times 1$  noise vector with zero mean and covariance matrix,  $\mathbf{C}_N = E[\mathbf{N}\mathbf{N}^T] = \sigma^2 \mathbf{I}$ . We need a linear estimator of the form

$$\hat{\mathbf{X}} = \mathbf{A} \mathbf{Y} \quad (17.6.2)$$

where  $\mathbf{A}$  is an  $n \times m$  matrix to be determined. We have to choose a scalar criterion function  $J$ , which is some measure of the estimation error  $\mathbf{\epsilon} = \mathbf{Y} - \mathbf{H}\hat{\mathbf{X}}$ , which can be minimized with respect to the vector  $\hat{\mathbf{X}}$ . It is reasonable to choose a sum-squared error function given by

$$J(\hat{\mathbf{X}}) = (\mathbf{Y} - \mathbf{H}\hat{\mathbf{X}})^T (\mathbf{Y} - \mathbf{H}\hat{\mathbf{X}}) \quad (17.6.3)$$

This is known as the *linear least-squares estimation*. Using Eq. (16.3.21), we can minimize Eq. (17.6.3) and write

$$\begin{aligned} \frac{\partial}{\partial \hat{\mathbf{X}}} J(\hat{\mathbf{X}}) &= \frac{\partial}{\partial \hat{\mathbf{X}}} \left[ (\mathbf{Y} - \mathbf{H}\hat{\mathbf{X}})^T (\mathbf{Y} - \mathbf{H}\hat{\mathbf{X}}) \right] \\ &= \frac{\partial}{\partial \hat{\mathbf{X}}} \left[ (\mathbf{Y}^T \mathbf{Y} - \hat{\mathbf{X}}^T \mathbf{H}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{H} \hat{\mathbf{X}} + \hat{\mathbf{X}}^T \mathbf{H}^T \mathbf{H} \hat{\mathbf{X}})^T \right] \\ &= \left[ 0 = 2\mathbf{H}^T \mathbf{Y} + 2\mathbf{H}^T \mathbf{H} \hat{\mathbf{X}} \right] = 0 \end{aligned} \quad (17.6.4)$$

or

$$\hat{\mathbf{X}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Y}$$

Thus, the desired  $n \times m$  matrix  $\mathbf{A}$  in Eq. (17.6.2) is given by

$$\mathbf{A} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \quad (17.6.5)$$

The quantity  $(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$  is called the *pseudoinverse*. We have essentially derived the estimator  $\hat{\mathbf{X}}$  deterministically without involving the noise covariance  $\mathbf{C}_N = \sigma^2 \mathbf{I}$ . In this case the estimation error  $\mathbf{\epsilon} = \mathbf{Y} - \mathbf{H}\hat{\mathbf{X}}$  is not random. We ask the question as to the meaning of this deterministic estimator  $\hat{\mathbf{X}}$ . If  $m > n$ , then we have an *overdetermined* system where the number of observations are more than the unknowns and the  $n \times n$  matrix  $\mathbf{H}^T \mathbf{H}$  is invertible. In this case among the many possible solutions,  $\hat{\mathbf{X}}$  is the one that gives a minimum mean-square solution. If  $m = n$ , then  $\mathbf{H}$  is a square matrix that may be invertible, in which case we have  $(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T = \mathbf{H}^{-1} \mathbf{H}^{-T} \mathbf{H}^T = \mathbf{H}^{-1}$  and  $\mathbf{X} = \mathbf{H}^{-1} \mathbf{Y}$  giving an *exact* solution. If  $m < n$ , then we have an *underdetermined* system where the number of observations is less than the number of unknowns and the  $n \times n$  matrix  $\mathbf{H}^T \mathbf{H}$  will be singular. The  $m$ -vector estimator  $\hat{\mathbf{X}}$  can be given only in terms of the  $n - m$  unknown quantities of the vector  $\mathbf{X}$ .

The presence of the noise vector  $\mathbf{N}$  makes the estimator  $\hat{\mathbf{X}}$  random that estimates the deterministic parameter vector  $\mathbf{X}$ , and the criterion function  $J$  is

$$J = E \left[ (\mathbf{Y} - \mathbf{H}\hat{\mathbf{X}})^T (\mathbf{Y} - \mathbf{H}\hat{\mathbf{X}}) \right] \quad (17.6.6)$$

and the estimator  $\hat{\mathbf{X}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Y}$  is in the minimum mean-square error sense.

By substituting  $\hat{\mathbf{X}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Y}$  in Eq. (17.6.3), we can find the least-square value as

$$\begin{aligned} J_{\min} &= (\mathbf{Y} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Y})^T (\mathbf{Y} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Y}) \\ &= \mathbf{Y}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{Y} \end{aligned} \quad (17.6.7)$$

since

$$(\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T)(\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) = \mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$$

or the matrix  $(\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T)$  is idempotent.

**Example 17.6.1** An observation system,  $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}$ , is given by

$$\begin{bmatrix} 3 \\ 5 \\ 10 \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 1 & 2 \\ \mathbf{H} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \mathbf{X} \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ \mathbf{N} \end{bmatrix}$$

We have to estimate in the least-squares sense the 2-vector  $\mathbf{X}$ . We also will find the noise sequence and the least squared error. From Eq. (17.6.4), we have

$$\begin{aligned} \mathbf{H}^T \mathbf{H} &= \begin{bmatrix} 1 & 2 & 1 \\ 1 & 4 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 6 & 11 \\ 11 & 6 \end{bmatrix} \\ (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T &= \begin{bmatrix} \frac{21}{5} & \frac{-11}{5} \\ \frac{-11}{5} & \frac{6}{5} \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 1 & 4 & 2 \end{bmatrix} = \begin{bmatrix} 2 & \frac{-2}{5} & \frac{-1}{5} \\ -1 & \frac{2}{5} & \frac{1}{5} \end{bmatrix} \end{aligned}$$

Hence the estimator  $\hat{\mathbf{X}}$  is given by

$$\hat{\mathbf{X}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Y} = \begin{bmatrix} 6 & 11 \\ 11 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 1 & 4 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 10 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

The error vector  $\mathbf{N}$  is given by

$$\mathbf{N} = \mathbf{Y} - \mathbf{H}\hat{\mathbf{X}} = \begin{bmatrix} 3 \\ 5 \\ 10 \end{bmatrix} - \begin{bmatrix} 3 \\ 8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ -3 \\ 6 \end{bmatrix}$$

and the least-squared error  $J_{\min}$  is obtained from Eq. (17.6.7) as

$$J_{\min} = \mathbf{Y}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{Y} = 45$$

# Estimation Theory

## 18.1 CRITERIA OF ESTIMATORS

*Estimation* is a process where we draw inferences about a set, random or otherwise, from knowledge of a suitably selected finite observation set. This inference is based on the minimization of some criterion function. The collection of elements in the entire set is called the *population*. The suitably selected observation set is called the *sample*. Any function of the elements of the sample is called an *estimator* or a *statistic*. Generally, estimators are random variables, but they can be constants under some circumstances. The criterion used for estimation in the previous chapter is the least-squares error. There are several other criteria for judging the proximity to the true value and the reasonableness of estimators. We shall enumerate some of these criteria assuming that  $\theta$  is a variable, random or otherwise, to be estimated by  $\hat{\theta}$  from a set of finite observations:

1. *Bias*. The bias  $B(\hat{\theta})$  of an estimator  $(\hat{\theta})$  is defined by

$$B(\hat{\theta}) = E[\hat{\theta} - \theta] \quad (18.1.1)$$

An estimator is said to be unbiased if  $B(\hat{\theta}) = 0$  and the pdf of the estimator is centered around  $\theta$ . Generally we would like to select an unbiased estimator. In some cases a biased estimator may be preferable to an unbiased one.

2. *Variance*. The variance of an estimator is defined by

$$\text{var}(\hat{\theta}) = \sigma_{\hat{\theta}}^2 = E\{\hat{\theta} - E[\hat{\theta}]\}^2 \quad (18.1.2)$$

Generally we would choose an estimator with minimum variance. However, this may not be compatible with a minimum bias.

3. *Mean-Square Error*. The mean-square error is defined by

$$\text{MSE}(\hat{\theta}) = E[\hat{\theta} - \theta]^2 \quad (18.1.3)$$

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= E\{\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta\}^2 \\ &= E\{(\hat{\theta} - E[\hat{\theta}])^2 + (E[\hat{\theta}] - \theta)^2 + (\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta)\} \\ &= \sigma_{\hat{\theta}}^2 + E\{(E[\hat{\theta}] - \theta)^2 + (\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta)\} \quad \text{if } \theta \text{ is a constant} \\ &= \sigma_{\hat{\theta}}^2 + B^2 + E\{(\hat{\theta} - E[\hat{\theta}])B\} = \sigma_{\hat{\theta}}^2 + B^2 \quad \text{if } \hat{\theta} \text{ is a constant} \end{aligned} \quad (18.1.4)$$

If the estimator is unbiased, then the minimum mean square is the same as minimum variance. However, in the presence of a bias we will be able to obtain a mean-square error smaller than the variance.

4. *Consistency.* Let  $\hat{\theta}_n$  be an estimator obtained from  $n$  observations. Then  $\hat{\theta}_n$  is a consistent estimator if

$$\lim_{n \rightarrow \infty} P[(\hat{\theta}_n - \theta) > \varepsilon] \rightarrow 0 \quad \text{for every } \varepsilon > 0 \quad (18.1.5)$$

or  $\hat{\theta}_n$  converges in probability to  $\theta$  as  $n \rightarrow \infty$ . We would prefer to have a consistent estimator; otherwise we will have a bias for infinite samples that may not be desirable.

5. *Maximum Likelihood.* In Example 4.6.2 we estimated the population  $N$  of tigers in a wildlife sanctuary by maximizing the probability of occurrence of  $N$  using the maximum-likelihood criterion. We will expand on this concept further. The maximum-likelihood estimate of the parameter  $\theta$  from a probability density function  $f(x, \theta)$  of a random variable  $X$  is obtained by maximizing the likelihood function formed from  $n$  independent observations of the population  $\{X_i = x_i, i = 1, \dots, n\}$ . The *likelihood function*  $L(\mathbf{X}, \theta)$  is the product of the density functions  $f(x_i, \theta)$ ,  $i = 1, \dots, n$ :

$$L(\mathbf{X}, \theta) = f(x_1, \theta) \cdots f(x_n, \theta) \quad (18.1.6)$$

Under certain regularity conditions the value of  $\hat{\theta}$  that maximizes Eq. (18.1.6) is called the *maximum-likelihood* (ML) estimator. The monotonicity of the logarithm allows us to maximize the *log likelihood* function  $\ln L(\mathbf{X}, \theta)$ . Thus the ML estimator is obtained from

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln[L(\mathbf{X}, \theta)] &= \frac{\partial}{\partial \theta} \ln [f(x_1, \theta) \cdots f(x_n, \theta)] = \frac{\partial}{\partial \theta} \sum_{i=1}^n \ln f(x_i, \theta) \\ &= \sum_{i=1}^n \frac{1}{f(x_i, \theta)} \frac{\partial f(x_i, \theta)}{\partial \theta} \end{aligned} \quad (18.1.7)$$

This method works well for large samples, but for small samples it may be biased.

6. *Efficiency.* The basic problem in estimation is to find the best estimator  $\hat{\theta}$ , and if it exists, it will be unique. However, in some problems finding the best estimator may not be simple. In such cases we are satisfied if we can find an estimator with the greatest lower bound. This bound is called the *Rao–Cramer bound* of an estimator, given by

$$\text{var}(\hat{\theta}) \geq \frac{1}{E \left\{ \left[ \frac{\partial}{\partial \theta} \ln f(\mathbf{X}, \theta) \right]^2 \right\}} \quad (18.1.8)$$

An unbiased estimate that satisfies the Rao–Cramer bound is called an *efficient* estimator.

## 18.2 ESTIMATION OF RANDOM VARIABLES

It is not possible for any estimator to satisfy all the criteria mentioned above. We will discuss unbiased minimum variance estimators under three broad categories:

1. Estimation of a random variable by a constant
2. Estimation of a random variable by a function of another random variable
3. Estimation of a constant by a random variable called *parameter estimation*

We will discuss the first two in this section and the third one in the next section.

### ***Estimation of a Random Variable by a Constant***

This is the simplest of the estimation problems. A random variable  $X$  is to be estimated by a constant  $a$  such that the mean-square error  $J = E[X - a]^2$  is minimized.

$$J = E[X - a]^2 = E[X^2 - 2Xa + a^2]$$

$$\frac{dJ}{dx} = \frac{d}{dx}[E[X^2] - 2\mu_X a + a^2] = -2\mu_X + 2a = 0$$

or

$$a = \mu_X. \quad (18.2.1)$$

Thus, the constant that minimizes the mean-square error is the expected value  $E[X]$ . The minimum mean-square error  $J_{\min} = E[X - \mu_X]^2 = \sigma_X^2$  is the variance. Equation (18.2.1) can be interpreted in a different way. Since

$$E[X - a]^2 = \int_{-\infty}^{\infty} (x - a)^2 f_X(x) dx \quad (18.2.2)$$

the value of  $a$  that minimizes the mean-square error is

$$\int_{-\infty}^{\infty} x f_X(x) dx = \mu_X \quad (18.2.3)$$

### ***Estimation of a Random Variable $Y$ by a Function of Another Random Variable $g(X)$***

Here we address the problem of estimating a random variable  $Y$  by a function of another random variable  $g(X)$  using the mean-square error criterion  $J = E[Y - g(X)]^2$ :

$$J = E[Y - g(X)]^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y - g(x)]^2 f_{XY}(x,y) dy dx \quad (18.2.4)$$

Inserting  $f_{XY}(x,y) = f_{Y|X}(y|x)f_X(x)$  in Eq. (18.2.4), we obtain

$$J = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} [y - g(x)]^2 f_{Y|X}(y|x) dy \right\} f_X(x) dx \quad (18.2.5)$$

Minimizing Eq. (18.2.5) amounts to minimizing the inner integral

$$\int_{-\infty}^{\infty} [y - g(x)]^2 f_{Y|X}(y|x) dy \quad (18.2.6)$$

From Eqs. (18.2.2) and (18.2.3), the value of  $g(x)$  that minimizes Eq. (18.2.6) is

$$g(x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$$

or

$$g(X) = \hat{Y} = E[Y|X] \quad (18.2.7)$$

Thus, the estimator  $\hat{Y}$  is the conditional expectation of  $Y$  given  $X$ . This estimator is unbiased because

$$E[\hat{Y}] = E[Y|X] = E[Y]$$

It is also minimum variance, and the minimum mean-square error is the conditional variance:

$$\sigma_{Y|X}^2 = E\{Y - E[Y|X]\}^2$$

In general, the conditional expectation  $E[Y|X]$  is a complicated nonlinear function of  $X$  that may not be expressible in an explicit form. We confine ourselves to a linear function of the form

$$E[Y|x] = g(x) = \hat{Y} = a_0 + a_1 x + \cdots + a_n x_n = \sum_{i=0}^n a_i x_i \quad (18.2.8)$$

This form is known as the *multiple linear regression*, and the coefficients  $\{a_i\}$  are called *linear regression coefficients*. It is linear because no powers of  $a_i$  are involved.

If  $E[Y|x]$  is of the form

$$E[Y|x] = g(x) = \hat{Y} = ax + b \quad (18.2.9)$$

then it is called *simple linear regression*. The equation for the regression line is

$$\hat{Y} = ax + b \quad (18.2.10)$$

The estimator  $\hat{Y}$  is unbiased because

$$E[\hat{Y}] = E[aX + b] = a\mu_X + \mu_Y - a\mu_X = E[Y]$$

It is also minimum variance.

### Evaluation of Regression Coefficients

We will now evaluate the linear regression coefficients  $a$  and  $b$  for the case where  $Y$  is regressed on  $X$ , using the minimum mean-square error

$$J = E[Y - aX - b]^2 \quad (18.2.11)$$

as the criterion function. We will first solve for  $b$  by differentiating  $J$  with respect to  $b$ :

$$\frac{\partial J}{\partial b} = -2E[Y - aX - b] = 0$$

Hence

$$b = \mu_Y - a\mu_X \quad (18.2.12)$$

We can now solve for  $a$  by substituting Eq. (18.2.12) into Eq.(18.2.11) and differentiating  $J$  with respect to  $a$ :

$$\begin{aligned} J &= E[Y - \mu_Y - a(X - \mu_X)]^2 \\ \frac{\partial J}{\partial a} &= -2E\{[Y - \mu_Y - a(X - \mu_X)](X - \mu_X)\} = 0 \\ &= E[(X - \mu_X)(Y - \mu_Y) - a(X - \mu_X)^2] = 0 \\ &= \sigma_{XY} - a\sigma_X^2 = 0 \end{aligned}$$

Hence

$$a = \frac{\sigma_{XY}}{\sigma_X^2} \quad \text{and substituting } \sigma_{XY} = \rho\sigma_X\sigma_Y, \quad a = \rho \frac{\sigma_Y}{\sigma_X} \quad (18.2.13)$$

The equations obtained by setting  $\partial J/\partial a = 0$  and  $\partial J/\partial b = 0$  are called *normal equations*. With these values of  $a$  and  $b$  the mean-square error can be found as follows:

$$\begin{aligned} J &= E[Y - \mu_Y - a(X - \mu_X)]^2 \\ &= \sigma_Y^2 - 2a\sigma_{XY} + a^2\sigma_X^2 = \sigma_Y^2 - 2\frac{\sigma_{XY}^2}{\sigma_X^2} + \sigma_{XY}^2 \\ &= \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} = \sigma_Y^2(1 - \rho^2) \end{aligned} \quad (18.2.14)$$

The simple linear regression plot shown in Fig.18.2.1 indicates the slope  $a$  and the offset  $b$ . The line of regression divides the cluster of points into two halves such that the mean square between the points and the line of regression is minimum.

In an analogous manner, we can also regress  $X$  on  $Y$  and find the corresponding regression coefficients  $\alpha$  and  $\beta$  in the expression

$$\hat{X} = \alpha Y + \beta \quad (18.2.15)$$

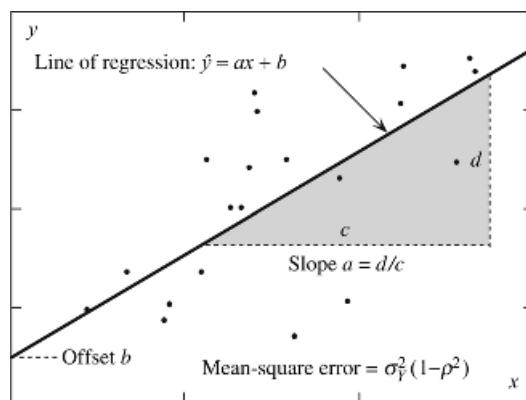


FIGURE 18.2.1

All the parameters such as the means, variances, and correlation coefficients will be the same. In much the same way as in Eqs. (18.2.12) and (18.2.13), the regression coefficients  $\alpha$  and  $\beta$  are given by

$$\alpha = \frac{\sigma_{XY}}{\sigma_Y^2} \quad \text{and} \quad \beta = \mu_X - \alpha \mu_Y \quad (18.2.16)$$

Substituting  $\sigma_{XY} = \rho \sigma_X \sigma_Y$  in Eq. (18.2.16),  $\alpha$  can also be given as

$$\alpha = \rho \frac{\sigma_X}{\sigma_Y} \quad (18.2.17)$$

Multiplying the two slopes  $a$  and  $\alpha$ , we obtain

$$a\alpha = \rho \frac{\sigma_Y}{\sigma_X} \rho \frac{\sigma_X}{\sigma_Y} = \rho^2 \quad (18.2.18)$$

Since the true means and variances are not known, Eq. (18.2.18) is a more accurate method for computing the correlation coefficient  $\rho$  in a real-life situation rather than using either Eq. (18.2.13) or (18.2.17). The sign of  $\rho$  can be obtained only from the physics of the problem.

The regression coefficients for a multiple linear regression problem

$$\hat{Y} = a_0 + a_1 X_1 + \cdots + a_n X_n \quad (18.2.19)$$

can be obtained in principle by forming

$$J = E \left[ Y - \sum_{i=0}^n a_i X_i \right]^2$$

and solving the  $(n+1)$  normal equations obtained from  $\partial J / \partial a_0 = 0, \partial J / \partial a_1 = 0, \dots, \partial J / \partial a_n = 0$ . A simpler method will be discussed in the next section.

**Example 18.2.1** The relationship between random variables  $X$  and  $Y$  is shown in Table 18.2.1. The parameters of  $X$  and  $Y$  are given by

$$\mu_X = 0.905, \sigma_X^2 = 0.198, \mu_Y = 1.939, \sigma_Y^2 = 0.308, \sigma_{XY} = 0.080.$$

TABLE 18.2.1

No.	X	Y	No.	X	Y
1	0.05	1.82	11	1.54	1.49
2	0.29	2.05	12	0.72	2.15
3	0.74	1.28	13	0.66	0.88
4	0.55	1.19	14	1.23	1.49
5	1.07	2.06	15	1.35	2.18
6	0.47	1.96	16	0.97	2.74
7	1.06	2.46	17	1.3	2.1
8	0.7	1.76	18	0.96	2.69
9	0.54	1.13	19	1.73	2.18
10	0.65	2.35	20	1.52	2.82

The correlation coefficient  $\rho_{XY} = \sigma_{XY}/\sigma_X\sigma_Y = 0.08/\sqrt{0.198 \times 0.308} = 0.324$ . We will find the regression coefficients  $a, b, \alpha, \beta$ :

$$a = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{0.08}{0.198} = 0.404, b = \mu_y - a\mu_x = 1.939 - 0.404 \times 0.905 = 1.574$$

$$\alpha = \frac{\sigma_{XY}}{\sigma_Y^2} = \frac{0.08}{0.308} = 0.26, \beta = \mu_x - \alpha\mu_y = 0.905 - 0.26 \times 1.939 = 0.40$$

The regression lines  $\hat{y} = ax + b$  and  $\hat{x} = \alpha y + \beta$  are shown in Figs. 18.2.2 and 18.2.3 respectively. The minimum mean-square error (MMSE) for  $Y$  on  $X$  regression  $= 0.308 \times (1 - 0.324^2) = 0.276$ , and the MMSE for  $X$  on  $Y$  regression  $= 0.194 \times (1 - 0.324^2) = 0.174$ .

The square root of the product  $\sqrt{a\alpha} = \sqrt{0.404 \times 0.26} = 0.324 = \rho_{XY}$  is the correlation coefficient obtained earlier. In general, the means and variances of random variables are not known a priori and we can only estimate them. We will revisit this example after discussing parameter estimation.

*Orthogonality Principle.* We will motivate the orthogonality principle by taking the following expectations

$$E[Y - aX - b]b \quad (18.2.20a)$$

$$E[(Y - aX - b)aX] \quad (18.2.20b)$$

with  $a = \sigma_{XY}/\sigma_X^2$  and  $b = \mu_y - a\mu_x$ . Equation (18.2.20a) becomes  $E[Y - aX - b]b = [\mu_y - a\mu_x - (\mu_y - a\mu_x)]b = 0$ . For Eq. (18.2.20b) we have

$$\begin{aligned} E[(Y - aX - b)aX] &= \{E[XY] - aE[X^2] - b\mu_x\}a \\ &= \{E[XY] - aE[X^2] - (\mu_y - a\mu_x)\mu_x\}a \\ &= \{\sigma_{XY} - a\sigma_X^2\}a = \{\sigma_{XY} - \sigma_{XY}\}a = 0 \end{aligned}$$

Combining both equations, we can write

$$E\left[\left(\begin{array}{c} Y - aX - b \\ \text{error } \varepsilon \end{array}\right) \left(\begin{array}{c} aX + b \\ \text{estimate } \hat{Y} \end{array}\right)\right] = 0$$

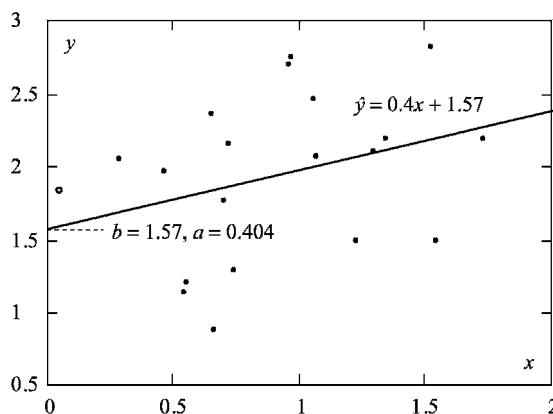


FIGURE 18.2.2

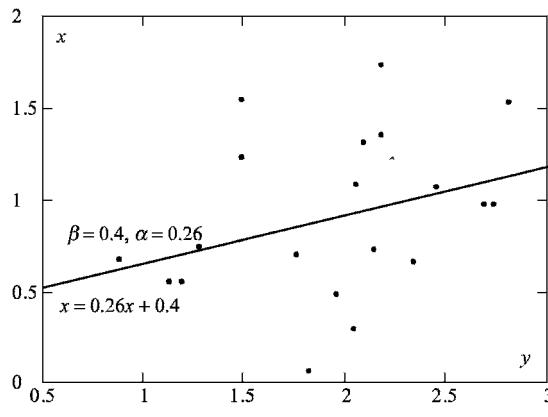


FIGURE 18.2.3

Equation (18.2.21) is the *orthogonality principle*, which states that under MMSE criterion the estimate  $\hat{Y}$  is orthogonal to the error  $\varepsilon$ . In other words

$$E[\varepsilon \hat{Y}] = 0 \quad (18.2.21)$$

This is a powerful result that is used extensively in estimation problems.

We can give a geometric interpretation to the orthogonality principle by means of Fig. 18.2.4. The two observations are  $b$  and  $a$ , represented by vectors  $x^0$  and  $x^1$  in a two-dimensional subspace. The unknown quantity  $y$  is not in the subspace. However, the estimate  $\hat{y}$ , which is a linear combination of  $x^0$  and  $x^1$  is in the subspace. The error vector  $\varepsilon$  is given by  $\varepsilon = y - \hat{y}$  is also shown in Fig. 18.2.4.

From the geometry of the diagram, it can be seen that  $\varepsilon$  is a minimum only when it is orthogonal to the subspace spanned by  $x^0 - x^1$ . Hence  $\varepsilon$  is orthogonal to the estimate  $\hat{y}$ .

We will use the orthogonality principle to find the regression coefficients for a multiple linear regression problem of Eq. (18.2.19) expressed in matrix form

$$\hat{Y} = a_0 + a_1 X_1 + \cdots + a_n X_n = \mathbf{a}^T \mathbf{X} \quad (18.2.19)$$

where  $\mathbf{a}$  is the  $(n + 1)$  column vector  $[a_0, a_1, \dots, a_n]^T$  and  $\mathbf{X}$  is the  $(n + 1)$  random variable vector  $[1, X_1, X_2, \dots, X_n]^T$ . Using the orthogonality principle that the error

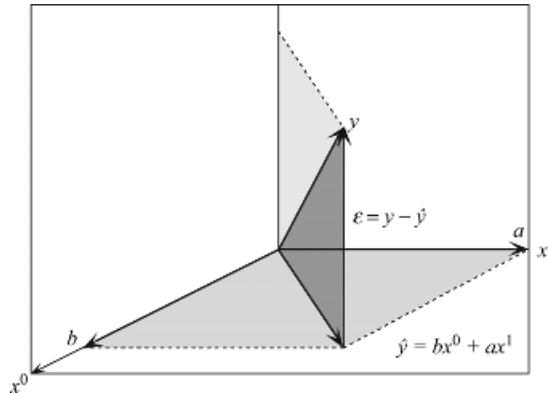


FIGURE 18.2.4

$\varepsilon = Y - \mathbf{a}^T \mathbf{X}$  is orthogonal to the subspace spanned by the vector  $\mathbf{X}$ , we have

$$E[(Y - \mathbf{a}^T \mathbf{X}) \mathbf{X}^T] = 0 \quad (18.2.22)$$

From Eq. (17.2.5) we can define  $E[\mathbf{X} \mathbf{X}^T] = \mathbf{R}_X$  and  $E[\mathbf{X} Y] = \mathbf{R}_{XY}$ , and Eq. (18.2.22) can be rewritten as

$$\begin{aligned} \mathbf{R}_X \mathbf{a} &= E[\mathbf{X} Y] \\ \mathbf{a} &= \mathbf{R}_X^{-1} E[\mathbf{X} Y] \end{aligned} \quad (18.2.23)$$

We can also find the minimum mean-square error using the orthogonality principle:

$$J = E[(Y - \mathbf{a}^T \mathbf{X})(Y - \mathbf{a}^T \mathbf{X})] = E[\varepsilon(Y - \mathbf{a}^T \mathbf{X})] \quad (18.2.24)$$

Using orthogonality, we obtain

$$\begin{aligned} J_{\min} &= E[\varepsilon Y] = E[(Y - \mathbf{a}^T \mathbf{X}) Y] \\ &= E[(Y^2) - \mathbf{a}^T E[\mathbf{X} Y]] \end{aligned} \quad (18.2.25)$$

a result arrived at with comparative ease.

**Example 18.2.2** We will now solve the simple linear regression problem by using Eq. (18.2.23). We are to estimate  $Y$  in  $Y = b + aX + \varepsilon$ . The estimator  $\hat{Y} = b + aX$ . The vector  $\mathbf{a} = [b, a]^T$ , and the vector  $\mathbf{X} = [1, X]^T$ . The correlation matrix  $\mathbf{R}_X$  is given by

$$\begin{aligned} \mathbf{R}_X &= E\left\{\begin{bmatrix} 1 \\ X \end{bmatrix} \begin{bmatrix} 1 & X \end{bmatrix}^T\right\} = \begin{bmatrix} 1 & \mu_X \\ \mu_X & m_2 \end{bmatrix} \\ E[\mathbf{X} Y] &= E\left\{\begin{bmatrix} 1 \\ X \end{bmatrix} Y\right\} = \begin{bmatrix} \mu_Y \\ E(XY) \end{bmatrix} \end{aligned}$$

Hence the regression coefficients from Eq. (18.2.23) are

$$\begin{aligned} \mathbf{a} &= \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 1 & \mu_X \\ \mu_X & m_2 \end{bmatrix}^{-1} \begin{bmatrix} \mu_Y \\ E(XY) \end{bmatrix} = \frac{1}{\sigma_X^2} \begin{bmatrix} m_2 & -\mu_X \\ -\mu_X & 1 \end{bmatrix} \begin{bmatrix} \mu_Y \\ E(XY) \end{bmatrix} \\ &= \frac{1}{\sigma_X^2} \begin{bmatrix} m_2 \mu_Y - \mu_X E(XY) \\ \sigma_{XY} \end{bmatrix} \\ &= \frac{1}{\sigma_X^2} \begin{bmatrix} \mu_Y(m_2 - \mu_X^2) - \mu_X(E(XY) - \mu_X \mu_Y) \\ \sigma_{XY} \end{bmatrix} \end{aligned}$$

or

$$\begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} \mu_Y - \mu_X a \\ \frac{\sigma_{XY}}{\sigma_X^2} \end{bmatrix}$$

which is the same result as obtained before but without differentiation.

We can also find the minimum mean-square error using the orthogonality principle:

$$\varepsilon^2 = E[(Y - aX - b)(Y - aX - b)] = E[\varepsilon(Y - aX - b)]$$

Using orthogonality, we obtain

$$\begin{aligned}\epsilon_{\min}^2 &= E[\epsilon Y] = E[(Y - aX - b)Y] \\ &= E[Y^2] - aE[XY] - \mu_Y[\mu_Y - a\mu_X] \\ &= E[Y^2] - \mu_y^2 - a\{E[XY] - \mu_x\mu_Y\} \\ &= \sigma_Y^2 - a\sigma_{XY} = \sigma_Y^2(1 - \rho^2)\end{aligned}$$

a result arrived at with comparative ease.

**Example 18.2.3** A random variable  $Y = aX^2 + bX + c + \epsilon$  is to be estimated by a linear regression line  $\hat{Y} = aX^2 + bX + c$ . We want to find the coefficients  $a, b, c$ . The column vector  $\mathbf{a} = [a, b, c]^T$ , and the column vector  $\mathbf{X} = [X^2, X, 1]^T$ . The correlation matrix  $\mathbf{R}_X$  is

$$\mathbf{R}_X = E \left\{ \begin{bmatrix} X^2 \\ X \\ 1 \end{bmatrix} \begin{bmatrix} X^2 & X & 1 \end{bmatrix} \right\} = \begin{bmatrix} m_4 & m_3 & m_2 \\ m_3 & m_2 & m \\ m_2 & m & 1 \end{bmatrix} \quad (18.2.26)$$

and

$$E[\mathbf{XY}] = E \left\{ \begin{bmatrix} X^2 \\ X \\ 1 \end{bmatrix} Y \right\} = \begin{bmatrix} E[X^2Y] \\ E[XY] \\ E[Y] \end{bmatrix} \quad (18.2.27)$$

Hence the regression coefficients are

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} m_4 & m_3 & m_2 \\ m_3 & m_2 & m \\ m_2 & m & 1 \end{bmatrix}^{-1} \begin{bmatrix} E[X^2Y] \\ E[XY] \\ E[Y] \end{bmatrix} \quad (18.2.28)$$

Using Eq. (18.2.25), we can derive the minimum mean-square error as follows:

$$\begin{aligned}\epsilon_{\min}^2 &= E[Y^2] - [a, b, c] \begin{bmatrix} E[X^2Y] \\ E[XY] \\ E[Y] \end{bmatrix} \\ &= E[Y^2] - \{aE[X^2Y] + bE[XY] + cE[Y]\}\end{aligned} \quad (18.2.29)$$

**Example 18.2.4** We will continue the previous example with observed data for random variables  $X$  and  $Y$  shown in Table 18.2.2. It is desired to estimate  $Y$  from  $X$ .

The correlation matrix  $\mathbf{R}_X$  is given as

$$\mathbf{R}_X = \begin{bmatrix} E[X^4] & E[X^3] & E[X^2] \\ E[X^3] & E[X^2] & E[X] \\ E[X^2] & E[X] & 1 \end{bmatrix} = \begin{bmatrix} 4.53 & 2.78 & 1.79 \\ 2.78 & 1.79 & 1.23 \\ 1.79 & 1.23 & 1 \end{bmatrix}$$

In addition, the cross-moments of  $X$  with  $Y$  are  $E[X^2Y] = 3.79$ ,  $E[XY] = 2.67$ , with  $E[Y] = 2.21$  and  $E[Y^2] = 4.92$ .

TABLE 18.2.2

No.	X	Y	No.	X	Y
1	0.91	2.6	12	1.47	2.25
2	0.22	2.21	13	1.37	2.28
3	0.23	2.2	14	0.63	2.26
4	1.92	2.03	15	1.39	2.23
5	1.87	2.07	16	0.79	2.27
6	0.55	2.38	17	1.53	2.11
7	0.94	2.48	18	1.51	2.1
8	1.4	2.37	19	1.35	2.17
9	1.06	2.44	20	1.86	1.77
10	1.38	2.34	21	1.64	1.95
11	1.91	1.9			

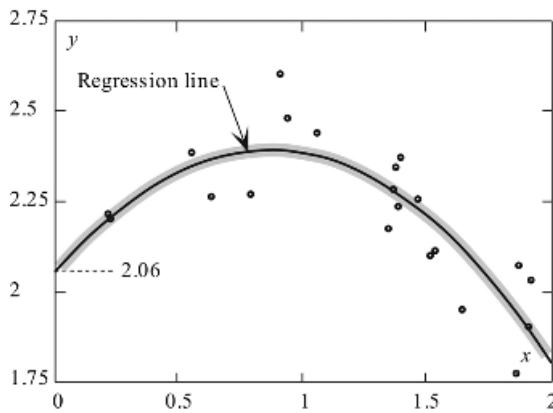


FIGURE 18.2.5

The regression coefficients are obtained from Eq. (18.2.28) as follows:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 4.53 & 2.78 & 1.79 \\ 2.78 & 1.79 & 1.23 \\ 1.79 & 1.23 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3.79 \\ 2.67 \\ 2.21 \end{bmatrix} = \begin{bmatrix} -0.46 \\ 0.79 \\ 2.06 \end{bmatrix}$$

The line of regression is  $y = -0.46x^2 + 0.79x + 2.06$ . The original data  $y$  and the line of regression are shown in Fig. 18.2.5.

The minimum mean-square error is obtained from Eq. (18.2.29) and is given by

$$\epsilon_{\min}^2 = E[Y^2] - \{aE[X^2Y] + bE[XY] + cE[Y]\} = 0.0109$$

### 18.3 ESTIMATION OF PARAMETERS (POINT ESTIMATION)

A population is represented by the random variable  $X$  characterized by a parameter  $\theta$ , which is an unknown constant, and whose distribution function  $F_X(x, \theta)$  is of a known form. We estimate this parameter by observing  $n$  realizations  $\{x_1, \dots, x_n\}$  of the

random variable  $X$ . Associated with each of these realizations is a set of random variables  $\{X_1, \dots, X_n\}$  having the same characteristics as the population. We form a function  $g(X_1, \dots, X_n)$  for the estimator  $\hat{\theta}$ . If  $\mathbf{X}$  is the  $n$ -vector equal to  $[X_1, \dots, X_n]^T$ , then  $\hat{\theta} = g(\mathbf{X})$  is known as the *point estimator* for  $\theta$ . As mentioned earlier, any function  $g(\mathbf{X})$  is also called a *statistic* for  $\theta$ . Having determined the estimator  $\hat{\theta}$ , we cannot conclude that it is exactly equal to the parameter  $\theta$ . However, we may be able to conclude that  $\theta$  is in the proximity of  $\hat{\theta}$  within the interval  $\theta_1 = g_1(\mathbf{X})$  and  $\theta_2 = g_2(\mathbf{X})$ . The *interval estimator* for  $\theta$  is the random interval,  $\{\Theta_1, \Theta_2\}$ , such that

$$P\{\Theta_1 < \theta < \Theta_2\} = \gamma \quad (18.3.1)$$

where  $\gamma$  is called the *confidence coefficient* and the interval  $(\theta_1, \theta_2)$  is the  $\gamma$ -*confidence interval*. The coefficient  $\alpha = (1 - \gamma)$  is called the *significance level*. We will discuss interval estimators in the next section.

### Estimation of Mean

The point estimator  $\hat{\mu}_X$ , called the *sample mean* for the mean  $\mu_X$  of the random variable  $X$ , can be given as the weighted sum of the independent observation random variables  $\{X_1, \dots, X_n\}$ . Thus

$$\hat{\mu}_X = \sum_{i=1}^n a_i X_i = \mathbf{a}^T \mathbf{X} \quad (18.3.2)$$

where the weight vector  $\mathbf{a}$  has to be determined according to the criteria of unbiasedness and minimum variance. The unbiasedness criterion applied to Eq. (18.3.2) yields

$$E[\hat{\mu}_X] = E\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i E[X_i] = \mu_X \sum_{i=1}^n a_i = \mu_X \quad (18.3.3)$$

Hence we have the first condition:

$$\sum_{i=1}^n a_i = 1 \quad (18.3.4)$$

The variance of the estimator  $\hat{\mu}_X$  of Eq. (18.3.2) under the assumption that the  $X_i$  values are independent is given by

$$\text{var}(\hat{\mu}_X) = \sigma_{\hat{\mu}_X}^2 = \sum_{i=1}^n a_i^2 \theta_X^2 \quad (18.3.5)$$

The problem is to minimize  $\text{var}(\hat{\mu}_X)$  subject to the constraint  $\sum_{i=1}^n a_i = 1$ . We make use of Lagrange multipliers to find a solution. We treat this problem as an unconstrained minimization by adjoining the constraint equation as shown below:

$$J = \sum_{i=1}^n a_i^2 \theta_X^2 + \lambda \left( \sum_{i=1}^n a_i - 1 \right) \quad (18.3.6)$$

where  $\lambda$  is the Lagrange multiplier that has to be determined. Since  $\lambda(\sum_{i=1}^n a_i - 1) = 0$ , we have not changed the minimization criterion. We treat Eq. (18.3.6) as unconstrained

and differentiate with respect to  $a_i$ :

$$\begin{aligned}\frac{\partial J}{\partial a_i} &= \frac{\partial}{\partial a_i} \left[ \sum_{i=1}^n a_i^2 \sigma_X^2 + \lambda \left( \sum_{i=1}^n a_i - 1 \right) \right] \\ &= 2a_i \sigma_X^2 + \lambda = 0 \quad \text{and} \quad a_i = \frac{-\lambda}{2\sigma_X^2}, \quad i = 1, \dots, n\end{aligned}\quad (18.3.7)$$

We have determined  $\{a_i\}$  in terms of the unknown Lagrange multiplier  $\lambda$ . Substituting for  $a_i$  from Eq. (18.3.7) into the constraint Eq. (18.3.4), we obtain

$$\sum_{i=1}^n \frac{-\lambda}{2\sigma_X^2} = 1 \quad \text{or} \quad \lambda = \frac{-2\sigma_X^2}{n} \quad (18.3.8)$$

Substituting Eq. (18.3.8) back into Eq. (18.3.7), we obtain the weighting coefficients  $a_i$  as

$$a_i = \frac{1}{n}, \quad i = 1, \dots, n \quad (18.3.9)$$

and the minimum variance unbiased estimator for  $\mu_X$  is given by

$$\hat{\mu}_X = \frac{1}{n} \sum_{i=1}^n X_i \quad (18.3.10)$$

We can compute the variance of  $\hat{\mu}_X$  from Eq. (18.3.5):

$$\text{var}(\hat{\mu}_X) = \sigma_{\hat{\mu}_X}^2 = \sum_{i=1}^n \frac{\sigma_X^2}{n^2} = \frac{\sigma_X^2}{n} \quad (18.3.11)$$

In Eq. (18.3.11), as  $n \rightarrow \infty$ , variance of  $\hat{\mu}_X$  tends to 0. We have already shown in Eq. (14.1.2) that if the variance is zero, then  $\hat{\mu}_X \rightarrow \mu_X$  with probability 1. Hence we can conclude that  $\hat{\mu}_X$  is a consistent estimator since it tends to  $\mu_X$  as  $n \rightarrow \infty$ . The density of  $\hat{\mu}_X$ , which is centered at  $\mu_X$ , is shown in Fig. 18.3.1 for different values of  $n$ .

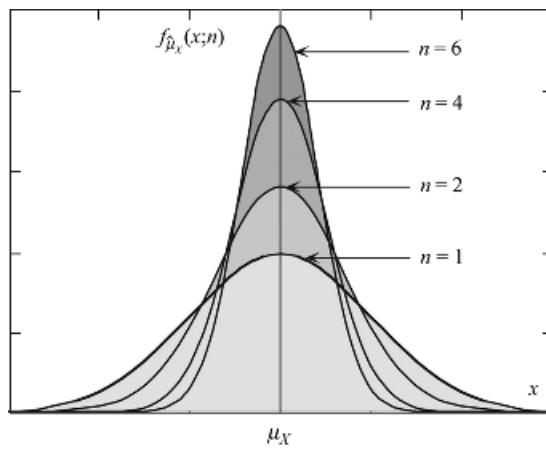


FIGURE 18.3.1

### Minimum Mean-Square Estimator for the Mean

Following the lines of Eq. (18.3.3), we will find a minimum mean-square estimator (MMSE)  $\bar{\mu}_X$  such that

$$\bar{\mu}_X = \sum_{i=1}^n a_i X_i \quad \text{and} \quad E[\bar{\mu}_X] = \sum_{i=1}^n a_i E[X_i] = k\mu_X \quad (18.3.12)$$

and the corresponding bias  $B_{\bar{\mu}_X}$  in the estimator  $\bar{\mu}_X$  is  $B_{\bar{\mu}_X} = (k - 1)\mu_X$ . We have to find the coefficients  $\{a_i\}$  such that the variance  $\sigma_{\bar{\mu}_X}^2 = \sum_{i=1}^n a_i^2 \sigma_X^2$  is minimized subject to the constraint  $\sum_{i=1}^n a_i = k$ . Using Lagrange multipliers as in the previous case, the coefficients are found to be

$$a_i = \frac{k}{n}, \quad i = 1, \dots, n \quad (18.3.13)$$

and the corresponding minimum variance is equal to

$$\sigma_{\bar{\mu}_X}^2 = \frac{k^2 \sigma_X^2}{n} \quad (18.3.14a)$$

and the biased estimator in terms of the unbiased estimator is given by

$$\bar{\mu}_X = \frac{k}{n} \hat{\mu}_X \quad (18.3.14b)$$

We will find  $k$  that minimizes the mean-square error criterion given by

$$\begin{aligned} \text{MSE} &= \bar{\varepsilon}^2 = E[\bar{\mu}_X - \mu_X]^2 = E[(\bar{\mu}_X - E[\bar{\mu}_X]) + (E[\bar{\mu}_X] - \mu_X)]^2 \\ &= \sigma_{\bar{\mu}_X}^2 + B_{\bar{\mu}_X}^2, \quad E[(\bar{\mu}_X - E[\bar{\mu}_X])] = 0 \end{aligned} \quad (18.3.15)$$

Substituting Eqs. (18.3.3) and (18.3.4) in Eq. (18.3.15), we have

$$\bar{\varepsilon}^2 = \sigma_{\bar{\mu}_X}^2 + B_{\bar{\mu}_X}^2 = \frac{k^2 \sigma_X^2}{n} + (k - 1)^2 \mu_X^2 \quad (18.3.16)$$

Differentiating Eq. (18.3.16) with respect to  $k$ , we obtain

$$\frac{\partial}{\partial k} [\bar{\varepsilon}^2] = 2k \frac{\sigma_X^2}{n} + 2(k - 1)\mu_X^2 = 0$$

or

$$k_{\min} = \frac{\mu_X^2}{\frac{\sigma_X^2}{n} + \mu_X^2} \quad \text{and} \quad \bar{\mu}_X = \frac{\mu_X^2}{\frac{\sigma_X^2}{n} + \mu_X^2} \hat{\mu}_X \quad (18.3.17)$$

The variance of the estimator  $\bar{\mu}_X$  is obtained by substituting for  $k$  in Eq. (18.3.14a) and

$$\sigma_{\bar{\mu}_X}^2 = k_{\min}^2 \frac{\sigma_X^2}{n} = \frac{\mu_X^4}{\left(\frac{\sigma_X^2}{n} + \mu_X^2\right)^2} \frac{\sigma_X^2}{n} \quad (18.3.18)$$

which is less than  $(\hat{\mu}_X)$  but biased with bias equal to

$$B_{\bar{\mu}_X} = (k_{\min} - 1)\mu_X = -\frac{\frac{\sigma_X^2}{n}\mu_X}{\frac{\sigma_X^2}{n} + \mu_X^2} \quad (18.3.19)$$

However,  $\bar{\mu}_X$  is consistent and asymptotically unbiased as  $n \rightarrow \infty$ . This may not be a useful estimator because it is dependent on  $\sigma_X^2$  and  $\mu_X$ , which are both unknown quantities. At the very least, the ratio  $\sigma_X/\mu_X$  must be known. This example illustrates that biased estimators in some cases may be better estimators than unbiased ones.

### Estimation of Variance

Analogous to formulation of the sample mean, we can express the sample variance to be of the form

$$\bar{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_X)^2 \quad (18.3.20)$$

and indeed  $\bar{\sigma}_X^2$  will be an unbiased minimum variance estimator; but  $\mu_X$  is unknown, and hence  $\bar{\sigma}_X^2$  may not be a useful estimator. However, if we substitute the sample mean  $\hat{\mu}_X$  instead of  $\mu_X$  in Eq. (18.3.20), then we can write the new estimator  $\tilde{\sigma}_X^2$  as follows:

$$\tilde{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu}_X)^2 \quad (18.3.21)$$

To determine whether  $\tilde{\sigma}_X^2$  is unbiased, we take expectations on both sides of Eq. (18.3.21) and write

$$\begin{aligned} E(\tilde{\sigma}_X^2) &= E\left\{ \frac{1}{n} \sum_{i=1}^n \left[ X_i^2 - \frac{2}{n} X_i \sum_{j=1}^n X_j + \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n X_k \right] \right\} \\ &= E\left\{ \frac{1}{n} \sum_{i=1}^n \left[ X_i^2 - \frac{2}{n} X_i^2 - \frac{2}{n} \sum_{j=1, j \neq i}^n X_i X_j + \frac{1}{n^2} \sum_{j=1}^n X_j^2 + \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1, k \neq j}^n X_j X_k \right] \right\} \quad (18.3.22) \\ &= \frac{1}{n} \sum_{i=1}^n \left[ \left( 1 - \frac{2}{n} + \frac{n}{n^2} \right) (\sigma_X^2 + \mu_X^2) - \frac{2(n-1)}{n} \mu_X^2 + \frac{n(n-1)}{n^2} \mu_X^2 \right] \\ &= \frac{1}{n} \sum_{i=1}^n \left[ \left( \frac{n-1}{n} \right) \sigma_X^2 + \left( \frac{n-1}{n} \right) \mu_X^2 - \frac{2(n-1)}{n} \mu_X^2 + \frac{(n-1)}{n} \mu_X^2 \right] = \frac{n-1}{n} \sigma_X^2 \end{aligned}$$

As we can see from Eq. (18.3.22), the estimator  $\tilde{\sigma}_X^2$  is biased. To obtain an unbiased estimator, we have to divide Eq. (18.3.21) by  $(n-1)$  instead of by  $n$  and write the unbiased minimum variance estimator  $\hat{\sigma}_X^2$ , also called the *sample variance*, as

$$\hat{\sigma}_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu}_X)^2 \quad (18.3.23)$$

In forming an unbiased estimator for the variance, the degrees of freedom in the estimator determines the divisor. In Eq. (18.3.20) there are  $n$  degrees of freedom since there are  $n$

independent values of  $X_i$ , and hence the divisor is  $n$ . However, in Eq. (18.3.23) there are only  $(n - 1)$  degrees of freedom since  $\hat{\mu}_X = \sum_{i=1}^n (X_i/n)$ , which takes away one degree of freedom. Thus, the divisor is  $(n - 1)$ . The square root of the sample variance  $\hat{\sigma}_X^2$  is called the *standard error*.

We will now compute the variance of the sample variance

$$\begin{aligned}
 \text{var}(\hat{\sigma}_X^2) &= E\left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu}_X)^2\right]^2 - (\sigma_X^2)^2 \\
 &= E\left\{\frac{1}{(n-1)^2} \sum_{i=1}^n (X_i - \hat{\mu}_X)^2 \sum_{j=1}^n (X_j - \hat{\mu}_X)^2\right\} - \sigma_X^4 \\
 &= E\left\{\frac{1}{(n-1)^2} \left[ \sum_{i=1}^n (X_i - \hat{\mu}_X)^4 + \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n (X_i - \hat{\mu}_X)^2 (X_j - \hat{\mu}_X)^2 \right]\right\} - \sigma_X^4 \\
 &\approx \frac{n}{(n-1)^2} v_4 + \frac{n(n-1)}{(n-1)^2} \sigma_X^4 - \sigma_X^4 = \frac{n}{(n-1)^2} v_4 + \frac{\sigma_X^4}{n-1} \tag{18.3.24}
 \end{aligned}$$

where  $v_4 = E(X_i - \mu_X)^4$  from Eq. (10.4.2). We can also confirm that  $\hat{\sigma}_X^2$  is a consistent estimator since  $\text{var}(\hat{\sigma}_X^2)$  tends to 0 as  $n$  tends to  $\infty$ . In fact, all three estimators,  $\bar{\sigma}_X^2$ ,  $\tilde{\sigma}_X^2$ , and  $\hat{\sigma}_X^2$ , are consistent while only  $\bar{\sigma}_X^2$  and  $\hat{\sigma}_X^2$  are unbiased. However, for large  $n$  all the estimators are asymptotically unbiased.

### Estimation of Covariance

The unbiased estimator for the covariance  $\sigma_{XY}$  between two random variables  $X$  and  $Y$  can be derived with the following assumptions. Corresponding to each of the random variables, we have two sets of independent observations  $\{X_1, \dots, X_n\}$  and  $\{Y_1, \dots, Y_n\}$  with  $E[X_i] = \mu_X$ ,  $E[Y_i] = \mu_Y$ ,  $E[X_i Y_i] = \sigma_{XY} + \mu_X \mu_Y$ , and  $E[X_i Y_j] = \mu_X \mu_Y$  for  $i \neq j$ . The estimator  $\hat{\sigma}_{XY}$  for the covariance can be expressed as

$$\hat{\sigma}_{XY} = \frac{1}{K} \sum_{i=1}^n (X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y) \tag{18.3.25}$$

where the divisor  $K$  is found from the following steps to satisfy the unbiasedness condition  $E(\hat{\sigma}_{XY}) = \sigma_{XY}$ :

$$\begin{aligned}
 E\left[\sum_{i=1}^n (X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y)\right] &= \sum_i E[X_i Y_i - X_i \hat{\mu}_Y - Y_i \hat{\mu}_X + \hat{\mu}_X \hat{\mu}_Y] \\
 &= \sum_i E\left[X_i Y_i - X_i \frac{1}{n} \sum Y_j - Y_i \frac{1}{n} \sum X_j + \frac{1}{n^2} \sum \sum X_i Y_j\right] \\
 &= \sum_i E\left[X_i Y_i - \frac{X_i Y_i}{n} - \sum_{j \neq i} \frac{X_i Y_j}{n} - \frac{X_i Y_i}{n} - \sum_{j \neq i} \frac{Y_i X_j}{n} + \frac{1}{n^2} \sum X_i Y_i\right. \\
 &\quad \left. + \frac{1}{n^2} \sum_{j \neq i} \sum_i X_i Y_j\right]
 \end{aligned}$$

$$\begin{aligned}
&= \sum_i \left[ (\sigma_{XY} + \mu_X \mu_Y) \left( 1 - \frac{1}{n} \right) - \frac{2(n-1)}{n} \mu_X \mu_Y + \frac{n(n-1)}{n^2} \mu_X \mu_Y \right] \\
&= \sum_i \left[ \frac{n-1}{n} \sigma_{XY} + \frac{n-1}{n} \mu_X \mu_Y - \frac{2(n-1)}{n} \mu_X \mu_Y + \frac{n-1}{n} \mu_X \mu_Y \right] \\
&= (n-1) \sigma_{XY}
\end{aligned} \tag{18.3.26}$$

From Eq. (18.3.26)  $K$  is found to be  $(n-1)$ , and the unbiased minimum variance estimator for the covariance is

$$\hat{\sigma}_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y) \tag{18.3.27}$$

a result very similar to that of the estimator  $\hat{\sigma}_X^2$  of the variance.

We will now compute the variance of the sample covariance

$$\begin{aligned}
\text{var}(\hat{\sigma}_{XY}) &= E \left[ \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y) \right]^2 - (\sigma_{XY})^2 \\
&= E \left\{ \frac{1}{(n-1)^2} \sum_{i=1}^n [(X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y)] \sum_{j=1}^n [(X_j - \hat{\mu}_X)(Y_j - \hat{\mu}_Y)] \right\} - \sigma_{XY}^2 \\
&= E \left\{ \frac{1}{(n-1)^2} \left[ \begin{array}{l} \sum_{i=1}^n [(X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y)]^2 \\ + \sum_{i=1}^n \sum_{j=1, j \neq i}^n [(X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y)][(X_j - \hat{\mu}_X)(Y_j - \hat{\mu}_Y)] \end{array} \right] \right\} - \sigma_{XY}^2 \\
&\approx \frac{n}{(n-1)^2} v_2^{XY} + \frac{n(n-1)}{(n-1)^2} \sigma_{XY}^2 - \sigma_{XY}^2 = \frac{n}{(n-1)^2} v_2^{XY} + \frac{\sigma_{XY}^2}{n-1}
\end{aligned} \tag{18.3.28}$$

where we have defined  $v_2^{XY} = E\{[(X_i - \mu_X)(Y_i - \mu_Y)]^2\}$  as the second cross-central moment between  $X$  and  $Y$ . Since  $\text{var}(\hat{\sigma}_{XY})$  tends to 0 as  $n$  tends to  $\infty$ , we conclude that  $(\hat{\sigma}_{XY})$  is also a consistent unbiased minimum variance estimator.

In Eqs. (18.2.13) and (18.2.17) the means and variances were explicitly stated. In general, these quantities can only be estimated. We will revisit the linear regression problem later.

**Example 18.3.1** We will find the estimated means and variances of the random variables  $X$  and  $Y$  from the data given in Example 18.2.1. From Eq. (18.3.10) the unbiased, minimum variance estimators for the mean of  $X$  and  $Y$  are

$$\hat{\mu}_X = \frac{1}{20} [0.05 + 0.29 + \dots + 1.52] = \frac{18.1}{20} = 0.905$$

$$\hat{\mu}_Y = \frac{1}{20} [1.82 + 2.05 + \dots + 2.82] = \frac{38.78}{20} = 1.939$$

Similarly, from Eq. (18.3.9) the estimated variances of  $X$  and  $Y$  are

$$\begin{aligned}\hat{\sigma}_X^2 &= \frac{1}{19}[(0.05 - 0.905)^2 + (0.29 - 0.905)^2 + \cdots + (1.52 - 0.905)^2] \\ &= \frac{3.767}{19} = 0.198 \\ \hat{\sigma}_Y^2 &= \frac{1}{19}[(1.82 - 1.939)^2 + (2.05 - 1.939)^2 + \cdots + (2.82 - 1.939)^2] \\ &= \frac{5.857}{19} = 0.308\end{aligned}$$

The estimated covariance between  $X$  and  $Y$  is

$$\begin{aligned}\hat{\sigma}_X Y &= \frac{1}{19}[(0.05 - 0.905)(1.82 - 1.939) + \cdots + (1.52 - 0.905)(2.82 - 1.939)] \\ &= \frac{1.521}{19} = 0.080\end{aligned}$$

The estimated correlation coefficient is

$$\hat{\rho}_{XY} = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X \hat{\sigma}_Y} = \frac{0.08}{0.247} = 0.324$$

These parameters are the same numbers that we used in Example 18.2.1. However, the regression coefficients  $a$  and  $b$  will be estimates and will be governed by a probability distribution.

**Example 18.3.2** We will now find the estimated means and variances of the random variables  $X$  and  $Y$  from the data given in Example 18.2.3. From Eq. (18.3.10) the unbiased, minimum variance estimators for the mean of  $X$  and  $Y$  are

$$\hat{\mu}_X = \frac{1}{21}[0.91 + 0.22 + \cdots + 1.64] = \frac{25.93}{21} = 1.235$$

$$\hat{\mu}_Y = \frac{1}{21}[2.6 + 2.21 + \cdots + 1.95] = \frac{46.41}{21} = 2.21$$

Similarly, from Eq. (18.3.9) the estimated variances of  $X$  and  $Y$  are

$$\begin{aligned}\hat{\sigma}_X^2 &= \frac{1}{20}[(0.91 - 1.235)^2 + (0.22 - 1.235)^2 + \cdots + (1.64 - 1.235)^2] \\ &= \frac{5.50}{20} = 0.275 \\ \hat{\sigma}_Y^2 &= \frac{1}{20}[(2.6 - 2.21)^2 + (2.21 - 2.21)^2 + \cdots + (1.95 - 2.21)^2] \\ &= \frac{0.795}{20} = 0.0398\end{aligned}$$

We will now find estimated values for the higher-order moments of  $X$ . They are given by

$$\hat{m}_{4X} = \hat{E}[X^4] = \frac{1}{21}[0.91^4 + 0.22^4 + \dots + 1.64^4] = \frac{95.1}{21} = 4.53$$

$$\hat{m}_{3X} = \hat{E}[X^3] = \frac{1}{21}[0.91^3 + 0.22^3 + \dots + 1.64^3] = \frac{58.43}{21} = 2.78$$

$$\hat{m}_{2X} = \hat{E}[X^2] = \frac{1}{21}[0.91^2 + 0.22^2 + \dots + 1.64^2] = \frac{37.52}{21} = 1.79$$

The estimated second moment of  $Y$  is given by

$$\hat{m}_{2Y} = \hat{E}[Y^2] = \frac{1}{21}[2.6^2 + 2.21^2 + \dots + 1.95^2] = \frac{103.36}{21} = 4.92$$

The estimated cross-moments of  $X$  with  $Y$  are

$$\hat{E}[XY] = \frac{1}{21}[0.91 \times 2.6 + 0.22 \times 2.21 + \dots + 1.64 \times 1.95] = \frac{56.1}{21} = 2.67$$

$$\hat{E}[X^2Y] = \frac{1}{21}[0.91^2 \times 2.6 + 0.22^2 \times 2.21 + \dots + 1.64^2 \times 1.95] = \frac{79.6}{21} = 3.79$$

The estimated correlation matrix  $\mathbf{R}_X$  is,

$$\hat{\mathbf{R}}_X = \begin{bmatrix} \hat{E}[X^4] & \hat{E}[X^3] & \hat{E}[X^2] \\ \hat{E}[X^3] & \hat{E}[X^2] & \hat{E}[X] \\ \hat{E}[X^2] & \hat{E}[X] & 1 \end{bmatrix} = \begin{bmatrix} 4.53 & 2.78 & 1.79 \\ 2.78 & 1.79 & 1.23 \\ 1.79 & 1.23 & 1 \end{bmatrix}$$

These values are the same values given in Example 18.2.3. Again, the regression coefficients  $a$ ,  $b$ , and  $c$  will be only estimates.

### ***Estimated Regression Coefficients from Data***

As stated in previous examples, the evaluated regression coefficients will be estimates of the true values since the means and variances are only estimates. We will now discuss the methodology of finding the estimated regression coefficients and determining their variances from the given data. The linear regression problem for each  $y_i$  can be stated as

$$y_i = a_1 x_{i1} + \dots + a_m x_{im} + w_i = \mathbf{x}_i^T \mathbf{a} + w_i, \quad i = 1, \dots, n$$

with  $n > m$ . Or more succinctly

$$\mathbf{y} = \mathbf{X}^T \mathbf{a} + \mathbf{w} \quad (18.3.29)$$

where the  $n$ -vector  $\mathbf{y}$  is the realization of the random variable  $Y$  with  $E(\mathbf{y}) = \mathbf{X}^T \mathbf{a}$  where  $\mathbf{a}$  is an unknown  $m$ -vector parameter to be estimated, and  $\mathbf{x}_i$  is the observation  $m$ -vector given by  $\mathbf{x}_i = [x_{i1}, \dots, x_{im}]^T$ . In linear regression, the coefficient  $a_1$  represents the offset and hence  $x_{i1}=1$  for all  $i$ . The  $n$ -vector  $\mathbf{w}$  is a realization of the measurement noise random variable  $W$ . The elements  $w_i$  of  $\mathbf{w}$  are assumed to be uncorrelated with zero mean and variance  $E[w_i^2] = \sigma_w^2$  for all  $i$ .  $\mathbf{X}$  is an  $m \times n$  ( $n > m$ ) matrix of

observations given by

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_i & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 & \cdots & 1 \\ x_{12} & x_{22} & \cdots & x_{i2} & \cdots & x_{n2} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{1j} & x_{2j} & \cdots & x_{ij} & \cdots & x_{nj} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{1m} & x_{2m} & \cdots & x_{im} & \cdots & x_{nm} \end{bmatrix} \quad (18.3.30)$$

Given this information, we have to estimate the regression coefficient vector  $\mathbf{a}$  using the least-squares error. Thus, the criterion function  $J(\mathbf{a})$  is given by

$$J(\mathbf{a}) = (\mathbf{y} - \mathbf{X}_a^T)^T(\mathbf{y} - \mathbf{X}_a^T) = \mathbf{y}^T \mathbf{y} - \mathbf{a}^T \mathbf{X} \mathbf{y} - \mathbf{y}^T \mathbf{X}^T \mathbf{a} + \mathbf{a}^T \mathbf{X} \mathbf{X}^T \mathbf{a} \quad (18.3.31)$$

From Eq. (16.3.4) the derivative with respect to the vector  $\mathbf{a}$  of Eq. (18.3.31) is

$$\begin{aligned} \frac{\partial}{\partial \mathbf{a}} J(\mathbf{a}) &= \frac{\partial}{\partial \mathbf{a}} [\mathbf{y}^T \mathbf{y} - \mathbf{a}^T \mathbf{X} \mathbf{y} - \mathbf{y}^T \mathbf{X}^T \mathbf{a} + \mathbf{a}^T \mathbf{X} \mathbf{X}^T \mathbf{a}] \\ &= -2\mathbf{X} \mathbf{y} + 2\mathbf{X} \mathbf{X}^T \mathbf{a} = 0 \end{aligned} \quad (18.3.32)$$

From Eq. (18.3.32) the estimated regression coefficients  $\hat{\mathbf{a}}$  are given by

$$\hat{\mathbf{a}} = (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{y} \quad (18.3.33)$$

where  $\mathbf{X} \mathbf{X}^T$  expanded in terms of the observation points  $\{x_{ij}\}$  is given by

$$\mathbf{X} \mathbf{X}^T = \begin{bmatrix} n & \sum_n x_{i2} & \cdots & \sum_n x_{ij} & \cdots & \sum_n x_{im} \\ \sum_n x_{i2} & \sum_n x_{i2}^2 & \cdots & \sum_n x_{i2} x_{ij} & \cdots & \sum_n x_{i2} x_{im} \\ \vdots & \vdots & & \vdots & & \vdots \\ \sum_n x_{ij} & \sum_n x_{ij} x_{i2} & \cdots & \sum_n x_{ij}^2 & \cdots & \sum_n x_{ij} x_{im} \\ \vdots & \vdots & & \vdots & & \vdots \\ \sum_n x_{im} & \sum_n x_{im} x_{i2} & \cdots & \sum_n x_{im} x_{ij} & \cdots & \sum_n x_{im}^2 \end{bmatrix} \quad (18.3.34)$$

The equation for the regression line is given by

$$\hat{\mathbf{y}} = \mathbf{X}^T \hat{\mathbf{a}} \quad (18.3.35)$$

The estimator  $\hat{\mathbf{a}}$  is an unbiased estimator of  $\mathbf{a}$  since

$$E(\hat{\mathbf{a}}) = (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} E(\mathbf{y}) = (\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{X}^T \mathbf{a} = \mathbf{a}$$

### Covariance of the Estimated Regression Coefficients $\hat{\mathbf{a}}$

Before finding the covariance of  $\mathbf{a}$ , we will express  $E[\hat{\mathbf{a}} - \mathbf{a}]$  in a more convenient form given by

$$\begin{aligned} E[\hat{\mathbf{a}} - \mathbf{a}] &= E[(\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{y} - \mathbf{a}] = E[(\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{a} + \mathbf{w}) - \mathbf{a}] \\ &= E[(\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{w}] \end{aligned} \quad (18.3.36)$$

The covariance matrix  $\mathbf{C}_{\hat{\mathbf{a}}}$  of the estimated regression coefficients can now be given by

$$\begin{aligned}\mathbf{C}_{\hat{\mathbf{a}}} &= \text{cov}(\hat{\mathbf{a}}) = \text{cov}[(\hat{\mathbf{a}} - \mathbf{a})(\hat{\mathbf{a}} - \mathbf{a})^T] \\ &= \text{cov}\{(\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{w}\mathbf{w}^T\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-T}\} \\ &= (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\text{cov}(\mathbf{w}\mathbf{w}^T)\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-T} \\ &= (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\sigma_w^2\mathbf{I}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-T} = (\mathbf{X}\mathbf{X}^T)^{-1}\sigma_w^2\end{aligned}\quad (18.3.37)$$

where  $\text{cov}(\mathbf{w}\mathbf{w}^T) = \sigma_w^2\mathbf{I}_m$ . The variances of the regression coefficients are given by the diagonal terms of the covariance matrix  $\mathbf{C}_{\hat{\mathbf{a}}}$ . Since we do not know the variance  $\sigma_w^2$  of the measurement noise, we use its estimate  $\hat{\sigma}_w^2$ , given by

$$\hat{\sigma}_w^2 = \frac{(\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}})}{n - m} = \frac{(\mathbf{y} - \mathbf{X}^T\hat{\mathbf{a}})(\mathbf{y} - \mathbf{X}^T\hat{\mathbf{a}})^T}{n - m} \quad (18.3.38)$$

In this equation, the divisor  $(n - m)$  reflects the reduction in the number of degrees of freedom  $n$  by the number of regression coefficients  $m$ . This yields an unbiased estimator for  $\sigma_w^2$ .

### **Variance of the Estimated Regression line $\hat{y}_i$**

We will now find the variance of the estimated mean  $\hat{y}_i$ , given by

$$\hat{y}_i = \hat{a}_1 x_{i1} + \cdots + \hat{a}_m x_{im} = \hat{\mathbf{a}}^T \mathbf{x}_i, \quad i = 1, \dots, n \quad (18.3.39)$$

where  $\hat{y}_i$  is an unbiased estimator of  $y_i$  since  $E[\hat{y}_i] = y_i$ . The variance of  $\hat{y}_i$  can be given by

$$\begin{aligned}E[\hat{y}_i - y_i]^2 &= E[(\hat{y}_i - y_i)^T(\hat{y}_i - y_i)] \\ &= \mathbf{x}_i^T E[(\hat{\mathbf{a}} - \mathbf{a})(\hat{\mathbf{a}} - \mathbf{a})^T] \mathbf{x}_i = \mathbf{x}_i^T \text{cov}(\hat{\mathbf{a}}) \mathbf{x}_i\end{aligned}\quad (18.3.40)$$

Substituting for  $\text{cov}(\hat{\mathbf{a}})$  from Eq. (18.3.37), we have

$$\text{var}(\hat{y}_i) = \sigma_{\hat{y}_i}^2 = \mathbf{x}_i^T (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{x}_i \hat{\sigma}_w^2, \quad i = 1, \dots, n \quad (18.3.41)$$

and since we do not know the variance, we substitute the estimated variance  $\hat{\sigma}_w^2$  and obtain the estimated variance of  $\hat{y}_i$  as

$$\hat{\sigma}_{\hat{y}_i}^2 = \mathbf{x}_i^T (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{x}_i \hat{\sigma}_w^2, \quad i = 1, \dots, n \quad (18.3.42)$$

which is a function of  $i$ , and is chi-square-distributed with  $(n - m)$  degrees of freedom.

### **Simple Linear Regression**

The governing equation for a simple linear regression is

$$y_i = a_1 + a_2 x_i + w_i, \quad i = 1, \dots, n \quad (18.3.43)$$

with the usual definitions. From the data, we will derive expressions for the estimated regression coefficients for  $a_1$  and  $a_2$ . The normal equations can be obtained from Eqs. (18.3.32) and (18.3.34) as follows:

$$\begin{bmatrix} n & \sum_n x_i \\ \sum_n x_i & \sum_n x_i^2 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \begin{bmatrix} \sum_n y_i \\ \sum_n x_i y_i \end{bmatrix} \quad (18.3.44)$$

The determinant of the rectangular matrix of Eq. (18.3.44) is

$$\begin{aligned} \begin{vmatrix} n & \sum_n x_i \\ \sum_n x_i & \sum_n x_i^2 \end{vmatrix} &= n \sum_n x_i^2 - \left( \sum_n x_i \right)^2 \\ &= n \sum_n x_i^2 - m^2 \hat{\mu}_x^2 = n \sum_n (x_i - \hat{\mu}_x)^2 \end{aligned} \quad (18.3.45)$$

where  $\hat{\mu}_x = (1/n) \sum_n x_i$ . Solving Eq. (18.3.44) and substituting in Eq. (18.3.45), the estimated coefficients  $\hat{a}_1$  and  $\hat{a}_2$  are

$$\begin{aligned} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} &= \frac{1}{n \sum_n (x_i - \hat{\mu}_x)^2} \begin{bmatrix} \sum_n x_i^2 & -\sum_n x_i \\ -\sum_n x_i & n \end{bmatrix} \begin{bmatrix} \sum_n y_i \\ \sum_n x_i y_i \end{bmatrix} \\ &= \frac{1}{n \sum_n (x_i - \hat{\mu}_x)^2} \begin{bmatrix} \sum_n x_i^2 \sum_n y_i - \sum_n x_i \sum_n x_i y_i \\ n \sum_n x_i y_i - \sum_n x_i \sum_n y_i \end{bmatrix} \end{aligned} \quad (18.3.46)$$

In Eq. (18.3.46)

$$n \sum_n x_i y_i - \sum_n x_i \sum_n y_i = n \sum_n x_i y_i - n^2 \hat{\mu}_x \hat{\mu}_y = n \sum_n (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)$$

Hence

$$\hat{a}_2 = \frac{\sum_n (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)}{\sum_n (x_i - \hat{\mu}_x)^2} \quad (18.3.47)$$

Also, in Eq. (18.3.46)

$$\begin{aligned} \sum_n x_i^2 \sum_n y_i - \sum_n x_i \sum_n x_i y_i &= n \hat{\mu}_y \sum_n x_i^2 - \sum_n x_i \left[ \sum_n (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y) + n \hat{\mu}_x \hat{\mu}_y \right] \\ &= n \hat{\mu}_y \left[ \sum_n (x_i^2 - \hat{\mu}_x^2) \right] - n \hat{\mu}_x \sum_n (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y) \end{aligned}$$

Hence

$$\hat{a}_1 = \hat{\mu}_y - \frac{\sum_n (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)}{\sum_n (x_i^2 - \hat{\mu}_x^2)} \hat{\mu}_x = \hat{\mu}_y - \hat{a}_2 \hat{\mu}_x \quad (18.3.48)$$

The estimated regression line equation is

$$\hat{y}_i = \hat{a}_1 + \hat{a}_2 x_i, \quad i = 1, \dots, n \quad (18.3.49)$$

We can now compute the variances of  $\hat{a}_1$  and  $\hat{a}_2$  from Eqs. (18.3.33)

$$\text{var}\left\{\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix}\right\} = \frac{1}{n \sum_n (x_i - \hat{\mu}_x)^2} \begin{bmatrix} \sum_n x_i^2 & -\sum_n x_i \\ -\sum_n x_i & n \end{bmatrix} \hat{\sigma}_w^2 \quad (18.3.50)$$

where we have substituted the estimated variance  $\hat{\sigma}_w^2$  from Eq. (18.3.38) for the true variance  $\sigma_w^2$ .

Simplifying Eq. (18.3.50), we obtain the variances of  $\hat{a}_1$ , and  $\hat{a}_2$  as

$$\begin{aligned} \text{var}(\hat{a}_1) &= \hat{\sigma}_{\hat{a}_1}^2 = \frac{\hat{\sigma}_w^2}{\sum_n (x_i - \hat{\mu}_x)^2} \\ \text{var}(\hat{a}_2) &= \hat{\sigma}_{\hat{a}_2}^2 = \hat{\sigma}_w^2 \left[ \frac{\sum_n (x_i - \hat{\mu}_x)^2 + n(\hat{\mu}_x^2)}{n \sum_n (x_i - \hat{\mu}_x)^2} \right] = \hat{\sigma}_w^2 \left[ \frac{1}{n} + \frac{\hat{\mu}_x^2}{\sum_n (x_i - \hat{\mu}_x)^2} \right] \end{aligned} \quad (18.3.51)$$

Finally, we can compute the estimated variance of the estimated regression line  $\hat{y}_i$ . From Eq. (18.3.42), we have

$$\hat{\sigma}_{\hat{y}_i}^2 = \frac{1}{n \sum_n (x_j - \hat{\mu}_x)^2} \begin{bmatrix} 1 & x_i \end{bmatrix} \begin{bmatrix} \sum_n x_j^2 & -\sum_n x_j \\ -\sum_n x_j & n \end{bmatrix} \begin{bmatrix} 1 \\ x_i \end{bmatrix} \hat{\sigma}_w^2, \quad i = 1, \dots, n \quad (18.3.52)$$

Equation (18.3.52) can be simplified as follows

$$\begin{aligned} \begin{bmatrix} 1 & x_i \end{bmatrix} \begin{bmatrix} \sum_n x_j^2 & -\sum_n x_j \\ -\sum_n x_j & n \end{bmatrix} \begin{bmatrix} 1 \\ x_i \end{bmatrix} &= \sum_n x_j^2 - 2x_i \sum_n x_j + nx_i^2 \\ &= \sum_n (x_j - \hat{\mu}_x)^2 + n(x_i - \hat{\mu}_x)^2 \end{aligned} \quad (18.3.53)$$

where we have subtracted and added  $\hat{\mu}_x^2$ . Substituting Eq. (18.3.53) in Eq. (18.3.52), we obtain

$$\begin{aligned} \hat{\sigma}_{\hat{y}_i}^2 &= \frac{1}{n \sum_n (x_j - \hat{\mu}_x)^2} \left[ \sum_n (x_j - \hat{\mu}_x)^2 + n(x_i - \hat{\mu}_x)^2 \right] \hat{\sigma}_w^2 \\ &= \left[ \frac{1}{n} + \frac{(x_i - \hat{\mu}_x)^2}{\sum_n (x_j - \hat{\mu}_x)^2} \right] \hat{\sigma}_w^2, \quad i = 1, \dots, n \end{aligned} \quad (18.3.54)$$

**Example 18.3.3** From the data in Example 18.2.1 we will determine the estimated regression coefficients and their variances using the techniques developed above. The linear equation used to solve for the regression coefficients is

$$y_i = a_1 + a_2 x_i + w_i, \quad i = 1, \dots, 20$$

In Eq. (18.3.29) the 20-vector  $\mathbf{y}$  is given by

$$\mathbf{y} = \begin{bmatrix} 1.82, 2.05, 1.28, 1.19, 2.06, 1.96, 2.46, 1.76, 1.13, 2.35, \\ 1.49, 2.15, 0.88, 1.49, 2.18, 2.74, 2.10, 2.69, 2.18, 2.82 \end{bmatrix}^T$$

and the 20-vector  $\mathbf{x}$  is given by,

$$\mathbf{x} = \begin{bmatrix} 0.05, 0.29, 0.74, 0.55, 1.07, 0.47, 1.06, 0.7, 0.54, 0.65, \\ 1.54, 0.72, 0.66, 1.23, 1.35, 0.97, 1.3, 0.96, 1.73, 1.52 \end{bmatrix}^T$$

$\mathbf{a}$  is a 2-vector of regression coefficients given by  $\mathbf{a} = [a_1, a_2]^T$  and  $\mathbf{w}$  is an unknown measurement 20-vector noise whose variance can only be estimated. We can now form the  $2 \times 20$  matrix  $\mathbf{X}$ , where the first row is always 1 since  $a_1$  multiplies 1 for all  $i$ . Hence the matrix  $\mathbf{X}$  is

$$\mathbf{X} = \left\{ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0.05 & 0.29 & 0.74 & 0.55 & 1.07 & 0.47 & 1.06 & 0.7 & 0.54 & 0.65 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1.54 & 0.72 & 0.66 & 1.23 & 1.35 & 0.97 & 1.3 & 0.96 & 1.73 & 1.52 \end{bmatrix} \right\}$$

The quantity  $\mathbf{X}\mathbf{X}^T$  and  $(\mathbf{X}\mathbf{X}^T)^{-1}$  are given by

$$\mathbf{X}\mathbf{X}^T = \begin{bmatrix} 20 & 18.1 \\ 18.1 & 20.147 \end{bmatrix} \quad (\mathbf{X}\mathbf{X}^T)^{-1} = \begin{bmatrix} 0.267 & -0.24 \\ -0.24 & 0.265 \end{bmatrix}$$

The estimated regression coefficients  $\hat{\mathbf{a}}$  are obtained from Eq. (18.3.33):

$$\hat{\mathbf{a}} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = (\mathbf{X}\mathbf{X}^T)^{-1}(\mathbf{X}\mathbf{y}) = \begin{bmatrix} 0.267 & -0.24 \\ -0.24 & 0.265 \end{bmatrix} \begin{bmatrix} 38.78 \\ 36.617 \end{bmatrix} = \begin{bmatrix} 1.574 \\ 0.404 \end{bmatrix}$$

with

$$\mathbf{X}\mathbf{y} = \begin{bmatrix} 38.78 \\ 36.617 \end{bmatrix}$$

where  $\hat{a}_1$  and  $\hat{a}_2$  exactly match the quantities  $b$  and  $a$  in Example 18.2.1. The equation of the regression line

$$\hat{y}_i = \hat{a}_1 + \hat{a}_2 x_i = 1.574 + 0.404 x_i, \quad i = 1, \dots, 20$$

has already been shown in Fig. 18.2.2. We will now compute the estimated variance of the noise term from Eq. (18.3.38):

$$\hat{\sigma}_w^2 = \frac{(\mathbf{y} - \mathbf{X}^T \hat{\mathbf{a}})^T (\mathbf{y} - \mathbf{X}^T \hat{\mathbf{a}})}{n - m} = \frac{4.74}{20 - 2} = 0.263$$

This compares within the limits of data analysis with MMSE = 0.276, obtained in Example 18.2.1. From Eq. (18.3.37), the covariance matrix of  $\hat{\mathbf{a}}$  is

$$\mathbf{C}_{\hat{\mathbf{a}}} = \text{cov}(\hat{\mathbf{a}}) = (\mathbf{X}\mathbf{X}^T)^{-1} \hat{\sigma}_w^2 = \begin{bmatrix} 0.267 & -0.24 \\ -0.24 & 0.266 \end{bmatrix} \cdot 0.263 = \begin{bmatrix} 0.0704 & -0.0633 \\ -0.0633 & 0.0699 \end{bmatrix}$$

Hence  $\text{var}(\hat{a}_1) = 0.0704$  and  $\text{var}(\hat{a}_2) = 0.0699$ .

We can now obtain the variance of  $\hat{y}_i$  using Eq. (18.3.54), where  $n = 20$ , the average value  $\hat{\mu}_x = 0.905$ , and the summation  $\sum_n (x_j - \hat{\mu}_x)^2 = 3.7665$ . Substituting these values, we obtain

$$\hat{\sigma}_{\hat{y}_i}^2 = \hat{\sigma}_w^2 \left[ \frac{1}{20} + \frac{(x_i - \hat{\mu}_x)^2}{\sum_{20} (x_j - \hat{\mu}_x)^2} \right] = 0.263 \left[ \frac{1}{20} + \frac{(x_i - 0.905)^2}{3.7665} \right], \quad i = 1, \dots, 20$$

## 18.4 INTERVAL ESTIMATION (CONFIDENCE INTERVALS)

In the previous section we discussed point estimators  $\hat{\theta}$  for a parameter  $\theta$ . However, a point estimator may not be meaningful without knowledge of the probability of how far away it is from the true value. Hence, we would also like to know an interval about the estimator  $\hat{\theta}$  and the probability that the true value  $\theta$  lies in that interval. Such an interval for  $\theta$  is the random interval,  $\{\Theta_1, \Theta_2\}$ , such that

$$P\{\Theta_1 < \theta < \Theta_2\} = \gamma \quad (18.4.1)$$

where  $\gamma$  is the *confidence coefficient*. The realized interval  $(\theta_1, \theta_2)$  is the  $\gamma$  *confidence interval*. The terms  $\theta_1$  and  $\theta_2$  could possibly be functions of the estimator  $\hat{\theta}$ . In essence, we can replace the empirical point estimator with an *interval estimator* that gives more information about the population parameter. The coefficient  $\alpha = (1 - \gamma)$  is called the *significance level*. In a two-sided confidence interval as in Eq. (18.4.1), it is usual to split the significance level  $\alpha$  on both sides of the density. We will apply the confidence intervals to various parameters such as the mean and variance. We have already come across the basic ideas of confidence intervals in Section 7.9 and Examples 14.8.1 and 14.8.4.

**Example 18.4.1** In a presidential election, a news agency polled 1000 probable voters, and 470 of them said that they would vote Democratic while 460 of them said that they would vote Republican. The agency reported that the Democrat winning was 47% and the Republican winning was 46%, with a margin of error of  $\pm 4\%$ . What this means is that the true probability of the Democrat winning lies in the interval (51%, 43%) and that of the Republican lies in the interval (50%, 42%):

$$\begin{aligned} P\{0.43 < P(\text{Democrat winning}) < 0.51\} &= x \\ P\{0.42 < P(\text{Republican winning}) < 0.50\} &= x \end{aligned}$$

The conclusion arrived at is that the results are within the margin of error and hence the election is a dead heat. What is never mentioned is  $x$ , the probability of these results occurring!

*Confidence Interval for the Unknown Mean of Population (Known Variance).* We have already established in Eq. (18.3.10) that  $\hat{\mu}_X = (1/n) \sum_{i=1}^n X_i$  is an unbiased minimum variance point estimator for the mean  $\mu_X$ . If  $n$  is large enough, we can assume that  $\hat{\mu}_X$  is Gaussian-distributed with variance  $\sigma_X^2/n$ . We will find the confidence interval under the assumption that  $\sigma_X^2$  is known. The random variable  $Z$  given by

$$Z = \frac{\hat{\mu}_X - \mu_X}{\sigma_X / \sqrt{n}} \quad (18.4.2)$$

is a standard Gaussian, and for a confidence coefficient  $\gamma = 1 - \alpha$  we can write

$$\begin{aligned} P\{Z_{\alpha/2} < Z \leq z_{1-\alpha/2}\} &= 1 - \alpha = \gamma \quad \text{or} \\ P\left\{z_{\alpha/2} < \frac{\hat{\mu}_X - \mu_X}{\sigma_X / \sqrt{n}} \leq z_{1-\alpha/2}\right\} &= \gamma \end{aligned} \quad (18.4.3)$$

where we have split the significance level  $\alpha$  equally on both sides of the density function. The values  $z_{\alpha/2} = P(Z \leq \alpha/2)$  and  $z_{1-\alpha/2} = P(Z \leq 1 - \alpha/2)$  for a confidence level of  $\gamma$  are extracted from the Gaussian table (Table 18.4.1) for

TABLE 18.4.1

Gaussian Tables				
$\gamma\%$	$P(Z \leq z)$	$z$	$P(Z \leq z)$	$z$
99	0.005	-2.57583	0.995	2.57583
98	0.01	-2.32635	0.99	2.32635
95	0.025	-1.95996	0.975	1.95996
90	0.05	-1.64485	0.95	1.64485
85	0.075	-1.43953	0.925	1.43953
80	0.1	-1.28155	0.9	1.28155
75	0.125	-1.15035	0.875	1.15035
70	0.15	-1.03643	0.85	1.03643
65	0.175	-0.93459	0.825	0.93459
60	0.2	-0.84162	0.8	0.84162
55	0.225	-0.75542	0.775	0.75542
50	0.25	-0.67449	0.75	0.67449

probabilities of  $\alpha/2$  and  $(1 - \alpha/2)$  respectively. Since the Gaussian curve is symmetric,  $z_{\alpha/2} = -z_{1-\alpha/2}$ .

The  $z$  values for 95% confidence are  $z_{0.025} = -1.96$  and  $z_{0.975} = 1.96$ . The corresponding  $z$  values for 99% confidence are  $z_{0.005} = -2.576$  and  $z_{0.995} = 2.576$  and for 90% confidence  $z_{0.05} = -1.645$  and  $z_{0.95} = 1.645$ .

From Eq. (18.4.3) we have

$$P\left\{z_{\alpha/2} < \frac{\hat{\mu}_X - \mu_X}{\sigma_X/\sqrt{n}} \leq z_{1-\alpha/2}\right\} = 1 - \alpha$$

from which it follows that

$$P\left\{\frac{z_{\alpha/2}\sigma_X}{\sqrt{n}} < \hat{\mu}_X - \mu_X \leq \frac{z_{1-\alpha/2}\sigma_X}{\sqrt{n}}\right\} = 1 - \alpha$$

and finally

$$P\left\{\hat{\mu}_X - \frac{z_{1-\alpha/2}\sigma_X}{\sqrt{n}} < \mu_X \leq \hat{\mu}_X - \frac{z_{\alpha/2}\sigma_X}{\sqrt{n}}\right\} = 1 - \alpha \quad (18.4.4)$$

Equation (18.4.4) is to be interpreted as the probability that the random interval  $(\hat{\mu}_X - z_{1-\alpha/2}\sigma_X/\sqrt{n}, \hat{\mu}_X - z_{\alpha/2}\sigma_X/\sqrt{n})$  will contain the true population mean  $\mu_X$  is  $100(1 - \alpha)\%$ . The probability density of  $\hat{\mu}_X$  centered about the true mean  $\mu_X$  and the confidence intervals are shown in Fig. 18.4.1.

The 99% confidence interval is given by

$$P\left\{\hat{\mu}_X - \frac{2.576\sigma_X}{\sqrt{n}} < \mu_X \leq \hat{\mu}_X + \frac{2.576\sigma_X}{\sqrt{n}}\right\} = 0.99 \quad (18.4.5)$$

In an analogous manner, the 95% confidence interval is given by

$$P\left\{\hat{\mu}_X - \frac{1.96\sigma_X}{\sqrt{n}} < \mu_X \leq \hat{\mu}_X + \frac{1.96\sigma_X}{\sqrt{n}}\right\} = 0.95 \quad (18.4.6)$$

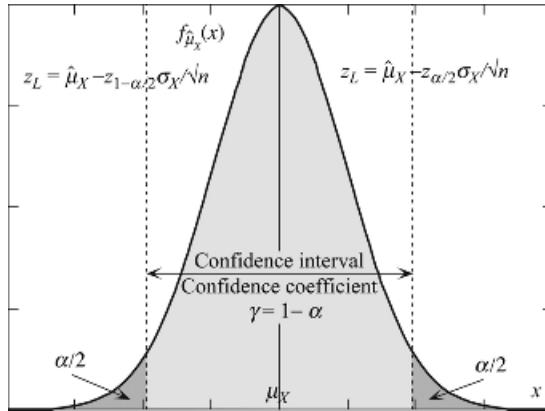


FIGURE 18.4.1

and the 90% confidence interval is given by

$$P\left\{ \hat{\mu}_X - \frac{1.645\sigma_X}{\sqrt{n}} < \mu_X \leq \hat{\mu}_X + \frac{1.645\sigma_X}{\sqrt{n}} \right\} = 0.90 \quad (18.4.7)$$

It can be seen from Eqs. (18.4.5)–(18.4.7) that as the confidence level  $\gamma$  *decreases* and conversely as the significance level  $\alpha$  *increases*, the bounds become tighter. If we want to have 100% confidence or 0% significance, then the confidence interval is  $\infty$ ! The equations also show that as the number of samples  $n$  increases, the bounds also become tighter. The confidence level is usually fixed a priori at 95% or 90% before sampling.

*Confidence Interval for the Unknown Mean of Population (Unknown Variance).* In the previous section we assumed that the variance is known. Since the true variance is known in very few cases, we use the estimated variance  $\hat{\sigma}_X^2$ . In that case we have to find the distribution of the random variable  $T$ , given by

$$T = \frac{\hat{\mu}_X - \mu_X}{\hat{\sigma}_X / \sqrt{n}} \quad (18.4.8)$$

which is a Student- $t$  distribution given by Eq. (7.9.2) with  $n$  degrees of freedom.

Equation (18.4.8) can be rewritten as

$$T = \frac{(\hat{\mu}_X - \mu_X)/\sigma_X / \sqrt{n}}{\{[(n-1)(\hat{\sigma}_X^2 / \sigma_X^2) / (n-1)]\}^{1/2}} \quad (18.4.9)$$

where the term  $(\hat{\mu}_X - \mu_X)/\sigma_X / \sqrt{n}$  is a standard Gaussian, and we will show that the term

$$\frac{(n-1)\hat{\sigma}_X^2}{\sigma_X^2} = \sum_{i=1}^n \left( \frac{X_i - \hat{\mu}_X}{\sigma_X} \right)^2$$

is chi-square-distributed with  $(n-1)$  degrees of freedom. Adding and subtracting  $\mu_X$  to  $(X_i - \hat{\mu}_X)$  in the numerator of the righthand side of the equation

above, we have

$$\begin{aligned}
 \sum_{i=1}^n \left( \frac{X_i - \hat{\mu}_X}{\sigma_X} \right)^2 &= \sum_{i=1}^n \left( \frac{X_i - \mu_X - (\hat{\mu}_X - \mu_X)}{\sigma_X} \right)^2 \\
 &= \sum_{i=1}^n \left\{ \left( \frac{X_i - \mu_X}{\sigma_X} \right)^2 + \left( \frac{\hat{\mu}_X - \mu_X}{\sigma_X} \right)^2 - \frac{2(X_i - \mu_X)(\hat{\mu}_X - \mu_X)}{\sigma_X^2} \right\} \\
 &= \left\{ \sum_{i=1}^n \left( \frac{X_i - \mu_X}{\sigma_X} \right)^2 \right\} - n \left( \frac{\hat{\mu}_X - \mu_X}{\sigma_X} \right)^2 \quad \text{since} \\
 \sum_{i=1}^n (X_i - \mu_X) &= n(\hat{\mu}_X - \mu_X) \tag{18.4.10}
 \end{aligned}$$

The first term on the righthand side of Eq. (18.4.10),  $\sum_{i=1}^n [(X_i - \mu_X)/\sigma_X]^2$  is chi-square-distributed with  $n$  degrees of freedom. Since  $(\hat{\mu}_X - \mu_X)/(\sigma_X/\sqrt{n})$  is Gaussian with mean 0 and variance  $\sigma_X/\sqrt{n}$ , the second term

$$n \left( \frac{\hat{\mu}_X - \mu_X}{\sigma_X} \right)^2 = \left( \frac{\hat{\mu}_X - \mu_X}{\sigma_X/\sqrt{n}} \right)^2$$

is chi-square-distributed with one degree of freedom. Hence  $\sum_{i=1}^n [(X_i - \hat{\mu}_X)/\sigma_X]^2$  is chi-square-distributed with  $(n - 1)$  degrees of freedom. Further, it can be shown that  $\hat{\sigma}_X^2$  and  $\hat{\mu}_X$  are independent random variables. Hence, from Example 13.4.5,  $T$  in Eq. (18.4.8) is indeed Student- $t$ -distributed with  $(n - 1)$  degrees of freedom.

We can now obtain the  $100(1 - \alpha)\%$  confidence intervals for the population mean  $\mu_X$  using the Student- $t$  as follows. Analogous to Eq. (18.4.4), we can write

$$P\{t_{\alpha/2} < T \leq t_{1-\alpha/2}\} = P\left\{t_{\alpha/2} < \frac{(\hat{\mu}_X - \mu_X)}{\hat{\sigma}_X/\sqrt{n}} \leq t_{1-\alpha/2}\right\} = 1 - \alpha = \gamma$$

from which it follows that

$$P\{t_{\alpha/2}\hat{\sigma}_X/\sqrt{n} < (\hat{\mu}_X - \mu_X) \leq t_{1-\alpha/2}\hat{\sigma}_X/\sqrt{n}\} = 1 - \alpha = \gamma$$

and finally

$$P\{\hat{\mu}_X - t_{1-\alpha/2}\hat{\sigma}_X/\sqrt{n} < \mu_X \leq \hat{\mu}_X - t_{\alpha/2}\hat{\sigma}_X/\sqrt{n}\} = 1 - \alpha = \gamma \tag{18.4.11}$$

where  $t_{\alpha/2} = P(T \leq \alpha/2)$  and  $t_{1-\alpha/2} = P(T \leq 1 - \alpha/2)$  and  $t_{\alpha/2} = -t_{1-\alpha/2}$  because of the symmetry of the Student- $t$  density. From the Student- $t$  table (Table 18.4.2),  $t_{1-\alpha/2}$  is extracted for  $n-1$  degrees of freedom and probability  $(1 - \alpha/2)$ . Equation (18.4.11) is to be read as the probability that the random interval  $(\hat{\mu}_X - t_{1-\alpha/2}\hat{\sigma}_X/\sqrt{n}, \hat{\mu}_X - t_{\alpha/2}\hat{\sigma}_X/\sqrt{n})$  includes the true population mean  $\mu_X$  is equal to  $(1 - \alpha)$ .

These confidence limits are shown in Fig. 18.4.2 on a Student- $t$  density with  $(n - 1)$  degrees of freedom.

**Example 18.4.2** The tensile strength of six samples of a material is given by 226, 223, 225, 227, 224, and 222 lb. The estimated mean  $\hat{\mu}_X = 224.5$  lb. The standard error,  $\hat{\sigma}_X = 1.871$  lb. Assuming a Gaussian distribution for the samples, the  $t_{0.95}$  value for  $(6-1) = 5$  degrees of freedom from Table 18.4.2 is given by  $t_{0.95} = 2.01505$ . Hence the

**TABLE 18.4.2**

<i>v</i>	$F_T(t) = 0.95$	$F_T(t) = 0.975$	$F_T(t) = 0.995$
1	6.31375	12.70620	63.65674
2	2.91999	4.30265	9.92484
3	2.35336	3.18245	5.84091
4	2.13185	2.77645	4.60409
5	2.01505	2.57058	4.03214
6	1.94318	2.44691	3.70743
7	1.89458	2.36462	3.49948
8	1.85955	2.30600	3.35539
9	1.83311	2.26216	3.24984
10	1.81246	2.22814	3.16927
11	1.79588	2.20099	3.10581
12	1.78229	2.17881	3.05454
13	1.77093	2.16037	3.01228
14	1.76131	2.14479	2.97684
15	1.75305	2.13145	2.94671
16	1.74588	2.11991	2.92078
17	1.73961	2.10982	2.89823
18	1.73406	2.10092	2.87844
19	1.72913	2.09302	2.86093
20	1.72472	2.08596	2.84534

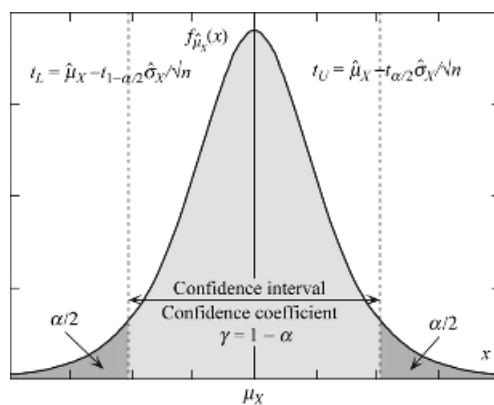
lower confidence limit  $t_L$  is

$$t_L = \hat{\mu}_X - t_{0.95} \hat{\sigma}_X / \sqrt{n} = 224.5 - 2.01505 \times \frac{1.871}{6} = 223.872$$

and the upper confidence limit  $t_U$  with  $t_{0.05} = -t_{0.95}$  is

$$t_U = \hat{\mu}_X - t_{0.05} \hat{\sigma}_X / \sqrt{n} = 224.5 + 2.01505 \times \frac{1.871}{6} = 225.128$$

Hence we are 90% confident that the true mean lies in the interval (223.872, 225.128) lb. On the other hand, if we are given that the true standard deviation  $\sigma_X = 1.871$  lb, then we

**FIGURE 18.4.2**

use the Gaussian distribution for confidence intervals given by Eq. (18.4.6), and the lower and upper confidence intervals are

$$z_L = \hat{\mu}_X - z_{0.95}\sigma_X/\sqrt{n} = 224.5 - 1.645 \times \frac{1.871}{6} = 223.987$$

$$z_U = \hat{\mu}_X - z_{0.05}\sigma_X/\sqrt{n} = 224.5 + 1.645 \times \frac{1.871}{6} = 225.013$$

The confidence interval (223.987, 225.013) for the Gaussian is a little tighter than (223.872, 225.128) as obtained for the Student-*t*, showing that the Student-*t* density is flatter than the Gaussian as shown in Fig. 7.9.1.

**Example 18.4.3** In a probability examination for a class of 16 students, the following scores were recorded:

$$X = [94 \ 70 \ 57 \ 77 \ 45 \ 43 \ 70 \ 67 \ 96 \ 89 \ 62 \ 98 \ 56 \ 68 \ 75 \ 71]$$

$$\text{Estimated mean } \hat{\mu}_X = \frac{1}{16} \sum_{i=1}^{16} X_i = 71.125$$

$$\text{Estimated variance } \hat{\sigma}_X^2 = \frac{1}{15} \sum_{i=1}^{16} (X_i - 71.125)^2 = 283.183$$

$$\text{Standard error } \hat{\sigma}_X = \sqrt{283.183} = 16.828$$

We have to find the 90% confidence limits for estimated mean  $\hat{\mu}_X = 71.125$ . From the Student-*t* tables (Table 18.4.2) the *t* value for 15 degrees of freedom and 95% probability is 1.753. Hence the lower confidence limit is given by

$$t_L = \hat{\mu}_X - t_{0.95}\hat{\sigma}_X/\sqrt{n} = 71.125 - 1.753 \times \frac{16.828}{16} = 63.75$$

and the upper confidence limit is

$$t_U = \hat{\mu}_X + t_{0.05}\hat{\sigma}_X/\sqrt{n} = 71.125 + 1.753 \times \frac{16.828}{16} = 78.5$$

Thus, we are 90% confident that the true mean will be in the interval (63.75, 78.5) or within  $\pm 7.375$  of the estimated mean 71.125.

*Confidence Interval for the Unknown Variance (Unknown Mean).* We have already seen in Eq. (18.4.10) that the random variable  $(n-1)\hat{\sigma}_X^2/\sigma_X^2$  is chi-square-distributed with  $(n-1)$  degrees of freedom. This density is not symmetric, and the interval estimate will not be centered at  $\sigma^2$ , unlike the Gaussian and Student-*t* densities. To determine the upper and lower confidence limits, we can write

$$P\left\{ \chi_{\alpha/2, (n-1)}^2 < \frac{(n-1)\hat{\sigma}_X^2}{\sigma_X^2} \leq \chi_{1-\alpha/2, (n-1)}^2 \right\} = 1 - \alpha = \gamma$$

from which it follows that

$$P\left\{ \frac{\chi_{\alpha/2, (n-1)}^2}{(n-1)\hat{\sigma}_X^2} < \frac{1}{\sigma_X^2} \leq \frac{\chi_{1-\alpha/2, (n-1)}^2}{(n-1)\hat{\sigma}_X^2} \right\} = \gamma$$

TABLE 18.4.3

Chi-Square Tables with $u = \chi^2$ Values Shown						
	$\gamma = 0.9$		$\gamma = 0.95$		$\gamma = 0.99$	
$v$	$F_U(u) = 0.05$	$F_U(u) = 0.95$	$F_U(u) = 0.025$	$F_U(u) = 0.975$	$F_U(u) = 0.005$	$F_U(u) = 0.995$
1	0.00393	3.84146	0.00098	5.02389	0.00004	7.87944
2	0.10259	5.99146	0.05064	7.37776	0.01003	10.59663
3	0.35185	7.81473	0.21580	9.34840	0.07172	12.83816
4	0.71072	9.48773	0.48442	11.14329	0.20699	14.86026
5	1.14548	11.0705	0.83121	12.83250	0.41174	16.74960
6	1.63538	12.59159	1.23734	14.44938	0.67573	18.54758
7	2.16735	14.06714	1.68987	16.01276	0.98926	20.27774
8	2.73264	15.50731	2.17973	17.53455	1.34441	21.95495
9	3.32511	16.91898	2.70039	19.02277	1.73493	23.58935
10	3.94030	18.30704	3.24697	20.48306	2.15586	25.18818
11	4.57481	19.67506	3.81575	21.91996	2.60322	26.75685
12	5.22603	21.02601	4.40379	23.33660	3.07382	28.29952
13	5.89186	22.36199	5.00875	24.73555	3.56503	29.81947
14	6.57063	23.68475	5.62873	26.11891	4.07467	31.31935
15	7.26094	24.99576	6.26214	27.48836	4.60092	32.80132
16	7.96165	26.29620	6.90766	28.84532	5.14221	34.26719
17	8.67176	27.58709	7.56419	30.19099	5.69722	35.71847
18	9.39046	28.86928	8.23075	31.52636	6.26480	37.15645
19	10.11701	30.14351	8.90652	32.85231	6.84397	38.58226
20	10.85081	31.41042	9.59078	34.16959	7.43384	39.99685

and finally

$$P\left\{\frac{(n-1)\hat{\sigma}_X^2}{\chi_{1-\alpha/2,(n-1)}^2} < \sigma_X^2 \leq \frac{(n-1)\hat{\sigma}_X^2}{\chi_{\alpha/2,(n-1)}^2}\right\} = \gamma \quad (18.4.12)$$

where  $\chi_{1-\alpha/2,(n-1)}^2$  and  $\chi_{\alpha/2,(n-1)}^2$  are constants to be determined from the chi-square table (Table 18.4.3) for  $(n-1)$  degrees of freedom at probability levels of  $1-\alpha/2$  and  $\alpha/2$  respectively. Equation (18.4.12) is to be read as the probability that the random interval

$$\left(\frac{(n-1)\hat{\sigma}_X^2}{\chi_{1-\alpha/2,(n-1)}^2}, \frac{(n-1)\hat{\sigma}_X^2}{\chi_{\alpha/2,(n-1)}^2}\right)$$

includes the population variance  $\sigma_X^2$  is equal to  $\gamma = (1-\alpha)$ .

These confidence limits are shown in Fig. 18.4.3 on a chi-square density with  $(n-1)$  degrees of freedom.

**Example 18.4.4** We will find the 90% confidence interval for the sample variance  $\hat{\sigma}_X^2 = 283.183$  calculated in Example 18.4.3. From the chi-square table (Table 18.4.3), the  $\chi^2$  value for probability 0.05 and 15 degrees of freedom is 7.261, and the corresponding value for probability of 0.95 and 15 degrees of freedom is 24.996. Hence, from Eq. (18.4.12), the upper confidence limit is given by

$$\chi_U^2 = \frac{(n-1)\hat{\sigma}_X^2}{\chi_{\alpha/2,(n-1)}^2} = \frac{15 \times 283.183}{7.261} = 585.008$$

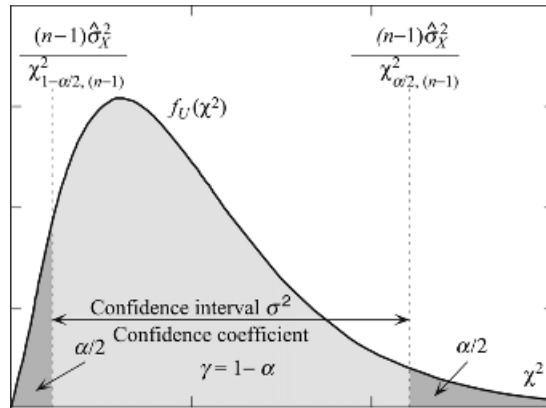


FIGURE 18.4.3

and the lower confidence limit is given by

$$\chi_L^2 = \frac{(n-1)\hat{\sigma}_x^2}{\chi_{1-\alpha/2, (n-1)}^2} = \frac{15 \times 283.183}{24.996} = 169.937$$

The corresponding 90% confidence limits for the standard error  $\hat{\sigma}_x = 16.828$  are ( $\sqrt{169.937} = 13.036$ ,  $\sqrt{585.008} = 24.187$ ). This result is to be interpreted as the probability of the standard deviation  $\sigma_x$  of the students' scores will be in the interval (13.036, 24.187) is 90%.

*Confidence Interval for the Coefficients of Simple Linear Regression.* Since the residual variance  $\hat{\sigma}_v^2$  is chi-square-distributed with  $(n-2)$  degrees of freedom the estimated variances  $\hat{\sigma}_{\hat{a}_1}^2$  and  $\hat{\sigma}_{\hat{a}_2}^2$  for the regression coefficients  $\hat{a}_1$  and  $\hat{a}_2$  of the regression line  $\hat{y}_i = \hat{a}_1 + \hat{a}_2 x_i$ ,  $i = 1, \dots, n$  are also chi-square-distributed with  $(n-2)$  degrees of freedom as given in Eqs. (18.3.51). The terms  $\hat{a}_1$  and  $\hat{a}_2$  are unbiased estimators of  $a_1$  and  $a_2$ , and hence the quantities  $(\hat{a}_1 - a_1)/(\hat{\sigma}_{\hat{a}_1}/\sqrt{n-2})$  and  $(\hat{a}_2 - a_2)/(\hat{\sigma}_{\hat{a}_2}/\sqrt{n-2})$  are Student-*t*-distributed with  $(n-2)$  degrees of freedom. From Eqs. (18.4.11), we can form equations for 100(1 -  $\alpha$ )% confidence intervals

$$\begin{aligned} P\left\{ t_{\alpha/2, (n-2)} < \frac{(\hat{a}_1 - a_1)}{\hat{\sigma}_{\hat{a}_1}} \leq t_{1-\alpha/2, (n-2)} \right\} &= 1 - \alpha = \gamma \\ P\left\{ t_{\alpha/2, (n-2)} < \frac{(\hat{a}_2 - a_2)}{\hat{\sigma}_{\hat{a}_2}} \leq t_{1-\alpha/2, (n-2)} \right\} &= 1 - \alpha = \gamma \end{aligned} \quad (18.4.13)$$

where  $t_{1-\alpha/2, (n-2)}$  is obtained from the *t* table (Table 18.4.2) for  $(n-2)$  degrees of freedom and probability  $(1 - \alpha/2)$ . Equation (18.4.13) can be stated more formally as

$$\begin{aligned} P\{\hat{a}_1 - t_{1-\alpha/2, (n-2)}\hat{\sigma}_{\hat{a}_1} < a_1 \leq \hat{a}_1 - t_{\alpha/2, (n-2)}\hat{\sigma}_{\hat{a}_1}\} &= 1 - \alpha \\ P\{\hat{a}_2 - t_{1-\alpha/2, (n-2)}\hat{\sigma}_{\hat{a}_2} < a_2 \leq \hat{a}_2 - t_{\alpha/2, (n-2)}\hat{\sigma}_{\hat{a}_2}\} &= 1 - \alpha \end{aligned} \quad (18.4.14)$$

and hence

$$\begin{aligned} \text{Confidence interval for } a_1 &= (\hat{a}_1 - t_{1-\alpha/2, (n-2)}\hat{\sigma}_{\hat{a}_1}, \hat{a}_1 - t_{\alpha/2, (n-2)}\hat{\sigma}_{\hat{a}_1}) \\ \text{Confidence interval for } a_2 &= (\hat{a}_2 - t_{1-\alpha/2, (n-2)}\hat{\sigma}_{\hat{a}_2}, \hat{a}_2 - t_{\alpha/2, (n-2)}\hat{\sigma}_{\hat{a}_2}) \end{aligned} \quad (18.4.15)$$

**Example 18.4.5** We will now find 90% confidence intervals for the estimated regression coefficients of Example 18.3.3. The estimated regression coefficients are given by  $\hat{a}_1 = 1.574$  and  $\hat{a}_2 = 0.404$ , with estimated variances  $\hat{\sigma}_{\hat{a}_1}^2 = 0.0704$   $\hat{\sigma}_{\hat{a}_2}^2 = 0.0699$ . The standard errors for  $\hat{a}_1$  and  $\hat{a}_2$  are  $\hat{\sigma}_{\hat{a}_1} = 0.265$  and  $\hat{\sigma}_{\hat{a}_2} = 0.264$ . From the  $t$  table (Table 18.4.2), for 18 degrees of freedom and probability of 0.95,  $t_{0.95,18} = 1.7341$ . From Eq. (18.4.15) we can compute the 90% confidence interval for  $\hat{a}_1$  as

$$\left. \begin{array}{l} L_{\hat{a}_1} \\ U_{\hat{a}_1} \end{array} \right\} = 1.574 - 1.734 \times 0.265 = \left\{ \begin{array}{l} 1.115 \\ 2.034 \end{array} \right.$$

and the 90% confidence interval for  $\hat{a}_2$  is

$$\left. \begin{array}{l} L_{\hat{a}_2} \\ U_{\hat{a}_2} \end{array} \right\} = 0.404 - 1.734 \times 0.264 = \left\{ \begin{array}{l} 1.115 \\ -0.054 \end{array} \right.$$

*Confidence Region for Simple Linear Regression Line.* The estimated regression line is given as follows:

$$\hat{y}_i = \hat{a}_1 + \hat{a}_2 x_i, \quad i = 1, \dots, n \quad (18.3.49)$$

Since  $\hat{y}_i$  is a Gaussian-distributed unbiased estimator of  $y_i$  with estimated variance  $\hat{\sigma}_{\hat{y}_i}^2$  given by Eq. (18.3.54), the quantity  $(\hat{y}_i - y_i)/(\hat{\sigma}_{\hat{y}_i}/\sqrt{n-2})$  is Student- $t$ -distributed with  $(n-2)$  degrees of freedom. Hence, we can form equations for  $100(1-\alpha)\%$  confidence intervals as

$$P\left\{ t_{\alpha/2,(n-2)} < \frac{\hat{y}_i - y_i}{\hat{\sigma}_{\hat{y}_i}} \leq t_{1-\alpha/2,(n-2)} \right\} = 1 - \alpha = \gamma, \quad i = 1, \dots, n \quad (18.4.16)$$

where  $t_{1-\alpha/2,(n-2)}$  is obtained from the  $t$  table (Table 18.4.2) for  $(n-2)$  degrees of freedom and probability  $(1-\alpha/2)$ . Equation (18.4.16) can be rewritten as

$$P\{\hat{y}_i - t_{1-\alpha/2,(n-2)}\hat{\sigma}_{\hat{y}_i} < y_i \leq \hat{y}_i + t_{\alpha/2,(n-2)}\hat{\sigma}_{\hat{y}_i}\} = 1 - \alpha, \quad i = 1, \dots, n \quad (18.4.17)$$

and the lower bound of the  $100(1-\alpha)\%$  confidence region for  $y_i$  is given by

$$L_{\hat{y}_i} = \hat{y}_i - t_{1-\alpha/2,(n-2)}\hat{\sigma}_{\hat{y}_i} \quad (18.4.18)$$

with the upper bound of the region given by

$$U_{\hat{y}_i} = \hat{y}_i + t_{\alpha/2,(n-2)}\hat{\sigma}_{\hat{y}_i} \quad (18.4.19)$$

**Example 18.4.6** In Example 18.4.5 we will find the 90% confidence region for the regression line given by

$$\hat{y}_i = 1.574 + 0.404x_i, \quad i = 1, \dots, 20$$

From Example 18.3.3 the variance of the regression line is given by

$$\hat{\sigma}_{\hat{y}_i}^2 = 0.263 \left[ \frac{1}{20} + \frac{(x_i - 0.905)^2}{3.7665} \right], \quad i = 1, \dots, 20$$

This value of  $\hat{\sigma}_{\hat{y}_i}^2$  can be substituted in Eq. (18.4.18) to obtain the lower bound of the 90% confidence region

$$L_{\hat{y}_i}(x_i) = 1.574 + 0.404x_i - 1.7341 \left\{ 0.263 \left[ \frac{1}{20} + \frac{(x_i - 0.905)^2}{3.7665} \right] \right\}^{1/2}, \quad i = 1, \dots, 2$$

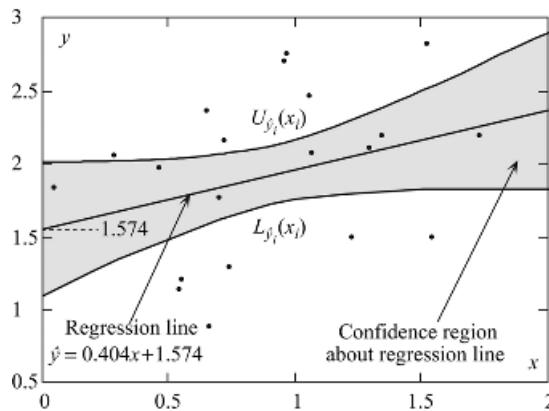


FIGURE 18.4.4

and the upper bound of the region can be obtained from Eq. (18.4.19):

$$U_{\hat{y}_i}(x_i) = 1.574 + 0.404x_i - 1.7341 \left\{ 0.263 \left[ \frac{1}{20} + \frac{(x_i - 0.905)^2}{3.7665} \right] \right\}^{1/2}, \quad i = 1, \dots, 20$$

The confidence region is shown in Fig. 18.4.4 along with the regression line  $\hat{y}_i$  and the data points.

## 18.5 HYPOTHESIS TESTING (BINARY)

We have discussed some estimation problems in the previous sections. However, many probability models require us to make decisions about populations on the basis of limited knowledge. For example, a physician has to make a decision based on knowledge of her patient whether to give a particular drug. Usually this decision is binary in nature, that is, to give or not to give. Thus, we have a decision problem as opposed to an estimation problem. Such decision problems come under the category of *hypothesis testing*. We postulate a carefully formed assertion based on available information. This assertion is called the *null hypothesis*,  $H_0$ . If we are interested only in accepting it or rejecting it at a significance level  $\alpha$ , then it is called *significance testing*. On the other hand, we may be interested in postulating an *alternate hypothesis*  $H_1$  and finding methods of accepting  $H_0$  or  $H_1$ . This is called *hypothesis testing*. For example, we may be interested in testing the null hypothesis  $H_0$  that the mean time between failures  $\theta$  (MTBF) of a memory chip is  $\theta_0$  hours. The alternate hypothesis may be  $\theta = \theta_1$  or  $\theta < \theta_0$ . We are not particularly interested in the alternate hypothesis  $\theta > \theta_0$  because it does not matter whether the chip overperforms. On the other hand, in the communication receiver we can hypothesize that the noise variance is  $\theta_0$ . We are now interested in the alternate hypothesis  $\theta > \theta_0$  because we do not want the variance to be too high, while we are not interested in  $\theta < \theta_0$  because it does not affect the performance for low variances.

The experiment on which we test a null hypothesis is based on an  $n$  random vector  $\mathbf{X}$ ,  $\{X_i, i = 1, \dots, n\}$  from a population as in the estimation problem. The  $n$  random vector will span an  $n$ -dimensional Euclidean space  $\mathbf{R}^n$ . We divide this  $n$ -space of observations into two mutually exclusive regions,  $R(H_0)$  and  $R(H_1)$ . If the observed vector  $\mathbf{x}$ ,  $\{x_i, i = 1, \dots, n\}$ , called the *test statistic*, lies in the space  $R(H_1)$ , we reject the null

**TABLE 18.5.1**

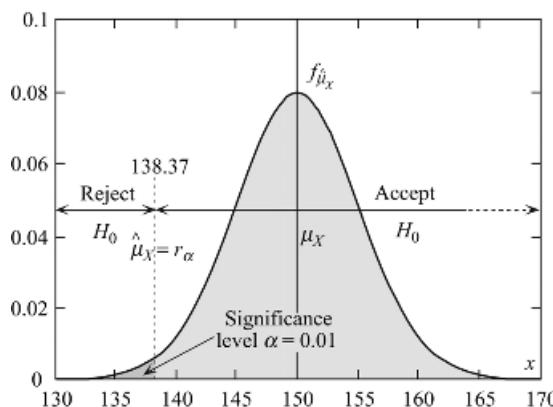
True Situation	Accept or Reject	Decision Type
$H_0$ true, $H_1$ false	Accept $H_0$	Correct decision—no error
$H_0$ true, $H_1$ false	Reject $H_0$	Type I error— $\alpha$
$H_0$ false, $H_1$ true	Accept $H_0$	Type II error— $\beta$
$H_0$ false, $H_1$ true	Reject $H_0$	Correct decision—no error

hypothesis  $H_0$ . On the other hand, if it falls in the region  $R(H_0)$ , we may not reject  $H_0$ . The region  $R(H_0)$  is called the *acceptance region*, and  $R(H_1)$  is called the *critical region* or *rejection region*. Whether we accept or reject the null hypothesis, there is always the possibility of errors. The error created when  $H_0$  is rejected when it is true is called *type I error* and is denoted by  $\alpha$ , which is also known as the *significance level*. Other names for this error are *false alarm* and *producer's risk*. The error created when  $H_0$  is accepted when it is false is called *type II error* and is denoted by  $\beta$ . Other terms for this error are *missed alarm* and *consumer's risk*. The four possibilities, two of which are correct decisions and the other two, erroneous decisions, are shown in Table 18.5.1.

A *simple hypothesis* is one in which the parameters are specified exactly. If they are not, it is called a *composite hypothesis*. A null hypothesis is almost always simple. For example,  $H_0 = \theta_0$  is simple whereas  $H_1 > \theta_0$  is composite. In testing hypotheses we usually control the significance level  $\alpha$  a priori and try to minimize the type II error  $\beta$ .

**Example 18.5.1 (Significance Testing)** A pharmaceutical company wants to test the effectiveness of the new antihypertensive drug that they are developing. A group of  $n = 25$  volunteers are selected with a mean systolic pressure of  $\mu_{25} = 150$  mmHg with a standard deviation of  $\sigma_{25} = 25$  mmHg. The company wants to see whether the new drug can bring the systolic pressure below 140 mmHg. We have to design a test to determine whether the drug is effective at a significance level of  $\alpha = 0.01$ .

We will assume that the systolic pressure is a Gaussian-distributed random variable  $X$  with mean  $\mu_X = 150$  and standard deviation  $\sigma_X = 25$ . Hence the sample mean  $\hat{\mu}_{25}(X)$  of systolic pressures after taking the drug is Gaussian-distributed with mean 150 and standard deviation  $\sigma_{\hat{\mu}_X} = 25/\sqrt{25} = 5$  as shown in Fig. 18.5.1. We have to form the null

**FIGURE 18.5.1**

hypothesis  $H_0$  carefully. If we choose the null hypothesis that the sample mean  $\hat{\mu}_{25}(X)$  after taking the drug is the same as the mean value of 150, then the drug has minimal effect on the systolic pressure. If we now choose a region  $R$  such that

$$P\{\hat{\mu}_{25}(X) \in R\} = P\{\hat{\mu}_{25}(X) \leq r_\alpha\} = 0.01$$

and the null hypothesis is rejected at a significance level of 1%, then we can conclude that the drug is effective in controlling systolic blood pressure. From Eq. (18.4.2) the test statistic is

$$Z = \frac{\hat{\mu}_X(X) - \mu_X}{\sigma_{\hat{\mu}_X}} = \frac{\hat{\mu}_X(X) - \mu_X}{\sigma_X/\sqrt{n}}$$

and we can write a one-sided equation:

$$P\left\{\frac{\hat{\mu}_X - \mu_X}{\sigma_X/\sqrt{n}} \leq z_\alpha\right\} = P\left\{\frac{\hat{\mu}_X - 150}{5} \leq z_\alpha\right\} = \alpha = 0.01$$

Referring to Table 18.4.1 the value of  $z_\alpha$  corresponding to a probability of 0.01 is  $-2.326$ . Hence

$$\frac{\hat{\mu}_X - 150}{5} = -2.326 \quad \text{or} \quad \hat{\mu}_X = 150 - 5 \times 2.326 = 138.37$$

so that  $r_\alpha = 138.37$  as shown in Fig. 18.5.1. Thus, we can reject the hypothesis  $H_0$  that the drug is ineffective in controlling the systolic blood pressure if the sample mean of the pressure measurements drops below 139.

On the other hand, if we increase the significance level to 2.5%, then  $z_{0.025} = -1.96$  and

$$\frac{\hat{\mu}_X - 150}{5} = -1.96 \quad \text{or} \quad \hat{\mu}_X = 150 - 5 \times 1.96 = 140.2$$

and we cannot reject the hypothesis at the higher significance level of 2.5% and conclude that the drug is ineffective at that level. Thus, the significance level has to be fixed before the test begins.

We shall now consider several cases of testing hypotheses. The form of the density function of a random variable  $X$  is a known function  $f_X(x, \theta)$  that depends on a parameter  $\theta$ .

*Case 1:  $H_0 = \theta_0$  and  $H_1 = \theta_1$ .* Assuming a Gaussian random variable, consider a simple hypothesis  $H_0 = \theta_0$  against a simple alternate  $H_1 = \theta_1$ . The decision to accept or reject  $H_0$  is dependent on the test statistic  $x$  obtained from the population. The probability densities  $f_X(x, \theta_0)$  and  $f_X(x, \theta_1)$  and the error regions  $\alpha$  and  $\beta$  are shown in Fig. 18.5.2. If  $x < c$ , the null hypothesis  $H_0$  will be accepted, and if  $x > c$ ,  $H_0$  will be rejected. The critical region  $R(H_1)$  is  $x > c$ . The type I error  $\alpha$  depends on the hypothesis under test  $H_0$ , and the type II error  $\beta$  depends on both  $H_0$  and the alternate hypothesis  $H_1$ . We can also see from Fig. 18.5.2 that we cannot minimize both types of error simultaneously since decreasing  $\alpha$  increases  $\beta$  and vice versa.

*Case 2:  $H_0 = \theta_0$  and  $H_1 = \theta > \theta_0$  (One-Sided Test).* In this case we have a simple hypothesis  $H_0 = \theta_0$  and a composite alternate  $H_1 = \theta > \theta_0$ . The acceptance region for  $H_0$  is  $x < c$ . Here  $\theta$  ranges from  $\theta_0$  to  $\infty$  and the error probability  $\beta$

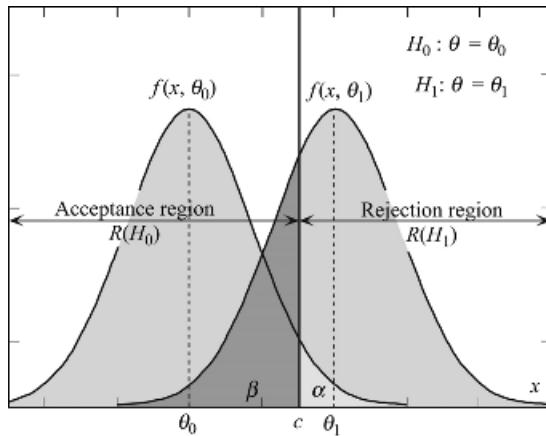


FIGURE 18.5.2

will be a function of  $\theta$ . The Gaussian probability densities along with the errors  $\alpha$  and  $\beta(\theta)$  are shown in Fig. 18.5.3. The acceptance and rejection regions are also shown in figure. The critical region will still be  $x > c$ , but the type II error function  $\beta(\theta)$  will depend on the closeness of  $\theta$  to  $\theta_0$ . The function  $\beta(\theta)$  is called the *operating characteristic* (OC) (Fig. 18.5.4) of the test. The OC shows the probability of a wrong decision. It can be visualized by sliding the density function  $f_X(x, \theta)$  under  $H_1$  to the left starting from  $\theta_0$  and integrating it from  $-\infty$  to  $c$ , or

$$\beta(\theta) = \int_{-\infty}^c f_X(x, \theta) dx = \int_{-\infty}^c \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(x-\theta)/\sigma]^2} dx \quad (18.5.1)$$

The complementary function  $P(\theta) = 1 - \beta(\theta)$  is called the *power of the test*. It shows the probability of a correct decision. The operating characteristic (OC) is shown in Fig. 18.5.4.

*Case 3:  $H_0 = \theta_0$  and  $H_1 \neq \theta_0$  (Two-Sided Test).* Here, we have a simple hypothesis  $H_0 = \theta_0$  and a composite alternate  $H_1 = \{\theta \neq \theta_0\}$ . The acceptance region is  $c_1 < x < c_2$ . The significance level  $\alpha$  is split between the two tails of the

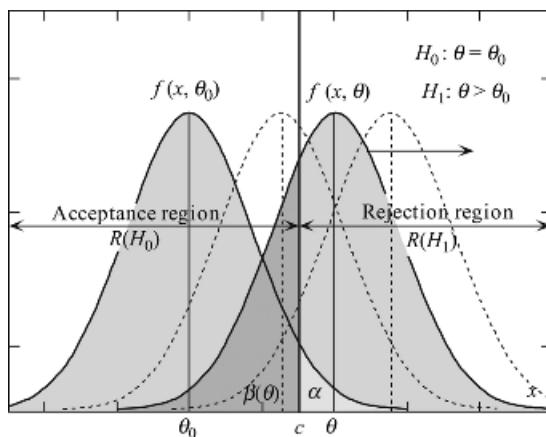


FIGURE 18.5.3

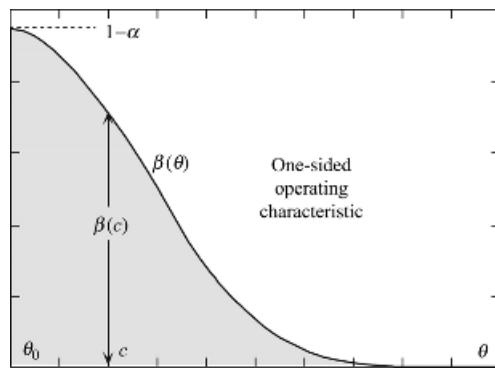


FIGURE 18.5.4

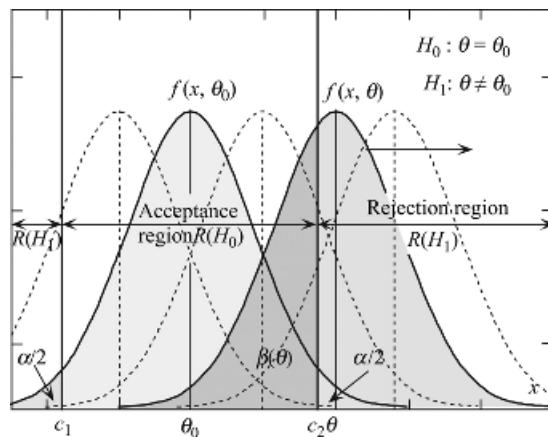


FIGURE 18.5.5

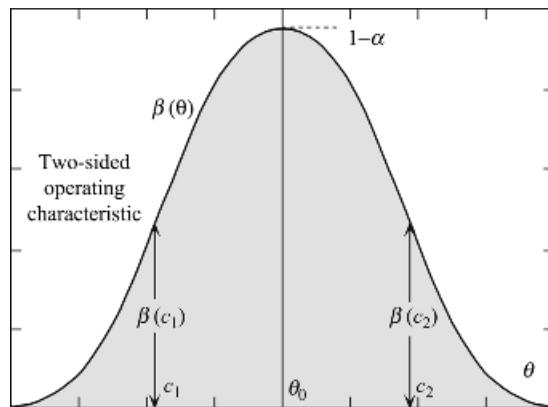


FIGURE 18.5.6

density under  $H_0$  so that  $P\{\hat{\mu}(x) < z_{\alpha/2}\} = c_1$  and  $P\{\hat{\mu}(x) < z_{1-\alpha/2}\} = c_2$ . Here  $\theta$  ranges from  $c_1$  to  $c_2$ , and the error probability  $\beta$  will be a function of  $\theta$ . The Gaussian probability densities along with the errors  $\alpha$  and  $\beta(\theta)$  are shown in Fig. 18.5.5.

The critical region is  $x < c_1$  and  $x > c_2$ , but the type II error function  $\beta(\theta)$  (OC) extends from  $-\infty$  to  $\infty$  (Fig. 18.5.6) and the value will depend on the closeness of  $\theta$  to  $\theta_0$ . It can again be visualized by sliding the density function  $f_X(x, \theta)$  under  $H_1$  to the left, starting from  $c_1$  and integrating it from  $c_1$  to  $c_2$ :

$$\beta(\theta) = \int_{c_1}^{c_2} f_X(x, \theta) dx = \int_{c_1}^{c_2} \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(x-\theta)/\sigma]^2} dx \quad (18.5.2)$$

The acceptance region is also shown in Fig. 18.5.5 and the two-sided OC curve, in Fig. 18.5.6.

### **Procedure for Hypothesis Testing**

The following procedures are established for credible testing of hypotheses:

1. Null hypothesis  $H_0$  must be carefully formulated.
2. Define the experiment to be performed.
3. Choose a proper test statistic.
4. Establish the significance level  $\alpha$ .
5. Choose a value for type II error  $\beta$ .
6. With  $\alpha$  and  $\beta$  specified, determine a suitable sample size  $n$ .

We will now apply hypothesis testing to the mean and variance of a population. We will have  $n$  realizations  $\{x_1, \dots, x_n\}$  of a Gaussian random variable  $X$  whose mean  $\mu$  we want to hypothesize. The sample mean  $\hat{\mu}$  is, as usual,  $\hat{\mu} = (1/n) \sum_n x_i$ , and the variance of the sample mean  $\hat{\sigma}^2 = \sigma^2/n$ . The density of  $\hat{\mu}$  will be given by

$$f_{\hat{\mu}}(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma / \sqrt{n}} e^{-(1/2)[(x-\mu)/(\sigma/\sqrt{n})]^2} \quad (18.5.3)$$

These quantities will be used to form the test statistic.

### **Mean $\mu$ —known Variance $\sigma$**

We will now test the hypothesis that the mean  $\mu$  is  $H_0 = \mu_0$  against the alternates: (1)  $H_1 = \mu < \mu_0$ , (2)  $H_1 = \mu > \mu_0$ , and (3)  $H_1 = \mu \neq \mu_0$ .

The test statistic we will use is the random variable

$$Z = \frac{\hat{\mu} - \mu_0}{\sigma / \sqrt{n}} \quad (18.5.4)$$

with a significance level of  $\alpha$ . Under the hypothesis  $H_0$ , the random variable  $Z$  is Gaussian with zero mean and unit variance. Note that  $z_\alpha$  for which  $P(Z \leq z_\alpha) = \alpha$  will be negative for  $\alpha < 0.5$  and  $z_{1-\alpha}$  for which  $P(Z \leq z_{1-\alpha}) = 1 - \alpha$  will be positive, and if the pdf is symmetric, then  $z_\alpha = -z_{1-\alpha}$ .

1.  $H_1 = \mu < \mu_0$ : Accept  $H_0$  if  $Z > z_\alpha$ . Or, from Eq. (18.5.4)

$$Z = \frac{\hat{\mu} - \mu_0}{\sigma / \sqrt{n}} > z_\alpha \quad \text{or} \quad \hat{\mu} > \mu_0 + z_\alpha \cdot \sigma / \sqrt{n} \quad (18.5.5)$$

The acceptance region  $R(H_0) = \{\hat{\mu} > \mu_0 + z_\alpha \cdot \sigma / \sqrt{n}\}$  and the rejection region  $R(H_1) = \{\hat{\mu} < \mu_0 + z_\alpha \cdot \sigma / \sqrt{n}\}$ . The critical point  $c = \mu_0 + z_\alpha \cdot \sigma / \sqrt{n}$  The OC curve

is obtained from Eq. (18.5.1):

$$\begin{aligned}\beta(\mu) &= \int_c^\infty f_{\hat{\mu}}(x, \mu) dx = \int_c^\infty \frac{1}{\sqrt{2\pi} \cdot \sigma / \sqrt{n}} e^{-(1/2)[(x-\mu)/(\sigma/\sqrt{n})]^2} dx \\ &= \frac{1}{2} \left\{ 1 - \operatorname{erf} \left[ \frac{1}{\sqrt{2}} \left( \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{1-\alpha} \right) \right] \right\}\end{aligned}\quad (18.5.6)$$

where

$$\operatorname{erf}(x) = \int_0^x \frac{2}{\sqrt{\pi}} e^{-\xi^2} d\xi \quad (18.5.7)$$

2.  $H_1 = \mu > \mu_0$ : In this case we accept  $H_0$  if  $Z > z_{1-\alpha}$ . Or

$$Z = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} < z_{1-\alpha} \quad \text{or} \quad \hat{\mu} > \mu_0 + z_{1-\alpha} \cdot \sigma/\sqrt{n} \quad (18.5.8)$$

The acceptance region  $R(H_0) = \{\hat{\mu} < \mu_0 + z_{1-\alpha} \cdot \sigma/\sqrt{n}\}$  and the rejection region  $R(H_1) = \{\hat{\mu} > \mu_0 + z_{1-\alpha} \cdot \sigma/\sqrt{n}\}$ . The critical point  $c = \mu_0 + z_{1-\alpha} \cdot \sigma/\sqrt{n}$

The OC curve is obtained from Eq. (18.5.1):

$$\begin{aligned}\beta(\mu) &= \int_{-\infty}^c f_{\hat{\mu}}(x, \mu) dx = \int_{-\infty}^c \frac{1}{\sqrt{2\pi} \cdot \sigma / \sqrt{n}} e^{-(1/2)[(x-\mu)/(\sigma/\sqrt{n})]^2} dx \\ &= \frac{1}{2} \left\{ 1 + \operatorname{erf} \left[ \frac{1}{\sqrt{2}} \left( \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{1-\alpha} \right) \right] \right\}\end{aligned}\quad (18.5.9)$$

where  $\operatorname{erf}(x)$  is given by Eq. (18.5.7)

3.  $H_1 = \mu \neq \mu_0$ : In the previous two cases we had one-sided hypothesis testing, whereas here we will have two-sided hypothesis testing. As seen previously, the significance level  $\alpha$  is split evenly between the two tails of the density under  $H_0$  so that the probability is  $\alpha/2$  on each tail. We now accept  $H_0$  if  $z_{\alpha/2} < Z < z_{1-\alpha/2}$ , or

$$z_{\alpha/2} < \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} < z_{1-\alpha/2} \quad \text{or} \quad \mu_0 + z_{\alpha/2} \cdot \sigma/\sqrt{n} < \hat{\mu} < \mu_0 + z_{1-\alpha/2} \cdot \sigma/\sqrt{n} \quad (18.5.10)$$

The acceptance region  $R(H_0) = \{\mu_0 + z_{\alpha/2} \cdot \sigma/\sqrt{n} < \hat{\mu} < \mu_0 + z_{1-\alpha/2} \cdot \sigma/\sqrt{n}\}$  and the rejection region  $R(H_1) = \{\hat{\mu} < \mu_0 + z_{\alpha/2} \cdot \sigma/\sqrt{n} \quad \hat{\mu} > \mu_0 + z_{1-\alpha/2} \cdot \sigma/\sqrt{n}\}$ . The critical points are  $c_1 = \mu_0 + z_{\alpha/2} \cdot \sigma/\sqrt{n}$  and  $c_2 = \mu_0 + z_{1-\alpha/2} \cdot \sigma/\sqrt{n}$ . The OC curve is obtained from Eq. (18.5.2):

$$\begin{aligned}\beta(\mu) &= \int_{c_1}^{c_2} f_{\hat{\mu}}(x, \mu) dx = \int_{c_1}^{c_2} \frac{1}{\sqrt{2\pi} \cdot \sigma / \sqrt{n}} e^{-(1/2)[(x-\mu)/(\sigma/\sqrt{n})]^2} dx \\ &= \frac{1}{2} \left\{ \operatorname{erf} \left[ \frac{1}{\sqrt{2}} \left( \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{1-\alpha/2} \right) \right] - \operatorname{erf} \left[ \frac{1}{\sqrt{2}} \left( \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha/2} \right) \right] \right\}\end{aligned}\quad (18.5.11)$$

where  $\operatorname{erf}(x)$  is as given by Eq. (18.5.7).

### Mean $\mu$ —Unknown Variance $\sigma^2$

The procedure with an unknown variance is similar to the one with the known variance discussed earlier. We replace the unknown variance with the estimated variance

$\hat{\sigma}_X^2$  and use the Student- $t$  distribution for analysis. The test statistic is

$$T = \frac{\hat{\mu} - \mu_0}{\hat{\sigma}/\sqrt{n}} \quad (18.5.12)$$

and this is Student- $t$ -distributed with  $(n - 1)$  degrees of freedom. The critical regions for the three cases discussed earlier are

1.  $H_1 = \mu < \mu_0$

$$\text{Acceptance region: } R(H_0) = \left\{ \hat{\mu} > \mu_0 + \frac{t_{(n-1),\alpha} \cdot \hat{\sigma}}{\sqrt{n}} \right\} \quad (18.5.13)$$

$$\text{Rejection region: } R(H_1) = \left\{ \hat{\mu} < \mu_0 + \frac{t_{(n-1),\alpha} \cdot \hat{\sigma}}{\sqrt{n}} \right\}$$

2.  $H_1 = \mu > \mu_0$

$$\text{Acceptance region: } R(H_0) = \left\{ \hat{\mu} < \mu_0 + \frac{t_{(n-1),(1-\alpha)} \cdot \hat{\sigma}}{\sqrt{n}} \right\} \quad (18.5.14)$$

$$\text{Rejection region: } R(H_1) = \left\{ \hat{\mu} > \mu_0 + \frac{t_{(n-1),(1-\alpha)} \cdot \hat{\sigma}}{\sqrt{n}} \right\}$$

3.  $H_1 = \mu \neq \mu_0$

$$\text{Acceptance region: } R(H_0) = \left\{ \mu_0 + \frac{t_{(n-1),\alpha/2} \cdot \hat{\sigma}}{\sqrt{n}} < \hat{\mu} < \mu_0 + \frac{t_{(n-1),(1-\alpha/2)} \cdot \hat{\sigma}}{\sqrt{n}} \right\} \quad (18.5.15)$$

$$\text{Rejection region: } R(H_1) = \left\{ \hat{\mu} < \mu_0 + \frac{t_{(n-1),(\alpha/2)} \cdot \hat{\sigma}}{\sqrt{n}}; \hat{\mu} > \mu_0 + \frac{t_{(n-1),(1-\alpha/2)} \cdot \hat{\sigma}}{\sqrt{n}} \right\} \quad (18.5.16)$$

To determine the OC curves in these three cases we have to integrate a noncentral  $t$  density, and this is beyond the scope of our discussion. However, for large  $n$ , the  $t$  distribution becomes Gaussian and the OC curves can be approximated by Eqs. (18.5.6), (18.5.9), and (18.5.11).

**Example 18.5.2** The following diastolic blood pressure measurements were recorded for a patient on 10 different occasions. The pressure measurements are assumed to be Gaussian-distributed:

BP measurements 78 85 88 85 78 86 79 82 80 81

We want to find whether these measurements support the hypothesis that the pressure is less than 80 at the significance levels of (1) 1%, (2) 2.5%, and (3) 5%.

This problem is a little different from what we discussed earlier because both hypotheses are composite. The estimated mean value  $\hat{\mu} = \frac{1}{10} \sum (78 + \dots + 81) = 82.2$  and the estimated standard deviation  $\hat{\sigma} = 3.584$ . We choose the statistic given by Eq. (18.5.12)

$$T = \frac{\hat{\mu} - \mu_0}{\hat{\sigma}/\sqrt{n}}$$

which is Student- $t$ -distributed. The  $t$  values for 1%, 2.5%, and 5% significance levels for  $(10 - 1) = 9$  degrees of freedom from the Student- $t$  table (Table 18.4.2) are 2.8214,

2.2622, and 1.8331 respectively. We can now formulate the composite hypotheses as (1) null hypothesis  $H_0: \mu < \mu_0 = 80$ ; and (2) alternate hypothesis  $H_1: \mu \geq \mu_0 = 80$ .

The acceptance regions from Eq. (18.5.14) are

$$1\% \text{ significance: } R(H_0) = \{\hat{\mu} < \mu_0 + t_{(n-1), (1-\alpha)} \cdot \hat{\sigma}/\sqrt{n}\}$$

$$= 80 + 2.8214 \times \frac{3.5839}{\sqrt{10}} = 83.2$$

$$2.5\% \text{ significance: } R(H_0) = \hat{\mu} < \left(80 + 2.2622 \times \frac{3.5839}{\sqrt{10}}\right) = 82.564$$

$$5\% \text{ significance: } R(H_0) = \hat{\mu} < \left(80 + 1.8331 \times \frac{3.5839}{\sqrt{10}}\right) = 82.078$$

Since the estimated mean is  $\hat{\mu} = 82.2 < 82.564 < 83.2$ , we conclude that the hypothesis with the diastolic pressure is 80 at significance levels of 1% and 2.5% can be accepted. It cannot be accepted at the 5% level because  $\hat{\mu} = 82.2 > 82.078$ . We will not calculate the OC curve  $\beta(\mu)$  since this involves noncentral Student- $t$  distribution.

**Example 18.5.3** An energy specialist recommends that a homeowner could save at least \$15 a month in utilities if energy appliances were upgraded. Before the upgrade was installed, the mean monthly utility expense  $\mu_0$  was \$255 with standard deviation  $\sigma = \$17$ . After the upgrade, the mean monthly expenses (in U.S. dollars) for a 10-year period is shown below:

243    252    246    262    244    219    228    212    264    260

We have to test whether the upgrade has saved \$15 a month at significance levels of (1) 1%, (2) 2.5%, (3) 5%.

We will assume that the utility expenses are Gaussian-distributed and formulate the null hypothesis  $H_0$  as mean monthly expense  $\mu_0 = \$255$  and the alternate hypothesis  $H_1$  as mean monthly expense  $\mu_1 = \$240$ . If  $H_0$  is rejected and  $H_1$  is accepted at significance levels of  $\alpha = 0.01, 0.025, 0.05$ , then the homeowner would have saved \$15 a month. The estimated mean value  $\hat{\mu}$  of the average monthly expense  $\hat{\mu} = \frac{1}{10}[243 + \dots + 260] = 243$  with  $\sigma$  given as 17. The density functions of  $\hat{\mu}$  for  $H_0$  and  $H_1$  are given by

$$f_{\hat{\mu}}(x | H_0) = \frac{1}{\sqrt{2\pi} \cdot 17/\sqrt{10}} e^{-(1/2)[(x-255)/(17/\sqrt{10})]^2}$$

$$f_{\hat{\mu}}(x | H_1) = \frac{1}{\sqrt{2\pi} \cdot 17/\sqrt{10}} e^{-(1/2)[(x-240)/(17/\sqrt{10})]^2}$$

and are shown in Fig. 18.5.7.

The test statistic is chosen as the normalized Gaussian

$$Z = \frac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} = \frac{\hat{\mu} - 255}{17/\sqrt{10}}$$

and the simple hypotheses  $H_0: \mu = \mu_0 = 255$  and  $H_1: \mu = \mu_1 = 240$  at significance levels of 1%, 2.5%, and 5% can be tested.

The corresponding  $z$  values from the Gaussian table (Table 18.4.1) are  $z_{0.01} = -2.326$ ,  $z_{0.025} = -1.96$ , and  $z_{0.05} = -1.645$  respectively. The acceptance regions from Eq. (18.5.6) are

$$1\% \text{ significance: } R(H_0) = \{\hat{\mu} > \mu_0 + z_{\alpha} \cdot \sigma/\sqrt{n}\}$$

$$= 255 - 2.326 \times (17/\sqrt{10}) = 242.49$$

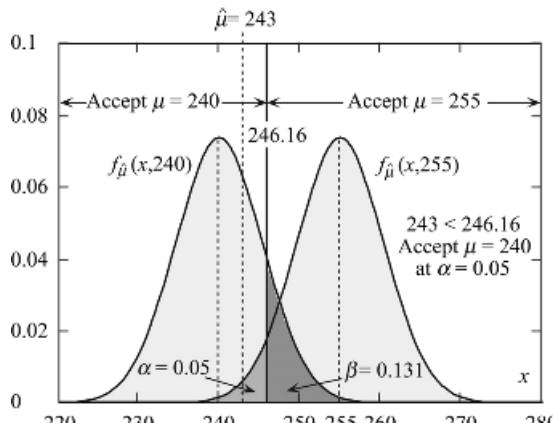


FIGURE 18.5.7

$$\begin{aligned}
 \text{2.5\% significance: } R(H_0) &= \{\hat{\mu} > \mu_0 + z_\alpha \cdot \sigma / \sqrt{n}\} \\
 &= 255 - 1.96 \times (17 / \sqrt{10}) = 244.47 \\
 \text{5\% significance: } R(H_0) &= \{\hat{\mu} > \mu_0 + z_\alpha \cdot \sigma / \sqrt{n}\} \\
 &= 255 - 1.645 \times (17 / \sqrt{10}) = 246.16
 \end{aligned}$$

The critical point  $c = \mu_0 + z_\alpha \cdot \sigma / \sqrt{n}$ .

The 5% significance region and the estimated mean  $\hat{\mu} = 243$  are shown in Fig. 18.5.7. Since the estimated mean is  $\hat{\mu} = 243 > 242.5$ , we conclude that the null hypothesis  $\mu_0 = \$255$  can be accepted at the 1% level with consequent no average monthly savings of \$15. On the other hand,  $\hat{\mu} = 243 < 244.47 < 246.16$ , and the null hypothesis is rejected at the 2.5% and 5% levels and the alternate hypothesis  $\mu_1 = 240$  is accepted, resulting in an average monthly savings of \$15.

The OC curve  $\beta(\mu, z_\alpha)$  for significance level  $\alpha$  is obtained from Eq. (18.5.6) as

$$\beta(\mu, z_\alpha) = \frac{1}{2} \left\{ 1 - \operatorname{erf} \left[ \frac{1}{\sqrt{2}} \left( \frac{\mu_0 - \mu}{\sigma / \sqrt{n}} + z_\alpha \right) \right] \right\} = \frac{1}{2} \left\{ 1 - \operatorname{erf} \left[ \frac{1}{\sqrt{2}} \left( \frac{255 - \mu}{5.4259} + z_\alpha \right) \right] \right\}$$

The type II errors for  $\alpha = 0.01, 0.025, 0.05$  are obtained by substituting  $\mu = 240$  and  $z_{0.01} = -2.326, z_{0.025} = -1.96, z_{0.05} = -1.645$  in the equation above. Thus,  $\beta(240, -2.326) = 0.3214, \beta(240, -1.96) = 0.2032, \beta(240, -1.645) = 0.1261$ . The operating characteristic curves shown in Fig. 18.5.8 are obtained from plotting  $\beta(\mu, z_\alpha)$  for 1%, 2.5%, and 5% values of  $z_\alpha$ .

*Number of Samples  $n$  for Given  $\alpha$  and  $\beta$ .* In testing hypotheses the usual practice is to fix the type I error  $\alpha$  at a value of 2.5% or 5% and minimize the type II error  $\beta$ . The only way to minimize  $\beta$  is to increase the sample size  $n$ . For the case of  $H_0: \mu = \mu_0$  and  $H_1: \mu < \mu_0$ , the critical point  $c$  for Gaussian distributions is found to be

$$c = \mu_0 + z_\alpha \cdot \sigma / \sqrt{n} \quad (18.5.17)$$

If the allowable  $\beta$  is also specified, we can then obtain the required number of samples  $n$ .

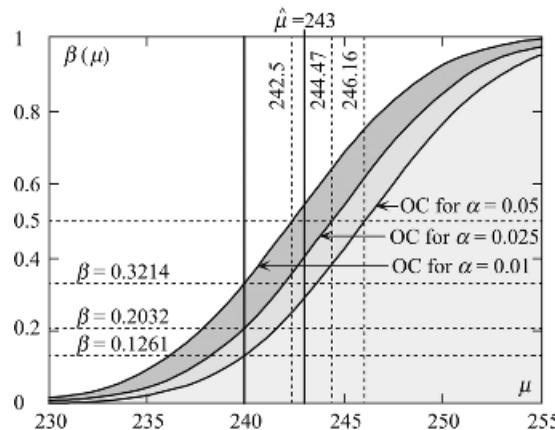


FIGURE 18.5.8

Referring to Fig. 18.5.9, we require  $n$  to be such that

$$\int_c^{\infty} \frac{1}{\sqrt{2\pi} \cdot \sigma / \sqrt{n}} e^{-(1/2)[(x-\mu_1)/(\sigma/\sqrt{n})]^2} dx \leq \beta \quad (18.5.18)$$

Substituting  $\xi = (x - \mu_1)/(\sigma/\sqrt{n})$  in this equation, we obtain

$$\int_{\frac{c-\mu_1}{\sigma/\sqrt{n}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1/2)\xi^2} d\xi \leq \beta \quad (18.5.19)$$

If the integral in Eq. (18.5.19) is to be  $\leq \beta$ , then the lower limit

$$\frac{c - \mu_1}{\sigma / \sqrt{n}} > z_{1-\beta} \quad (18.5.20)$$

where  $z_{1-\beta}$  is obtained from the Gaussian tables for a probability of  $\beta$ . Substituting  $c$  from Eq. (18.5.17) in Eq. (18.5.20) and simplifying, we have

$$\frac{\mu_0 - \mu_1}{\sigma / \sqrt{n}} > z_{1-\beta} - z_{\alpha}$$

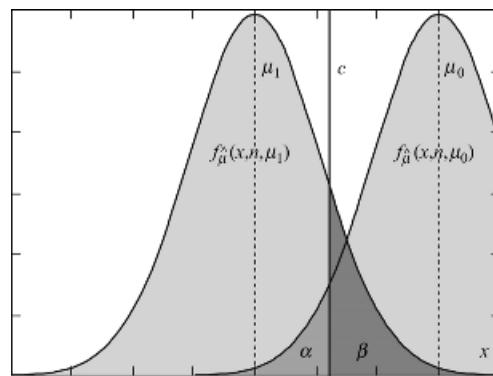
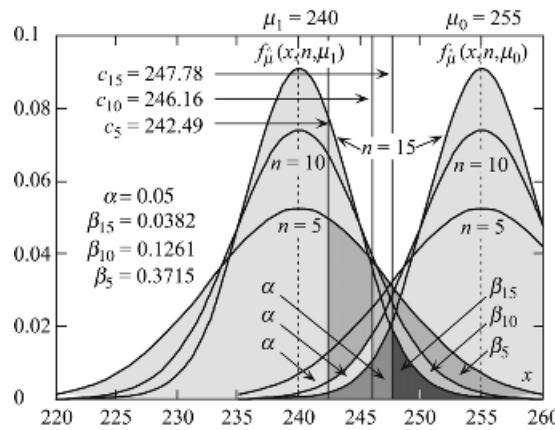


FIGURE 18.5.9



**FIGURE 18.5.10**

Thus the sample size  $n$  is given by

$$n > \frac{\sigma^2(z_{1-\beta} - z_\alpha)^2}{(\mu_0 - \mu_1)^2} \quad (18.5.21)$$

**Example 18.5.4** We will use the data from Example 18.5.2 and find the required sample size for  $\alpha = 0.05$ , and we are also given  $\sigma = 17$ ,  $\mu_0 = 255$ , and  $\mu_1 = 240$ . As the number of samples  $n$  is increased, the critical point  $c$  increases and  $\beta$  decreases and we stop at the value of  $n$  that gives the desired value of  $\beta$ . In this example, for  $n = 5, 10, 15$  we obtain  $c = 242.49, 246.16, 247.78$  and  $\beta = 0.3715, 0.1261, 0.0382$  respectively. These are shown in Fig. 18.5.10. If we choose the value of  $\beta = 0.1261$  from the previous example, we should get  $n = 10$ . From Gaussian tables  $z_{0.05} = -1.645$  and  $z_{1-\beta} = z_{0.8739} = 1.145$ . Substituting these values in Eq. (18.5.21), we obtain  $n > [17^2(1.145 + 1.645)^2]/(255 - 240)^2 = 9.9982$  or  $n = 10$ .

This corresponds to the 10 samples of Example 18.5.2.

## 18.6. BAYESIAN ESTIMATION

The previous approaches of estimating a parameter  $\theta$  in the probability density  $f(x, \theta)$  were based solely on the random samples  $\{X_1, \dots, X_n\}$  that are incorporated in the test statistic. This methodology is called the *classical* or the *objective* approach. However, in some problems we may have *prior* information about  $\theta$  that can be used to refine its estimate. For example, the classical estimate of the probability  $\theta$  in a binomial distribution  $b(k; n, \theta)$ , where  $k$  is the number of successes in  $n$  trials, is  $\hat{\theta} = k/n$ . We can then view  $\theta$  as the realization of a random variable  $\Theta$  whose density function  $f(\theta)$  is the prior information. The density of  $X$  will be a conditional density  $f(x | \theta)$ . The joint density of the sample  $\{X_1, \dots, X_n\}$  and the parameter  $\Theta$  can be given in terms of the conditional density as

$$f(x_1, \dots, x_n, \theta) = f(x_1, \dots, x_n \mid \theta) f(\theta) \quad (18.6.1)$$

Using Bayes' theorem, we can obtain the conditional density  $f(\theta | x_1, \dots, x_n)$  as

$$f(\theta | x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n, \theta)}{\text{Posterior}} = \frac{f(x_1, \dots, x_n, \theta)}{f(x_1, \dots, x_n)} f(\theta) \text{ Prior} \quad (18.6.2)$$

where the marginal density  $f(x_1, \dots, x_n)$  is given by

$$f(x_1, \dots, x_n) = \int_{-\infty}^{\infty} f(x_1, \dots, x_n, \theta) d\theta \quad (18.6.3)$$

The density  $f(\theta | x_1, \dots, x_n)$  is called the *posterior* density. What we have accomplished is modifying the *prior density*  $f(\theta)$  by knowledge of the joint density of the random samples  $\{X_1, \dots, X_n\}$  to the *posterior density*. This method of estimation is called the *subjective* approach to estimation or *Bayesian* estimation. It is used in medical imaging problems, which are discussed in Chapter 23.

We will now present Bayes' estimation of the probability  $p$  of an event where  $p$  is a realization of a random variable  $X$  with probability density function  $f_X(p)$  whose range is  $[0,1]$ . Prior estimates of  $p$  can be obtained from

$$\hat{p} = \int_0^1 p f_X(p) dp \quad (18.6.4)$$

To improve on the estimate of  $p$ , we conduct an experiment of tossing a die  $n$  times and observing the number of aces to be  $k$ . We can apply Bayes' theorem from Eq. (8.3.4) and write the posterior density as

$$f_X(p | B) = \frac{P(B | X = p) f_X(p)}{\int_0^1 P(B | X = p) f_X(p) dp} \quad (18.6.5)$$

where  $B = \{k \text{ aces in } n \text{ trials}\}$ . From binomial probability, we obtain

$$P(B | X = p) = \binom{n}{k} p^k (1-p)^{n-k} \quad (18.6.6)$$

Substituting Eq. (18.6.6) into Eq. (18.6.5), we obtain

$$f_X(p | B) = \frac{\binom{n}{k} p^k (1-p)^{n-k} f_X(p)}{\int_0^1 \binom{n}{k} p^k (1-p)^{n-k} f_X(p) dp} \quad (18.6.7)$$

The updated estimate of  $p$  can be obtained by substituting  $f_X(p | B)$  from Eq. (18.6.7) for  $f_X(p)$  in Eq. (18.6.4), and the result is

$$\hat{p}_{\text{post}} = \int_0^1 p f_X(p | B) dp \quad (18.6.8)$$

If we now assume that  $f_X(p)$  is uniformly distributed in  $\{0,1\}$  instead of a general distribution in the range  $[0,1]$ , Eq. (18.6.7) can be simplified. The integral

$$\int_0^1 p^m (1-p)^{n-k} dp = \frac{m!(n-k)!}{(n+m-k+1)!} \quad (18.6.9)$$

can be shown to be true using mathematical induction.

Substituting  $f_X(p) = 1$  and from Eq. (18.6.9), we can evaluate  $f_X(B)$  as

$$f_X(B) = \binom{n}{k} \int_0^1 p^k (1-p)^{n-k} \cdot 1 dp = \frac{n!}{k!(n-k)!} \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1} \quad (18.6.10)$$

we can express the conditional density  $f_X(p | B)$  as follows:

$$f_X(p | B) = \frac{\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \cdot 1}{\frac{1}{(n+1)}} = \frac{(n+1)!}{k!(n-k)!} p^k (1-p)^{n-k} \quad (18.6.11)$$

The posterior estimate for  $p$  is obtained from Eq. (18.6.8) as

$$\hat{p}_{\text{post}} = \frac{(n+1)!}{k!(n-k)!} \int_0^1 p^{k+1} (1-p)^{n-k} dp$$

and from Eq. (18.6.9)  $\hat{p}_{\text{post}}$  is given by

$$\hat{p}_{\text{post}} = \frac{(n+1)!}{k!(n-k)!} \frac{(k+1)!(n-k)!}{(n+2)!} = \frac{k+1}{n+2} \quad (18.6.12)$$

**Example 18.6.1** We have to estimate the parameter  $\Theta$  in the following problem

$$Y_i = \Theta + W_i, \quad i = 1, \dots, n$$

where  $\{W_i\}$  are independent identically distributed zero mean Gaussian random variables with variance  $\sigma_W^2$  and  $\Theta$  is also a Gaussian-distributed random variable with mean  $\mu_\Theta$  and variance  $\sigma_\Theta^2$ . We want to find the Bayesian estimate  $\hat{\mu}_\Theta$  of  $\Theta$ .

From Eq. (18.6.2), the conditional density of  $\Theta$  conditioned on the  $n$ -vector random variable  $\mathbf{Y}$  can be given as

$$f(\theta | \mathbf{y}) = \frac{f(\mathbf{y}, \theta)}{f(\mathbf{y})} = \frac{f(\mathbf{y} | \theta) f(\theta)}{f(\mathbf{y})} = \frac{f(\mathbf{y} | \theta) f(\theta)}{\int_{-\infty}^{\infty} f(\mathbf{y} | \theta) f(\theta) d\theta}$$

Since  $\int_{-\infty}^{\infty} f(\theta | \mathbf{y}) d\theta = [1/f(\mathbf{y})] \int_{-\infty}^{\infty} f(\mathbf{y} | \theta) f(\theta) d\theta = 1$ , the factor  $f(\mathbf{y})$  is independent of  $\theta$  and normalizes the integral  $\int_{-\infty}^{\infty} f(\mathbf{y} | \theta) f(\theta) d\theta$  to 1. The joint density function  $f(\mathbf{y} | \Theta = \theta)$  is governed by the joint density of  $W$ , and hence

$$\begin{aligned} f(\mathbf{y} | \theta) &= \prod_{i=1}^n f(y_i | \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_W} e^{-(1/2)[(y_i - \theta)/\sigma_W]^2} \\ &= \frac{1}{(2\pi)^{n/2} \sigma_W^n} e^{-(1/2) \sum_{i=1}^n [(y_i - \theta)/\sigma_W]^2} \end{aligned}$$

with

$$f(\theta) = \frac{1}{\sqrt{2\pi}\sigma_\Theta} e^{-(1/2)(\theta/\sigma_\Theta)^2}$$

Multiplying  $f(\mathbf{y} | \theta)$  and  $f(\theta)$ , we have

$$f(\mathbf{y} | \theta) f(\theta) = f(\mathbf{y}, \theta) = \frac{1}{(2\pi)^{(n+1)/2} \sigma_W^n \sigma_\Theta} \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^n \left( \frac{y_i - \theta}{\sigma_W} \right)^2 + \left( \frac{\theta}{\sigma_\Theta} \right)^2 \right] \right\}$$

The factor  $\sum_{i=1}^n (y_i - \theta)^2$  can be written as,

$$\sum_{i=1}^n (y_i - \theta)^2 = \sum_{i=1}^n (y_i - \hat{\mu}_Y + \hat{\mu}_Y - \theta)^2 = \sum_{i=1}^n (y_i - \hat{\mu}_Y)^2 + n(\hat{\mu}_Y - \theta)$$

since the cross product vanishes. Substitution of this result in  $f(\mathbf{y}, \theta)$  yields

$$\begin{aligned} f(\mathbf{y}, \theta) &= \frac{1}{(2\pi)^{(n+1)/2} \sigma_W^n \sigma_\Theta} \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^n \left( \frac{y_i - \hat{\mu}_Y}{\sigma_W} \right)^2 \right] \right\} \\ &\quad \times \exp \left[ -\frac{1}{2} \left( \frac{n}{\sigma_W^2} (\hat{\mu}_Y - \theta)^2 + \frac{1}{\sigma_\Theta^2} \theta^2 \right) \right] \\ &= C \exp \left\{ -\frac{1}{2} \left( \frac{n}{\sigma_W^2} (\hat{\mu}_Y - \theta)^2 + \frac{1}{\sigma_\Theta^2} \theta^2 \right) \right\} \\ &= C \exp \left\{ -\frac{1}{2} \left[ \frac{n}{\sigma_W^2} (\hat{\mu}_Y^2 - 2\hat{\mu}_Y \theta) + \theta^2 \left( \frac{n\sigma_\Theta^2 + \sigma_W^2}{\sigma_W^2 \sigma_\Theta^2} \right) \right] \right\} \\ &= C \exp \left\{ -\frac{1}{2} \frac{n\sigma_\Theta^2 + \sigma_W^2}{\sigma_W^2 \sigma_\Theta^2} \left[ \theta^2 - \frac{2n\sigma_\Theta^2 \hat{\mu}_Y \theta}{n\sigma_\Theta^2 + \sigma_W^2} + \frac{n\sigma_\Theta^2 \hat{\mu}_Y^2}{n\sigma_\Theta^2 + \sigma_W^2} \right] \right\} \end{aligned}$$

where  $C$  is a function of only the sample values  $\{y_i\}$  and the parameters. Completing the squares in  $\theta$  in the exponent and after some algebra, there results

$$\begin{aligned} f(\mathbf{y}, \theta) &= C \exp \left\{ -\frac{1}{2} \left[ \frac{\theta - \frac{n\sigma_\Theta^2 \hat{\mu}_Y}{n\sigma_\Theta^2 + \sigma_W^2}}{\frac{\sigma_W \sigma_\Theta}{\sqrt{n\sigma_\Theta^2 + \sigma_W^2}}} \right]^2 \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2} \left[ \frac{n\hat{\mu}_Y^2}{n\sigma_\Theta^2 + \sigma_W^2} \right] \right\} = K e^{-(1/2)[(\theta - m_\theta)/s_\theta]^2} \end{aligned}$$

where the quantities,  $m_\theta$  and  $s_\theta$  are defined by

$$m_\theta = \frac{n\sigma_\Theta^2 \hat{\mu}_Y}{n\sigma_\Theta^2 + \sigma_W^2}, \quad s_\theta = \frac{\sigma_W \sigma_\Theta}{\sqrt{n\sigma_\Theta^2 + \sigma_W^2}}$$

and  $K$  is another function independent of  $\theta$ . The marginal density  $f(\mathbf{y})$  can be given as

$$f(\mathbf{y}) = K \sqrt{2\pi} \cdot s_\theta \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \cdot s_\theta} e^{-(1/2)[(\theta - m_\theta)/s_\theta]^2} d\theta = K \sqrt{2\pi} \cdot s_\theta$$

Substituting for  $f(\mathbf{y}, \theta)$  and  $f(\mathbf{y})$  in the posterior density  $f(\theta | \mathbf{y})$ , we obtain

$$f(\theta | \mathbf{y}) = \frac{f(\mathbf{y}, \theta)}{f(\mathbf{y})} = \frac{K e^{-(1/2)[(\theta - m_\theta)/s_\theta]^2}}{K \sqrt{2\pi} \cdot s_\theta} = \frac{1}{\sqrt{2\pi} \cdot s_\theta} e^{-(1/2)[(\theta - m_\theta)/s_\theta]^2}$$

Hence, the posterior density is also Gaussian-distributed with mean  $m_\theta$  and variance  $s_\theta^2$ , and the conditional expectation of  $\Theta$  given the observed values of  $\mathbf{y}$  using Bayesian estimation techniques is

$$E[\Theta | \mathbf{y}] = m_\theta = \frac{n\sigma_\theta^2 \hat{\mu}_Y}{n\sigma_\theta^2 + \sigma_W^2}$$

Note that using classical estimation techniques where there is no knowledge of the parameters of  $\Theta$ , the conditional expectation of  $\Theta$  given the observed values of  $\mathbf{y}$  is

$$E[\Theta | \mathbf{y}] = \hat{\Theta} = \hat{\mu}_Y$$

obtained by substituting  $\sigma_\Theta^2 = \infty$  in the Bayesian estimate  $m_\theta$ .

**Example 18.6.2** A machine of uncertain quality manufactures microchips. In the absence of any information, we shall assume a priori that the bad microchips are uniformly distributed in the interval  $[0,1]$ . We take a sample of 10 microchips and find that 3 are bad. We want to estimate the probability of the bad microchips manufactured.

The prior probability of bad microchips from classical method  $= \frac{3}{10} = 0.3$ . We find the posterior probability from Eq. (18.6.2). Here  $n = 10$  and  $k = 3$ ; hence the posterior probability is  $(3 + 1)/(10 + 2) = 0.3333$ , indicating that the proportion of bad chips is worse than what we might obtain by the classical method. On the other hand, if the number of bad chips out of 10 is only 2, then the prior will be  $\frac{2}{10} = 0.2$ . The Bayesian probability will yield  $(2 + 1)/(10 + 2) = 0.25$ . In this case also, the proportion of bad chips is worse than that obtained from the classical method.

### **Maximum A Posteriori Probability (MAP) Estimation**

We will now discuss the important problem of maximizing the posterior probability to arrive at a decision. A communication receiver with additive noise is given by

$$x_i = s_i + w_i, \quad i = 1, \dots, n, \quad \text{or} \quad \mathbf{x} = \mathbf{s} + \mathbf{w} \quad (18.6.13)$$

where  $\{x_i\}$  is the observation process defined by the  $n$ -vector  $\mathbf{x}$ ,  $\{s_i\}$  is the signal process defined by the  $n$ -vector  $\mathbf{s}$ , and  $\{w_i\}$  is a zero mean random noise defined by the  $n$ -vector  $\mathbf{w}$ . Without loss in generality, we will assume that the signal vector is either all 0 or 1. We observe  $\mathbf{x}$  and decide on one of the following two hypotheses:

$$\begin{aligned} H_0: x_i &= w_i, \quad i = 1, \dots, n, \quad \text{or} \quad \mathbf{x} = \mathbf{w}: \quad 0 \text{ transmitted} \\ H_1: x_i &= s_i + w_i, \quad i = 1, \dots, n \quad \text{or} \quad \mathbf{x} = \mathbf{s} + \mathbf{w}: \quad 1 \text{ transmitted} \end{aligned}$$

where  $H_0$  is the null hypothesis and  $H_1$  is the alternate hypothesis. The prior probabilities of these hypotheses are  $P(H_0)$  and  $P(H_1)$ . The probability density functions of observing  $\mathbf{x}$  conditioned on  $H_0$  and  $H_1$  are  $f_{\mathbf{x}}(\mathbf{x} | H_0)$  and  $f_{\mathbf{x}}(\mathbf{x} | H_1)$ . The posterior probabilities of  $H_0$  and  $H_1$  after observing  $\mathbf{x}$  are  $P(H_0 | \mathbf{x})$  and  $P(H_1 | \mathbf{x})$ . The decision will be based on maximizing the posterior probabilities. We choose the following decision criterion:

$$\begin{aligned} P(H_0 | \mathbf{x}) &> P(H_1 | \mathbf{x}): \quad \text{choose } H_0 \\ P(H_0 | \mathbf{x}) &< P(H_1 | \mathbf{x}): \quad \text{choose } H_1 \end{aligned}$$

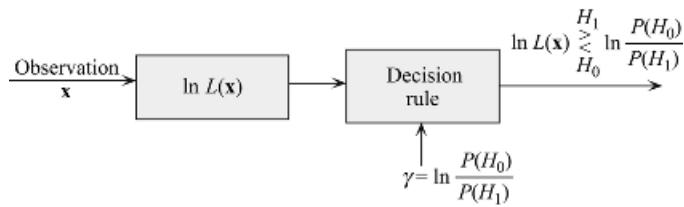


FIGURE 18.6.1

Or more succinctly, the decision criterion is

$$\frac{P(H_1 | \mathbf{x})}{P(H_0 | \mathbf{x})} \frac{H_1}{H_0} \geq 1 \quad (18.6.14)$$

If  $\{H_i, i = 1,2\}$  are the two hypotheses, then, from Bayes' rule, we obtain

$$P(H_i | \mathbf{x}) = \frac{P(H_i) f_{\mathbf{x}}(\mathbf{x} | H_i)}{f_{\mathbf{x}}(\mathbf{x})} = \frac{f_{\mathbf{x}}(\mathbf{x} | H_i) P(H_i)}{f_{\mathbf{x}}(\mathbf{x})}, \quad i = 0, 1 \quad (18.6.15)$$

Substituting Eq. (18.6.15) in Eq. (18.6.14), the decision rule becomes

$$\frac{\frac{f_{\mathbf{x}}(\mathbf{x} | H_1) P(H_1)}{f_{\mathbf{x}}(\mathbf{x})} H_1}{\frac{f_{\mathbf{x}}(\mathbf{x} | H_0) P(H_0)}{f_{\mathbf{x}}(\mathbf{x})} H_0} \geq 1 \quad \text{or} \quad \frac{f_{\mathbf{x}}(\mathbf{x} | H_1) P(H_1)}{f_{\mathbf{x}}(\mathbf{x} | H_0) P(H_0)} \frac{H_1}{H_0} \geq 1 \quad (18.6.16)$$

Transposing the prior probabilities to the right of the inequality, we have

$$L(\mathbf{x}) = \frac{f_{\mathbf{x}}(\mathbf{x} | H_1) P(H_1)}{f_{\mathbf{x}}(\mathbf{x} | H_0) P(H_0)} \frac{H_1}{H_0} \geq \frac{P(H_0)}{P(H_1)} = \gamma \quad (18.6.17)$$

where  $L(\mathbf{x})$  is defined as the *likelihood function* or *likelihood ratio* and  $\gamma$  is the threshold value or the critical point. Since  $L(\mathbf{x})$  is the ratio of  $n$ -fold probability densities, we prefer to take logarithm of  $L(\mathbf{x})$ . The inequality of Eq. (18.6.17) will be preserved since the logarithm is a monotone function. Thus the log likelihood ratio  $\ln[L(\mathbf{x})]$  is given by

$$\ln[L(\mathbf{x})] \geq \frac{H_1}{H_0} \ln \left[ \frac{P(H_0)}{P(H_1)} \right] = \ln \gamma \quad (18.6.18)$$

The inequality given by this equation maximizes the posterior probability.

We can now represent the decision receiver as shown in Fig. 18.6.1.

### Minimization of Average Probability of Error

In the previous section the posterior probability was maximized to arrive at a decision rule. We can approach this problem through minimizing the *average* probability of error. The probabilities of missed alarm  $\alpha$  and false alarm  $\beta$  are given by

$$\begin{aligned} \alpha &= P(\text{accept } H_1 | H_0 \text{ true}) = P(H_1 | H_0) \\ \beta &= P(\text{accept } H_0 | H_1 \text{ true}) = P(H_0 | H_1) \end{aligned} \quad (18.6.19)$$

$\alpha$ ,  $\beta$ , and the acceptance and rejection regions are as shown in Fig. 18.6.2.

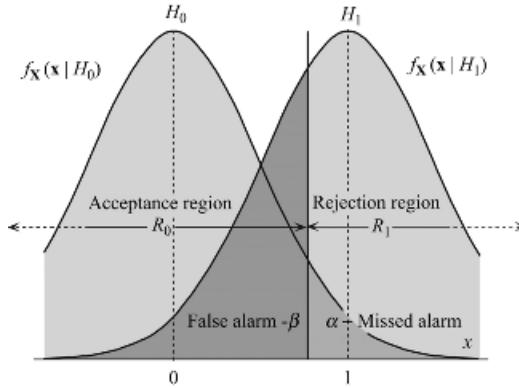


FIGURE 18.6.2

The probability of average error  $P_e$  is given by

$$\begin{aligned} P_e &= P(\text{average error}) = P(H_1 | H_0)P(H_0) + P(H_0 | H_1)P(H_1) \\ &= P(H_0) \int_{R_1} f_X(x | H_0) dx + P(H_1) \int_{R_0} f_X(x | H_1) dx \end{aligned} \quad (18.6.20)$$

If  $R$  is the entire region of integration, then  $R = R_0 + R_1$  and

$$\int_R f_X(x | H_0) dx = \int_R f_X(x | H_1) dx = 1 \quad (18.6.21)$$

Using Eq. (18.6.21), we can convert Eq. (18.6.20) into integration over the region  $R_0$

$$\begin{aligned} P_e &= P(H_0) \left[ \int_R f_X(x | H_0) dx - \int_{R_0} f_X(x | H_0) dx \right] \\ &\quad + P(H_1) \int_{R_0} f_X(x | H_1) dx \\ &= P(H_0) + \int_{R_0} [P(H_1)f_X(x | H_1) - P(H_0)f_X(x | H_0)] dx \end{aligned} \quad (18.6.22)$$

where  $P(H_0)$  is always positive, and minimizing  $P_e$  depends on the function  $g(x) = P(H_1)f_X(x | H_1) - P(H_0)f_X(x | H_0)$  under the integral sign. A plot of  $g(x)$  is shown in Fig. 18.6.3. From the figure, we obtain

$$\begin{aligned} P(H_1)f_X(x | H_1) &> P(H_0)f_X(x | H_0): \quad g(x) \text{ is positive} \Rightarrow x \in R_1 \\ P(H_1)f_X(x | H_1) &< P(H_0)f_X(x | H_0): \quad g(x) \text{ is negative} \Rightarrow x \in R_0 \end{aligned} \quad (18.6.23)$$

Hence we have the decision criterion

$$\frac{f_X(x | H_1)P(H_1)}{f_X(x | H_0)P(H_0)} \stackrel{H_1}{\gtrless} 1 \quad \text{or} \quad \frac{f_X(x | H_1)}{f_X(x | H_0)} \stackrel{H_1}{\gtrless} \frac{P(H_0)}{P(H_1)} \quad (18.6.24)$$

which is exactly the same as Eq. (18.6.17) derived for the MAP criterion.

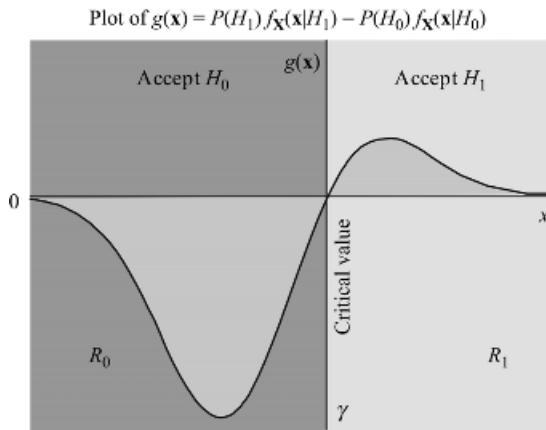


FIGURE 18.6.3

### Maximum-Likelihood Estimation

We have already formulated the basic ML criterion in Eq. (18.1.7) as the maximization of the loglikelihood function  $\ln L(\mathbf{x}; \boldsymbol{\theta})$ . We will discuss this important estimation criterion more formally.  $X$  is a continuous random variable with density function  $f_X(x; \boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  is a  $k$ -vector given by  $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_k]^T$  are parameters to be estimated by observing  $n$  ( $n > k$ ) independent observations  $x_1, x_2, \dots, x_n$  of the random variable  $X$ . The likelihood function  $L(\mathbf{x}; \boldsymbol{\theta})$  is defined by

$$L(\mathbf{x}; \boldsymbol{\theta}) = \prod_{i=1}^n f_X(x_i; \boldsymbol{\theta})$$

and the loglikelihood function, by

$$\ln[L(\mathbf{x}; \boldsymbol{\theta})] = \sum_{i=1}^n \ln [f_X(x_i; \boldsymbol{\theta})] \quad (18.6.25)$$

Maximizing the loglikelihood function  $\ln[L(\mathbf{x}; \boldsymbol{\theta})]$  with respect to  $\{\theta_i, i = 1, \dots, k\}$  and setting them equal to zero yields  $k$  equations given by

$$\frac{\partial}{\partial \theta_i} \{\ln[L(\mathbf{x}; \theta_1, \dots, \theta_k)]\} = 0, \quad i = 1, \dots, k \quad (18.6.26)$$

The estimates of the  $k$  parameters  $(\theta_1, \theta_2, \dots, \theta_k)$  are obtained by solving the  $k$  equations in Eq. (18.6.26). We will briefly discuss the properties of ML estimators:

1. They provide a consistent approach to parameter estimation problems.
2. They can be used to generate hypothesis tests for parameters.
3. The likelihood equations need to be specifically worked out for any given distribution and estimation problems.
4. The numerical estimation problem is usually nontrivial. Simple solutions are possible in only a few cases.
5. Maximum-likelihood estimates can be heavily biased for small samples.

**Example 18.6.3** In an election we want to estimate the proportion  $p$  of voters who will vote Democratic. We will assign 1 if a person votes Democratic and 0 if he

does not vote Democratic. The probability  $f(x;p)$  for each person is a Bernoulli distribution with

$$f(x;p) = p^x(1-p)^{1-x}, \quad x = 0,1; \quad 0 \leq p \leq 1$$

The value of  $p$  has to be estimated using the ML rule based on an opinion poll of  $n$  voters chosen at random. Thus, there will be  $n$  sample values  $x_1, x_2, \dots, x_n$ , each with probability

$$f(x_i;p) = p^{x_i}(1-p)^{1-x_i}, \quad x_i = 0,1; \quad 0 \leq p \leq 1; \quad i = 1, \dots, n$$

The likelihood function and the loglikelihood functions are given by

$$L(\mathbf{x};p) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}$$

$$\ln[L(\mathbf{x};p)] = \sum_{i=1}^n x_i \ln(p) + \left( n - \sum_{i=1}^n x_i \right) \ln(1-p)$$

Differentiating  $\ln[L(\mathbf{x};p)]$  with respect to  $p$ , we have

$$\frac{\partial}{\partial p} \ln[L(\mathbf{x};p)] = \frac{\sum_{i=1}^n x_i}{p} - \frac{\left( n - \sum_{i=1}^n x_i \right)}{1-p} = 0$$

We obtain the estimate  $\hat{p}$  by solving for  $p$  in this equation:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i = \hat{\mu}_X$$

where  $\hat{p}$  is an unbiased estimator because  $E[\hat{p}] = E[\hat{\mu}_X] = p \cdot 1 + (1-p) \cdot 0 = p$ .

**Example 18.6.4** Telephone calls coming into an exchange are Poisson-distributed with parameter  $\lambda$  and are given by

$$f_X(x;\lambda) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0,1,\dots$$

where  $\lambda$  is to be estimated using the ML rule by observing a random sample  $x_1, x_2, \dots, x_n$ , of  $n$  telephone calls, each with probability

$$f_X(x_i;\lambda) = e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}, \quad i = 1,2,\dots,n$$

The likelihood function and the loglikelihood functions are given by

$$L(\mathbf{x};\lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$$

$$\ln[L(\mathbf{x};\lambda)] = -n\lambda + \sum_{i=1}^n x_i \ln \lambda - \sum_{i=1}^n \ln(x_i!)$$

Differentiating  $\ln[L(\mathbf{x};\lambda)]$  with respect to  $\lambda$ , we have

$$\frac{\partial}{\partial \lambda} \ln[L(\mathbf{x};\lambda)] = -n + \frac{\sum_{i=1}^n x_i}{\lambda} = 0$$

We obtain the estimate  $\hat{\lambda}$  by solving for  $\lambda$  in the preceding equation:

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \hat{\mu}_X$$

where  $\hat{\lambda}$  is an unbiased estimator because  $E[\hat{\lambda}] = E[\hat{\mu}_X] = \lambda$ .

**Example 18.6.5** A sample of observations  $\{x_i, i = 1, \dots, n\}$  drawn from a population governed by a form of Weibull density is given by  $f_X(x) = 2\lambda x e^{-\lambda x^2} u(x)$  with parameter  $\lambda$ . We want to estimate  $\lambda$  using the ML criterion. The likelihood function is given by

$$L(\mathbf{x}; \lambda) = \prod_{i=1}^n f_X(x_i) = (2\lambda)^n \prod_{i=1}^n x_i e^{-\lambda \sum_{i=1}^n x_i^2}, \quad x_i > 0$$

and the loglikelihood function by

$$\ln[L(\mathbf{x}; \lambda)] = n \ln(2) + n \ln(\lambda) + \sum_{i=1}^n \ln(x_i) - \lambda \sum_{i=1}^n x_i^2, \quad x_i > 0$$

Differentiating  $\ln[L(\mathbf{x}; \lambda)]$  with respect to  $\lambda$ , we obtain

$$\frac{\partial}{\partial \lambda} \ln[L(\mathbf{x}; \lambda)] = \frac{n}{\lambda} - \sum_{i=1}^n x_i^2 = 0$$

Solving for  $\lambda$  in this equation, we obtain the estimate  $\hat{\lambda}$  as

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i^2} = \frac{1}{\hat{m}_2} \quad \text{where} \quad m_2 = E[X^2]$$

We have  $E[X^2] = m_2 = 1/\lambda$  and  $E[\hat{m}_2] = m_2$ , but  $E[1/\hat{m}_2] \neq 1/E[\hat{m}_2] = 1/m_2 = \lambda$ , and hence  $\hat{\lambda}$  is a biased estimator.

**Example 18.6.6** A random sample of  $n$  is drawn from a Gaussian-distributed population of mean  $\mu_X$  and variance  $\sigma_X^2$ . We want to estimate the mean and the variance using the ML rule. The likelihood function and the loglikelihood function are shown below:

$$\begin{aligned} L(\mathbf{x}; \mu_X, \sigma_X^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-(1/2\sigma_X^2)(x_i - \mu_X)^2} \\ &= \frac{1}{(2\pi)^{n/2} \sigma_X^n} e^{-(1/2\sigma_X^2) \sum_{i=1}^n (x_i - \mu_X)^2} \\ \ln[L(\mathbf{x}; \mu_X, \sigma_X^2)] &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma_X^2) - \frac{1}{2\sigma_X^2} \sum_{i=1}^n (x_i - \mu_X)^2 \end{aligned}$$

Differentiating with respect to  $\mu_X$  and  $\sigma_X^2$ , and setting them equal to zero, we have

$$\begin{aligned}\frac{\partial}{\partial \mu_X} \ln[L(\mathbf{x}; \mu_X, \sigma_X^2)] &= \frac{1}{\sigma_X^2} \sum_{i=1}^n (x_i - \mu_X) = 0 \\ \frac{\partial}{\partial \sigma_X^2} \ln[L(\mathbf{x}; \mu_X, \sigma_X^2)] &= -\frac{n}{2\sigma_X^2} + \frac{1}{2\sigma_X^4} \sum_{i=1}^n (x_i - \mu_X)^2 = 0\end{aligned}$$

Solving these equations, we find the estimators as follows:

$$\hat{\mu}_X = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad \hat{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_X)^2$$

In these estimators, while the estimator for the mean  $\hat{\mu}_X$  is unbiased, the estimator for the variance  $\hat{\sigma}_X^2$  is biased because  $E[\hat{\sigma}_X^2] = [(n-1)/n]\sigma_X^2$ .

### **ML Rule in Hypothesis Testing**

In the previous section we saw that the MAP rule minimizes the average probability of wrong decisions using the prior probabilities  $P\{H_i, i = 0, 1\}$  of the two hypotheses. In many cases the priors may not be exactly known. In such a case, we choose whichever conditional probability  $P(\mathbf{x} | H_0)$  or  $P(\mathbf{x} | H_1)$  is greater, thereby making the assumption that both hypotheses  $H_0$  and  $H_1$  are equiprobable. We will show later in an example that when the priors are not known precisely, the decision errors will be more than that when the priors are equiprobable. This decision rule where the priors are assumed equiprobable is called the *maximum-likelihood* (ML) rule for testing hypotheses and is stated as follows:

$$\begin{aligned}P(\mathbf{x} | H_0) > P(\mathbf{x} | H_1): & \text{ accept } H_0 \\ P(\mathbf{x} | H_0) < P(\mathbf{x} | H_1): & \text{ accept } H_1\end{aligned}\tag{18.6.27}$$

In terms of density functions Eq. (18.6.27) can be given more succinctly as

$$\frac{f_{\mathbf{x}}(\mathbf{x} | H_1)}{f_{\mathbf{x}}(\mathbf{x} | H_0)} \stackrel{H_1}{\gtrless} 1 \tag{18.6.28}$$

In many applications the ML rule is widely used, especially when the average probability of error  $P_e$  is low. Further, the missed-alarm probability is much more important than the false-alarm probability. For example, missing the diagnosis of cancer has more serious consequences than does diagnosing cancer when there is none. The MAP rule is prone to yielding higher miss probabilities than is the ML rule as we will see in the following example.

**Example 18.6.7** The output  $x$  of a system consists of an input  $s = 0$  or  $1$  and a zero mean additive noise  $w$  with variance  $\sigma_w^2 = \frac{1}{4}$ . The prior probability of sending  $s = 0$  is  $P(s = 0) = \frac{3}{4}$  and that of  $s = 1$  is  $P(s = 1) = \frac{1}{4}$ . We have to find the following using both the MAP and the ML rules: (1) the decision threshold  $\gamma$ , (2) the average probability of error  $P_e(x)$ , and (3) the probabilities of missed alarm and false alarm. We can formulate the hypotheses as

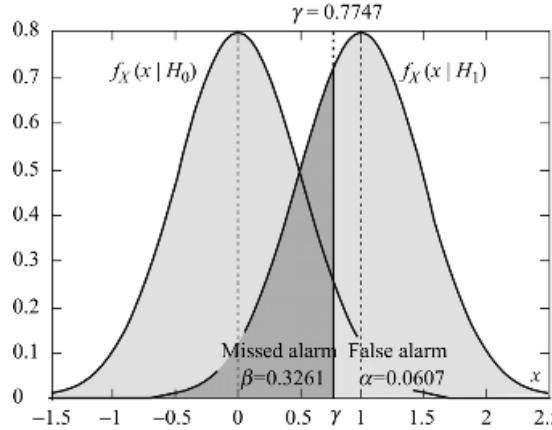


FIGURE 18.6.4

follows:

$$H_0: \{0 \text{ transmitted}\} \text{ with } P(H_0) = \frac{3}{4}$$

$$H_1: \{1 \text{ transmitted}\} \text{ with } P(H_1) = \frac{1}{4}$$

The conditional probabilities of  $x$  under the two hypotheses are

$$f_X(x | H_0) = \frac{2}{\sqrt{2\pi}} e^{-(1/2)[4x^2]} \quad \text{and} \quad f_X(x | H_1) = \frac{2}{\sqrt{2\pi}} e^{-(1/2)[4(x-1)^2]}$$

These probability density functions are shown in Fig. 18.6.4.

The likelihood ratio  $L(x)$  is given by

$$L(x) = \frac{f_X(x | H_1)}{f_X(x | H_0)} = \frac{e^{-(1/2)[4(x-1)^2]}}{e^{-(1/2)[4x^2]}} = e^{-2[x^2 - 2x + 1 - x^2]} = e^{2(2x-1)}$$

*MAP Criterion.* The decision rule from the loglikelihood function of Eq. (18.6.18) is

$$\frac{\ln[e^{2(2x-1)}]}{H_0} \geq \frac{\ln\left[\frac{3/4}{1/4}\right]}{H_0} \quad \text{or} \quad \frac{H_1}{H_0} \geq \frac{1}{2} + \frac{\ln 3}{4} \quad \text{or} \quad \frac{H_1}{H_0} \geq 0.7747$$

Thus, the critical point for MAP is  $\gamma = 0.7747$  as shown in Fig. 18.6.4. The probability of average error  $P_e(x)$  for any value of  $x$  from Eq. (18.6.22) is given by

$$\begin{aligned} P_{e\text{MAP}}(x) &= P(H_0) + \int_{-\infty}^x [P(H_1)f_X(\xi | H_1) - P(H_0)f_X(\xi | H_0)]d\xi \\ &= \frac{3}{4} + \frac{2}{\sqrt{2\pi}} \int_{-\infty}^x \left[ \frac{1}{4}e^{-2(\xi-1)^2} - \frac{3}{4}e^{-2\xi^2} \right] d\xi \\ &= \frac{1}{2} + \frac{1}{8} \left\{ \text{erf}[\sqrt{2}(x-1) - 3\sqrt{2\pi}] - \text{erf}[\sqrt{2\pi}] \right\}; \quad \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} d\xi \end{aligned}$$

The minimum error is obtained when  $x = \gamma = 0.7747$  and  $P_e(\min) = 0.127$ . The probability of missed alarm is obtained from the following integral

$$P_{\text{MAMAP}} = \int_{-\infty}^{\gamma} \frac{2}{\sqrt{2\pi}} e^{-2(x-1)^2} dx = 0.3261$$

and the probability of false alarm from

$$P_{\text{FAMAP}} = \int_{\gamma}^{\infty} \frac{2}{\sqrt{2\pi}} e^{-2x^2} dx = 0.0607$$

These probabilities are also shown in Fig. 18.6.4. The function  $g_{\text{MAP}}(x)$  under the integral, that governs the extremum for  $P_{e\text{MAP}}(x)$  is given by

$$g_{\text{MAP}}(x) = \frac{2}{\sqrt{2\pi}} \left[ \frac{1}{4} e^{-2(x-1)^2} - \frac{3}{4} e^{-2x^2} \right]$$

*ML Criterion.* We will now obtain the same quantities using the ML criterion. Here the decision criterion becomes

$$\begin{matrix} H_1 \\ x \leq \frac{1}{2} \\ H_0 \end{matrix}$$

and the critical point for ML is 0.5. Hence the probability of error  $P_{e\text{ML}}(x)$  is given by

$$\begin{aligned} P_{e\text{ML}}(x) &= \frac{1}{2} + \frac{1}{2} \int_{-\infty}^x [f_X(\xi | H_1) - f_X(\xi | H_0)] d\xi \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x [e^{-2(\xi-1)^2} - e^{-2\xi^2}] d\xi \\ &= \frac{1}{2} + \frac{1}{4} \left\{ \text{erf} \left[ \sqrt{2}(x-1) \right] - \text{erf} \left[ \sqrt{2x} \right] \right\} \end{aligned}$$

The minimum error is obtained when  $x = \gamma = 0.5$  and  $P_e(\min) = 0.159$ . The probability of missed alarm is obtained from the following integral

$$P_{\text{MAML}} = \int_{-\infty}^{0.5} \frac{2}{\sqrt{2\pi}} e^{-2(x-1)^2} dx = 0.1587$$

and the probability of false alarm from,

$$P_{\text{FAML}} = \int_{0.5}^{\infty} \frac{2}{\sqrt{2\pi}} e^{-2x^2} dx = 0.1587$$

Note that both probabilities are the same for the maximum-likelihood criterion. The function  $g_{\text{ML}}(x)$  under the integral that governs the extremum for  $P_{e\text{ML}}(x)$  is given by

$$g_{\text{ML}}(x) = \frac{2}{\sqrt{2\pi}} \left[ e^{-2(x-1)^2} - e^{-2x^2} \right]$$

The average probabilities of error  $P_{e\text{MAP}}(x)$  and  $P_{e\text{ML}}(x)$  are shown in Fig. 18.6.5.

Table 18.6.1 compares the two decision criteria.

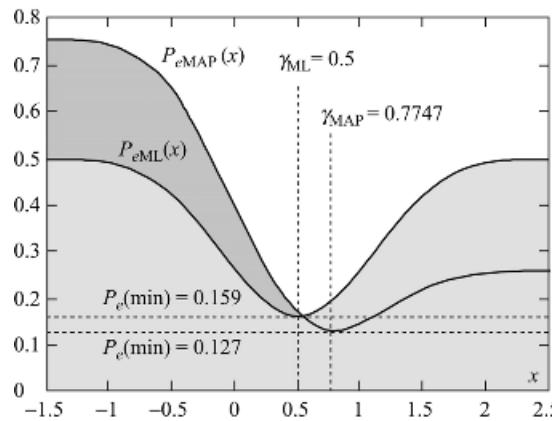


FIGURE 18.6.5

TABLE 18.6.1

Category	MAP Rule	ML Rule
$P(H_0)$ —probability of null hypothesis	0.75	0.5
$P(H_1)$ —probability of alternate hypothesis	0.25	0.5
$\gamma$ —critical point	0.7747	0.5
$P_e(\min)$ —minimum probability of average error	0.127	0.159
$P_{MA}$ —probability of missed alarm	0.3261	0.1587
$P_{FA}$ —probability of false alarm	0.0607	0.1587

Comparing the two methods, we see that the minimum probability of average error is smaller for the MAP than for the ML. However, the probability of the more important missed alarm is higher for MAP than for ML. Thus, in most communication problems the ML criterion is the preferred method of estimation. The functions  $g_{MAP}(x)$  and  $g_{ML}(x)$  for both MAP and ML criteria are shown in Fig. 18.6.6.

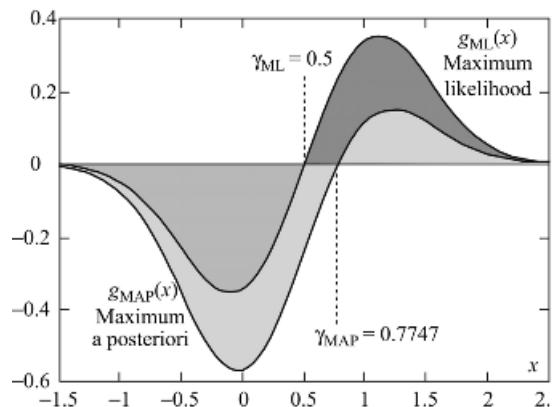


FIGURE 18.6.6

In Fig. 18.6.6 the function  $g_{\text{ML}}(x)$  corresponding to the ML rule has an odd symmetry, and hence it integrates to zero. The critical point is  $\frac{1}{2}$ , as seen before.

**Example 18.6.8 (Discrete Probability)** An extrusion company has contracted to supply plastic shopping bags to supermarkets. Ordinarily, the probability of tearing in this group  $A_0$  of bags is 0.01. Because of a production malfunction, a second group  $A_1$  of bags have a higher probability of 0.05 of tearing. The prior probabilities of choosing  $A_0$  or  $A_1$  are 0.7 and 0.3 respectively. To find out which group of bags came from  $A_1$ , the company tests a sample of  $n$  bags sequentially until it comes across a first failure. From this information it must decide whether the bags came from group  $A_0$  or  $A_1$ . Using both the MAP and the ML criteria, we have to find (1) the number  $n$  of bags required to be tested, (2) the probability of average error  $P_e(n)$ , and (3) probabilities of missed and false alarms  $P_{\text{MA}}$  and  $P_{\text{FA}}$ .

We will first approach the problem in generality and later substitute numbers to obtain results. Let  $q_0$  be the tearing probability in group  $A_0$  and  $q_1$  be the tearing probability in group  $A_1$ . The two hypotheses can be defined by

$$\begin{aligned} H_0: & \text{ bag came from group } A_0 \\ H_1: & \text{ bag came from group } A_1 \end{aligned} \quad (18.6.29)$$

The conditional probabilities of the waiting times upto the first tearing are geometrically distributed and are given by

$$P(n | H_0) = q_0(1 - q_0)^{n-1}; \quad P(n | H_1) = q_1(1 - q_1)^{n-1} \quad (18.6.30)$$

*MAP Criterion.* According to the MAP criterion of Eq. (18.6.14), the decision rule is

$$\begin{aligned} \frac{P(H_1 | n) H_1}{P(H_0 | n) H_0} \gtrless 1 & \quad \text{or} \quad \frac{\frac{P(n | H_1) P(H_1)}{P(n)} H_1}{\frac{P(n | H_0) P(H_0)}{P(n)} H_0} \gtrless 1 \quad \text{or} \\ \frac{P(n | H_1) P(H_1) H_1}{P(n | H_0) P(H_0) H_0} \gtrless 1 & \quad \text{or} \quad \frac{P(n | H_1) H_1 P(H_0)}{P(n | H_0) H_0 P(H_1)} \gtrless 1 \end{aligned} \quad (18.6.31)$$

Substituting the conditional probabilities from Eq. (18.6.30) into Eq. (18.6.31), we have

$$\frac{q_1}{q_0} \left( \frac{1 - q_1}{1 - q_0} \right)^{n-1} \frac{H_1 P(H_0)}{H_0 P(H_1)} \gtrless 1 \quad (18.6.32)$$

Taking logarithms on both sides of Eq. (18.6.32) and simplifying, we obtain

$$(n - 1) \frac{H_1}{H_0} \frac{\ln \left( \frac{q_0 P(H_0)}{q_1 P(H_1)} \right)}{\ln \left( \frac{1 - q_1}{1 - q_0} \right)} \quad \text{or} \quad n \frac{H_1}{H_0} n_{\text{MAP}} = 1 + \frac{\ln \left( \frac{q_0 P(H_0)}{q_1 P(H_1)} \right)}{\ln \left( \frac{1 - q_1}{1 - q_0} \right)} \quad (18.6.33)$$

where  $n_{\text{MAP}}$  is the critical value of  $n$ . Thus we have the decision criterion for  $n$ :

$$\begin{aligned} \text{If } n > n_{\text{MAP}}: & \text{ accept hypothesis } H_1 \\ \text{If } n < n_{\text{MAP}}: & \text{ accept hypothesis } H_0 \end{aligned} \quad (18.6.34)$$

This MAP decision criterion gives us the minimum number  $n$  that are needed to be tested sequentially.

*Probabilities of Missed and False Alarms and Average Error.* We will now calculate the probabilities of missed and false alarms as functions of  $n$ . The probability of missed alarm  $P_{\text{MA}}(n)$  is given by

$$\begin{aligned} P_{\text{MA}}(n) &= P(N > n | H_1) = \sum_{k=n+1}^{\infty} q_1(1 - q_1)^{k-1} \\ &= 1 - \sum_{k=1}^n q_1(1 - q_1)^{k-1} = 1 - q_1 \frac{1 - (1 - q_1)^n}{1 - (1 - q_1)} = (1 - q_1)^n \end{aligned} \quad (18.6.35)$$

where we have used the result for the finite summation of a geometric series. Similarly, the probability of false alarm  $P_{\text{FA}}(n)$  is given by

$$\begin{aligned} P_{\text{FA}}(n) &= P(N \leq n | H_0) = \sum_{k=1}^n q_0(1 - q_0)^{k-1} \\ &= q_0 \frac{1 - (1 - q_0)^n}{1 - (1 - q_0)} = 1 - (1 - q_0)^n \end{aligned} \quad (18.6.36)$$

Finally, the probability of average error  $P_{e\text{MAP}}(n)$  is given by

$$P_{e\text{MAP}}(n) = (1 - q_1)^n P(H_1) + [1 - (1 - q_0)^n] P(H_0) \quad (18.6.37)$$

We obtain the various probabilities by substituting  $n = n_{\text{MAP}}$  in Eqs. (18.6.35)–(18.6.37).

*ML Criterion.* For the ML criterion, the hypotheses  $H_0$  and  $H_1$  are equiprobable, and substituting  $P(H_0) = P(H_1)$  in Eq. (18.6.36), we obtain

$$n \frac{H_1}{H_0} \geq n_{\text{ML}} = 1 + \frac{\ln\left(\frac{q_0}{q_1}\right)}{\ln\left(\frac{1 - q_1}{1 - q_0}\right)} \quad (18.6.38)$$

with the decision criterion similar to that of Eq. (18.6.37).

Equations for the probabilities of missed and false alarms for the ML criterion as functions of  $n$  are exactly the same as those of the MAP criterion except that we substitute  $n = n_{\text{ML}}$ . The probability of average error for the ML criterion can be obtained from Eq. (18.6.37) by substituting  $P(H_0) = P(H_1) = \frac{1}{2}$ , and the resulting equation is

$$P_{e\text{ML}}(n) = \frac{1}{2} + \frac{1}{2} \left\{ (1 - q_1)^n - (1 - q_0)^n \right\} \quad (18.6.39)$$

*Numerical Solutions.* The conditional probability functions  $P(n | H_0) = 0.01 \times (1 - 0.01)^n$  and  $P(n | H_1) = 0.05 \times (1 - 0.05)^n$  are shown in Fig. 18.6.7.

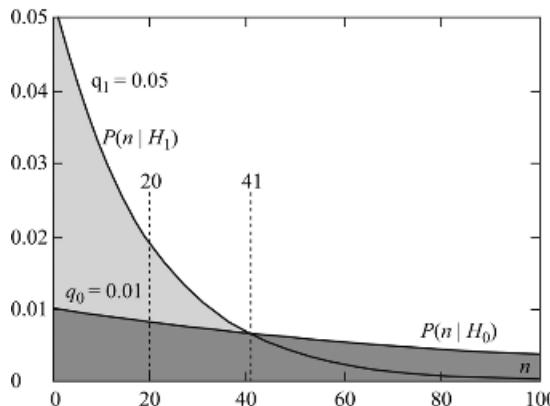


FIGURE 18.6.7

Substituting for the MAP criterion,  $q_0 = 0.01$ ,  $q_1 = 0.05$ ,  $P(H_0) = 0.7$ , and  $P(H_1) = 0.3$  in Eq. (18.6.39), we obtain the critical value  $n_{\text{MAP}}$ :

$$n_{\text{MAP}} = 1 + \frac{\ln\left(\frac{0.01 \times 0.7}{0.05 \times 0.3}\right)}{\ln\left(\frac{1 - 0.05}{1 - 0.01}\right)} = 19.4793$$

Or, using the MAP criterion, the minimum number of bags to test sequentially is  $n = 20$ . The missed-alarm probability is obtained from Eq. (18.6.38):  $P_{\text{MA}}(n) = (1 - q_1)^n = 0.95^n$ , and the  $\alpha$  error is  $0.95^{20} = 0.3585$ . The false-alarm probability is obtained from Eq. (18.6.39):  $P_{\text{FA}}(n) = 1 - (1 - q_0)^n = 1 - 0.99^n$ , and the  $\beta$  error is  $1 - 0.99^{20} = 0.1821$ . The probability of average error  $P_{e\text{MAP}}(n)$  is obtained from Eq. (18.6.37):  $P_{e\text{MAP}}(n) = 0.95^n \times 0.3 + (1 - 0.99^n) \times 0.7$  and  $P_{e\text{MAP}}(\text{min}) = P_{e\text{MAP}}(20) = 0.235$ .

For the ML criterion, we substitute  $q_0 = 0.01$ ,  $q_1 = 0.05$ ,  $P(H_0) = 0.5$ , and  $P(H_1) = 0.5$  in Eq. (18.6.38) and obtain the critical value  $n_{\text{ML}}$ :

$$n_{\text{ML}} = 1 + \frac{\ln\left(\frac{0.01}{0.05}\right)}{\ln\left(\frac{1 - 0.05}{1 - 0.01}\right)} = 40.0233$$

Or, using the ML criterion, the minimum number of bags to test sequentially is  $n = 41$ .

The missed-alarm probability is obtained from Eq. (18.6.38):  $P_{\text{MA}}(n) = 1 - (1 - q_1)^n = 0.95^n$ , and the  $\alpha$  error is  $0.95^{41} = 0.1221$ . The false-alarm probability is obtained from Eq. (18.6.39):  $P_{\text{FA}}(n) = 1 - (1 - q_0)^n = 1 - 0.99^n$ , and the  $\beta$  error is  $1 - 0.99^{41} = 0.3377$ . The probability of average error  $P_{e\text{ML}}(n)$  is obtained from Eq. (18.6.39):  $P_{e\text{ML}}(n) = \frac{1}{2} + \frac{1}{2}[0.95^n + 0.99^n]$  and  $P_{e\text{ML}}(\text{min}) = P_{e\text{ML}}(41) = 0.2299$ .

Probabilities  $P_{e\text{MAP}}(n)$  and  $P_{e\text{ML}}(n)$  are shown in Fig. 18.6.8 along with the minimum number of bags needed for both the MAP and ML criteria. The results indicate that the number of bags needed using the MAP is half the number needed for ML. The more important missed-alarm probability for the ML criterion

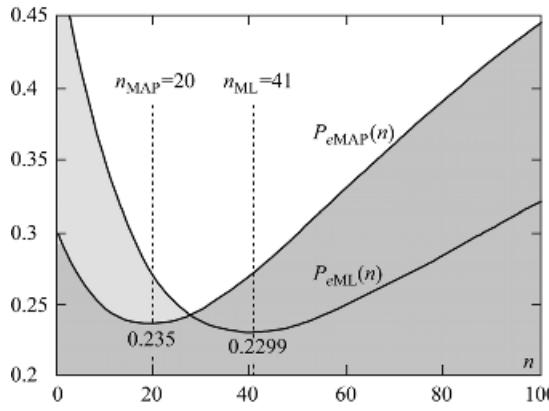


FIGURE 18.6.8

is considerably lower than that of the MAP criterion. The ML criterion is more often used in communication problems. The MAP criterion does find applications in medical imaging problems such as single-photon emission computed tomography (SPECT).

#### ***Hypothesis Testing with Added Cost for Errors***

In the previous sections we never considered the costs associated with type I and type II errors. We have already mentioned that the missed-alarm or type I error is more serious than the false-alarm or type II error. We will now form a decision rule that takes into account the costs associated with the errors. We shall postulate the true hypotheses by  $H_0$  and  $H_1$  and the define the accepted hypotheses by  $A_0$  and  $A_1$ . There will be four pairs of decisions, two of which,  $\{A_0, H_0\}$  and  $\{A_1, H_1\}$ , are correct decisions and the other two of which,  $\{A_0, H_1\}$  (missed alarm) and  $\{A_1, H_0\}$  (false alarm) are incorrect decisions. We will associate a cost  $C_{ij}$  for each pair  $\{A_i, H_j, i, j = 1, 2\}$  of the four decisions. The average cost of the decisions  $\bar{C}$  for the MAP criterion is

$$\bar{C} = C_{00}P(A_0, H_0) + C_{11}P(A_1, H_1) + C_{01}P(A_0, H_1) + C_{10}P(A_1, H_0) \quad (18.6.40)$$

where  $P(A_i, H_j)$   $i, j = 1, 2$  represents the joint probability of accepting  $A_i$ , given that  $H_j$  is true. We have to form a decision rule that minimizes the average cost. Expressing Eq. (18.6.40) in terms of conditional probabilities, we have

$$\begin{aligned} \bar{C} = & C_{00}P(A_0 | H_0)P(H_0) + C_{01}P(A_0 | H_1)P(H_1) \\ & + C_{11}P(A_1 | H_1)P(H_1) + C_{10}P(A_1 | H_0)P(H_0) \end{aligned} \quad (18.6.41)$$

where  $C_{01}$  is the cost of the missed alarm and  $C_{10}$  is the cost of the false alarm. Substituting conditional density functions for the conditional probabilities, there results

$$\begin{aligned} \bar{C}_{\text{MAP}} = & C_{00}P(H_0) \int_{R_0} f_{\mathbf{X}}(\mathbf{x} | H_0) d\mathbf{x} + C_{01}P(H_1) \int_{R_0} f_{\mathbf{X}}(\mathbf{x} | H_1) d\mathbf{x} \\ & + C_{11}P(H_1) \int_{R_1} f_{\mathbf{X}}(\mathbf{x} | H_1) d\mathbf{x} + C_{10}P(H_0) \int_{R_1} f_{\mathbf{X}}(\mathbf{x} | H_0) d\mathbf{x} \end{aligned} \quad (18.6.42)$$

We can transform the integrals in Eq. (18.6.42) to be over the same region  $R_0$ :

$$\begin{aligned}\bar{C}_{\text{MAP}} &= C_{00}P(H_0) \int_{R_0} f_{\mathbf{X}}(\mathbf{x} | H_0) d\mathbf{x} + C_{01}P(H_1) \int_{R_0} f_{\mathbf{X}}(\mathbf{x} | H_1) d\mathbf{x} \\ &\quad + C_{11}P(H_1) \left[ 1 - \int_{R_0} f_{\mathbf{X}}(\mathbf{x} | H_1) d\mathbf{x} \right] + C_{10}P(H_0) \left[ 1 - \int_{R_0} f_{\mathbf{X}}(\mathbf{x} | H_0) d\mathbf{x} \right]\end{aligned}\quad (18.6.43)$$

Combining like terms in Eq. (18.6.43), we obtain

$$\begin{aligned}\bar{C}_{\text{MAP}} &= C_{11}P(H_1) + C_{10}P(H_0) + \int_{R_0} [P(H_1)(C_{01} - C_{11})f_{\mathbf{X}}(\mathbf{x} | H_1) \\ &\quad - P(H_0)(C_{10} - C_{00})f_{\mathbf{X}}(\mathbf{x} | H_0)] d\mathbf{x}\end{aligned}\quad (18.6.44)$$

At this point it is natural to assume that the cost for an incorrect decision exceeds the cost for a correct decision, or  $C_{10} > C_{00}$  and  $C_{01} > C_{11}$ . With this assumption, the terms  $P(H_1)(C_{01} - C_{11})f_{\mathbf{X}}(\mathbf{x} | H_1)$  and  $P(H_0)(C_{10} - C_{00})f_{\mathbf{X}}(\mathbf{x} | H_0)$  are both positive. Hence, if the integrand  $I = P(H_1)(C_{01} - C_{11})f_{\mathbf{X}}(\mathbf{x} | H_1) - P(H_0)(C_{10} - C_{00})f_{\mathbf{X}}(\mathbf{x} | H_0)$  is negative, then we can conclude that  $\mathbf{x} \in R_0$  and choose  $H_0$ . Thus the decision rule for the MAP criterion is

$$\frac{f_{\mathbf{X}}(\mathbf{x} | H_1)P(H_1)(C_{01} - C_{11})}{f_{\mathbf{X}}(\mathbf{x} | H_0)P(H_0)(C_{10} - C_{00})} \frac{H_1}{H_0} \gtrless 1$$

or

$$L(\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x} | H_1)}{f_{\mathbf{X}}(\mathbf{x} | H_0)} \frac{H_1}{H_0} \gtrless \frac{P(H_0)(C_{10} - C_{00})}{P(H_1)(C_{01} - C_{11})} \quad (18.6.45)$$

and the decision rule for the ML criterion can be obtained by substituting  $P(H_0) = P(H_1)$  in Eq. (18.6.45):

$$L(\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x} | H_1)}{f_{\mathbf{X}}(\mathbf{x} | H_0)} \frac{H_1}{H_0} \gtrless \frac{C_{10} - C_{00}}{C_{01} - C_{11}} \quad (18.6.46)$$

In Eqs. (18.6.45) and (18.6.46),  $C_{10}$  is the cost associated with a false alarm and  $C_{01}$  is for a missed alarm. We weigh the missed alarm more heavily than the false alarm with the result  $C_{01} > C_{10}$ . Hence the ratio  $(C_{10} - C_{00})/(C_{01} - C_{11})$  will always be lesser than 1 if the costs associated with right decisions are the same.

**Example 18.6.9** We will reexamine Example 18.6.7, assigning cost functions to wrong decisions. We will assume that the costs for the right decisions are the same,  $C_{00} = C_{11} = 1$ , and the cost for a missed alarm to be  $C_{01} = 4C_{00}$  and the cost for a false alarm,  $C_{10} = 2C_{00}$ . The prior probabilities are  $P(H_0) = \frac{3}{4}$  and  $P(H_1) = \frac{1}{4}$ . The conditional probabilities of  $x$  from Example 18.6.7 under the two hypotheses are

$$f_X(x | H_0) = \frac{2}{\sqrt{2\pi}} e^{-(1/2)[4x^2]} \quad \text{and} \quad f_X(x | H_1) = \frac{2}{\sqrt{2\pi}} e^{-(1/2)[4(x-1)^2]}$$

The likelihood ratio  $L(x)$  is

$$L(x) = \frac{f_X(x | H_1)}{f_X(x | H_0)} = e^{2(2x-1)}$$

From Eq. (18.6.45), the MAP criterion for minimization of cost functions is

$$L(\mathbf{x}) = e^{2(2x-1)} \frac{H_1}{H_0} \frac{P(H_0)(C_{10} - C_{00})}{P(H_1)(C_{01} - C_{11})} = \frac{3C_{00}(2-1)}{C_{00}(4-1)} = 1$$

The critical point from the loglikelihood function can be obtained as follows:

$$2(2x-1) \frac{H_1}{H_0} \geq 0 \quad \text{or} \quad x \frac{H_1}{H_0} \geq \frac{1}{2} \quad \text{or} \quad \gamma_{\text{MAP}} = \frac{1}{2}$$

The average cost function  $\bar{C}_{\text{MAP}}(x)$  obtained from Eq. (18.6.44) is given by

$$\begin{aligned} \bar{C}_{\text{MAP}}(x) &= C_{00} \left( \frac{1}{4} + 2 \frac{3}{4} \right) + \frac{2}{\sqrt{2\pi}} \int_{-\infty}^x C_{00} \\ &\quad \times \left( \frac{1}{4}(4-1)e^{-2(\xi-1)^2} - \frac{3}{4}(2-1)e^{-2\xi^2} \right) dx \\ &= C_{00} \left\{ \frac{7}{4} + \frac{3}{8} \left\{ \text{erf} \left[ \sqrt{2}(x-1) \right] - \text{erf} \sqrt{2}x \right\} \right\} \end{aligned}$$

Assuming that  $C_{00} = 1$ , the minimum cost function  $\bar{C}_{\text{MAP}(\min)} = \bar{C}_{\text{MAP}}(0.5) = 1.238$ .

The probability of missed alarm is obtained from the following integral

$$P_{\text{MAMAP}} = \int_{-\infty}^{\gamma_{\text{MAP}}} \frac{2}{\sqrt{2\pi}} e^{-2(x-1)^2} dx = \frac{1}{2} \left[ 1 - \text{erf} \left( \frac{1}{\sqrt{2}} \right) \right] = 0.1587$$

and the probability of false alarm, from

$$P_{\text{FAMAP}} = \int_{\gamma_{\text{MAP}}}^{\infty} \frac{2}{\sqrt{2\pi}} e^{-2x^2} dx = \frac{1}{2} \left[ 1 - \text{erf} \left( \frac{1}{\sqrt{2}} \right) \right] = 0.1587$$

Similarly, the ML criterion for minimization of cost functions can be obtained by substituting  $P(H_0) = P(H_1) = \frac{1}{2}$  in Eq. (18.6.44). The result is

$$L(\mathbf{x}) = e^{2(2x-1)} \frac{H_1}{H_0} \frac{C_{10} - C_{00}}{C_{01} - C_{11}} = \frac{C_{00}(2-1)}{C_{00}(4-1)} = \frac{1}{3}$$

The critical point from the loglikelihood function can be obtained as follows:

$$2(2x-1) \frac{H_1}{H_0} \ln \left( \frac{1}{3} \right) \quad \text{or} \quad \frac{H_1}{H_0} \frac{1}{2} - \frac{\ln(3)}{4} \quad \text{or} \quad \gamma_{\text{ML}} = 0.2253$$

The average cost function  $\bar{C}_{\text{ML}}(x)$  is

$$\begin{aligned} \bar{C}_{\text{ML}}(x) &= \frac{C_{00}}{2} (1+2) + \frac{2}{2\sqrt{2\pi}} \int_{-\infty}^x C_{00} \left( (4-1)e^{-2(\xi-1)^2} - (2-1)e^{-2\xi^2} \right) dx \\ &= C_{00} \left\{ 2 + \frac{3}{4} \text{erf} \left[ \sqrt{2}(x-1) \right] - \frac{1}{4} \text{erf} \left[ \sqrt{2}x \right] \right\} \end{aligned}$$

Assuming that  $C_{00} = 1$ , the minimum cost function  $\bar{C}_{\text{ML}(\min)} = \bar{C}_{\text{ML}}(0.2253) = 1.254$ .

The probability of missed alarm is obtained from the following integral

$$P_{\text{MAML}} = \int_{-\infty}^{\gamma_{\text{ML}}} \frac{2}{\sqrt{2\pi}} e^{-2(x-1)^2} dx = 0.0607$$

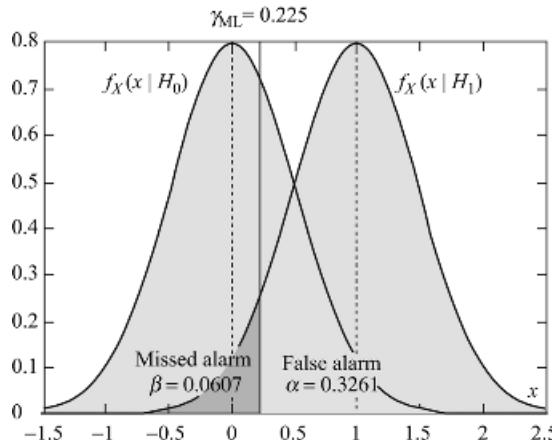


FIGURE 18.6.9

and the probability of false alarm, from

$$P_{\text{FAML}} = \int_{\gamma_{\text{ML}}}^{\infty} \frac{2}{\sqrt{2\pi}} e^{-2x^2} dx = 0.3261$$

The probability density functions, the critical point  $\gamma_{\text{ML}}$ , and the error regions for the ML criterion are shown in Fig. 18.6.9. The cost functions  $\bar{C}_{\text{MAP}}(x)$  and  $\bar{C}_{\text{ML}}(x)$  along with their critical points  $\gamma_{\text{MAP}}$  and  $\gamma_{\text{ML}}$  are shown in Fig. 18.6.10.

The function  $g_{\text{MAP}}(x)$  under the integral that governs the extremum for  $P_{e\text{MAP}}(x)$  is

$$g_{\text{MAP}}(x) = \frac{1}{2\sqrt{2\pi}} \left( e^{-2(\xi-1)^2} - 3e^{-2\xi^2} \right)$$

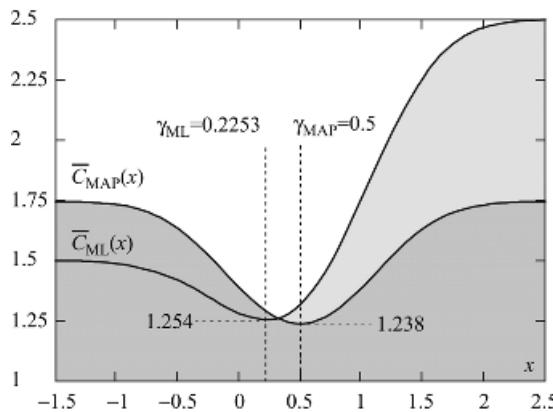


FIGURE 18.6.10

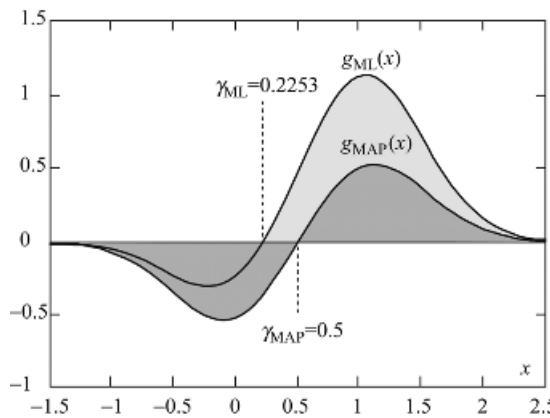


FIGURE 18.6.11

TABLE 18.6.2

Category	MAP Rule	ML Rule
$P(H_0)$ —probability of null hypothesis	0.75	0.5
$P(H_1)$ —probability of alternate hypothesis	0.25	0.5
$C_{00} = C_{11}$ —cost of correct decisions	1	1
$C_{01}$ —cost of missed alarm	4	4
$C_{10}$ —cost of false alarm	2	2
$\gamma$ —critical point	0.5	0.2253
$C_{\min}$ —minimum cost of average error	1.238	1.254
$P_{\text{MA}}$ —probability of missed alarm	0.1587	0.0607
$P_{\text{FA}}$ —probability of false alarm	0.1587	0.3261

The function  $g_{\text{ML}}(x)$  under the integral, which governs the extremum for  $P_{e\text{ML}}(x)$ , is

$$g_{\text{ML}}(x) = \frac{1}{2\sqrt{2\pi}} \left( 3e^{-2(\xi-1)^2} - e^{-2\xi^2} \right)$$

These two functions  $g_{\text{MAP}}(x)$  and  $g_{\text{ML}}(x)$  are shown in Fig. 18.6.11.

Table 18.6.2 compares the two decision criteria.

# Random Processes

## 19.1 BASIC DEFINITIONS

In the earlier sections a random variable  $X$  was defined as a function that maps every outcome  $\xi_i$  of points in the sample space  $S$  to a number  $X(\xi_i)$  on the real line  $R$ . A *random process*  $X(t)$  is a mapping that assigns a time function  $X(t, \xi_i)$  to every outcome  $\xi_i$  of points in the sample space  $S$ . Alternate names for random processes are *stochastic processes* and *time series*. More formally, a random process is a time function assigned for *every* outcome  $\xi \in S$  according to some rule  $X(t, \xi)$ ,  $t \in T$ ,  $\xi \in S$ , where  $T$  is an index set of time. As in the case of a random variable, we suppress  $\xi$  and define a random process by  $X(t)$ . If the index set  $T$  is countably infinite, the random process is called a discrete-time process and is denoted by  $X_n$ .

Referring to Fig. 19.1.1, a random process has the following interpretations:

1.  $X(\xi, t_1)$  is random variable for a fixed time  $t_1$ .
2.  $X(\xi_i, t)$  is a *sample realization* for any point  $\xi_i$  in the sample space  $S$ .
3.  $X(\xi_i, t_2)$  is a number.
4.  $X(\xi, t)$  is a collection or *ensemble* of realizations and is called a *random process*.

An important point to emphasize is that a random process is a finite or an infinite ensemble of time functions and is not a single time function.

**Example 19.1.1** A fair coin is tossed. If heads come up, a sine wave  $x_1(t) = \sin(5\pi t)$  is sent. If tails come up, then a ramp  $x_2(t) = t$  is sent. The resulting random process  $X(t)$  is an ensemble of two realizations, a sine wave and a ramp, and is shown in Fig. 19.1.2. The sample space  $S$  is discrete.

**Example 19.1.2** In this example a sine wave is in the form  $X(t) = A \sin(\omega t + \Phi)$ , where  $\Phi$  is a random variable uniformly distributed in the interval  $(0, 2\pi)$ . Here the sample space is continuous, and the sequence of sine functions is shown in Fig. 19.1.3.

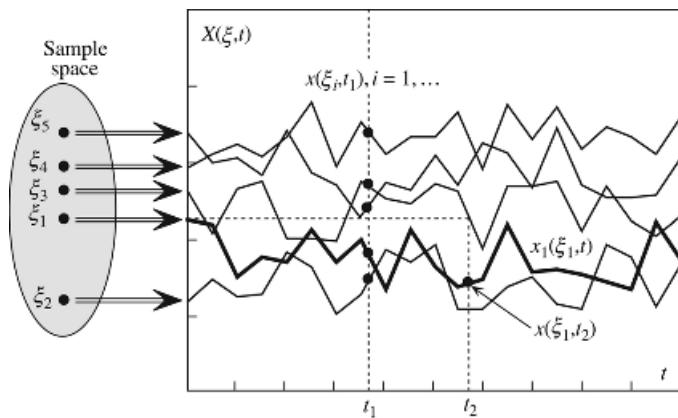


FIGURE 19.1.1

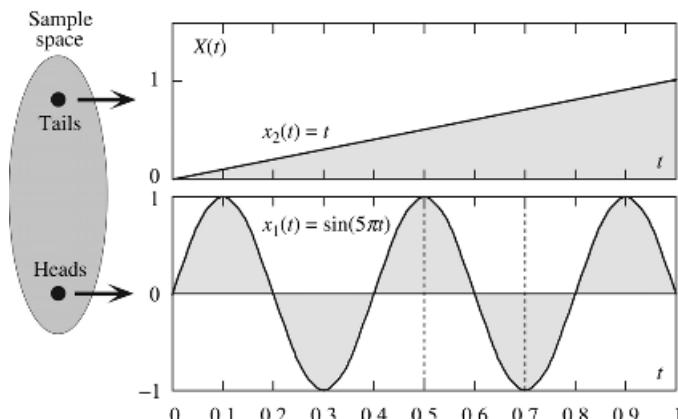


FIGURE 19.1.2

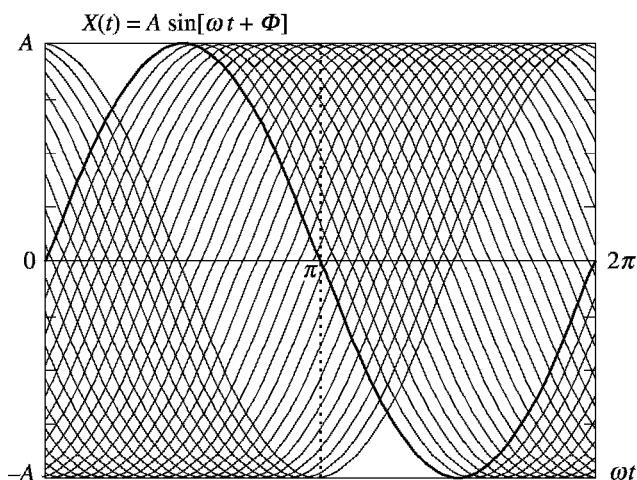


FIGURE 19.1.3

*Distribution and Density Functions.* Since a random process is a random variable for any fixed time  $t$ , we can define a probability distribution and density functions as

$$F_X(x; t) = P(\xi, t: X(\xi; t) \leq x) \text{ for a fixed } t \quad (19.1.1)$$

and

$$\begin{aligned} f_X(x; t) &= \frac{\partial}{\partial x} F_X(x; t) = \lim_{\Delta x \rightarrow 0} \frac{F_X(x + \Delta x; t) - F_X(x; t)}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} P(x < X(t) \leq x + \Delta x) \end{aligned} \quad (19.1.2)$$

These are also called *first-order* distribution and density functions, and in general, they are functions of time.

*Means and Variances.* Analogous to random variables, we can define the mean of a random process as

$$\mu_X(t) = E[X(t)] = \int_{-\infty}^{\infty} x f_X(x; t) dx \quad (19.1.3)$$

and the variance as

$$\begin{aligned} \sigma_X^2(t) &= E[X(t) - \mu_X(t)]^2 = E[X^2(t)] - \mu_X^2(t) \\ &= \int_{-\infty}^{\infty} [x - \mu_X(t)]^2 f_X(x; t) dx \end{aligned} \quad (19.1.4)$$

where

$$E[X^2(t)] = \int_{-\infty}^{\infty} x^2 f_X(x; t) dx \quad (19.1.5)$$

Since the density is a function of time, the means and variances of random processes are also functions of time.

**Example 19.1.3** We shall now find the distribution and density functions along with the mean and variance for the random process of Example 19.1.1 for times  $t = 0, \frac{1}{2}, \frac{7}{10}$ :

$$t = 0, \quad x_1(0) = 0, \quad x_2(0) = 0$$

At  $t = 0$  the mapping diagram from the sample space to the real line is shown in Fig. 19.1.4a along with the corresponding distribution and density functions.

The mean value is given by  $\mu_X(0) = 0; \frac{1}{2} + 0 \cdot \frac{1}{2} = 0$ . The variance is given by  $\sigma_X^2(0) = (0 - 0)^2 \frac{1}{2} + (0 - 0)^2 \frac{1}{2} = 0$ :

$$t = \frac{1}{2}, \quad x_1\left(\frac{1}{2}\right) = 1, \quad x_2\left(\frac{1}{2}\right) = \frac{1}{2}$$

At  $t = \frac{1}{2}$  the mapping diagram from the sample space to the real line is shown in Fig. 19.1.4b along with the corresponding distribution and density functions.

The mean value is given by  $\mu_X\left(\frac{1}{2}\right) = \frac{1}{2} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{3}{4} = 0.75$ .

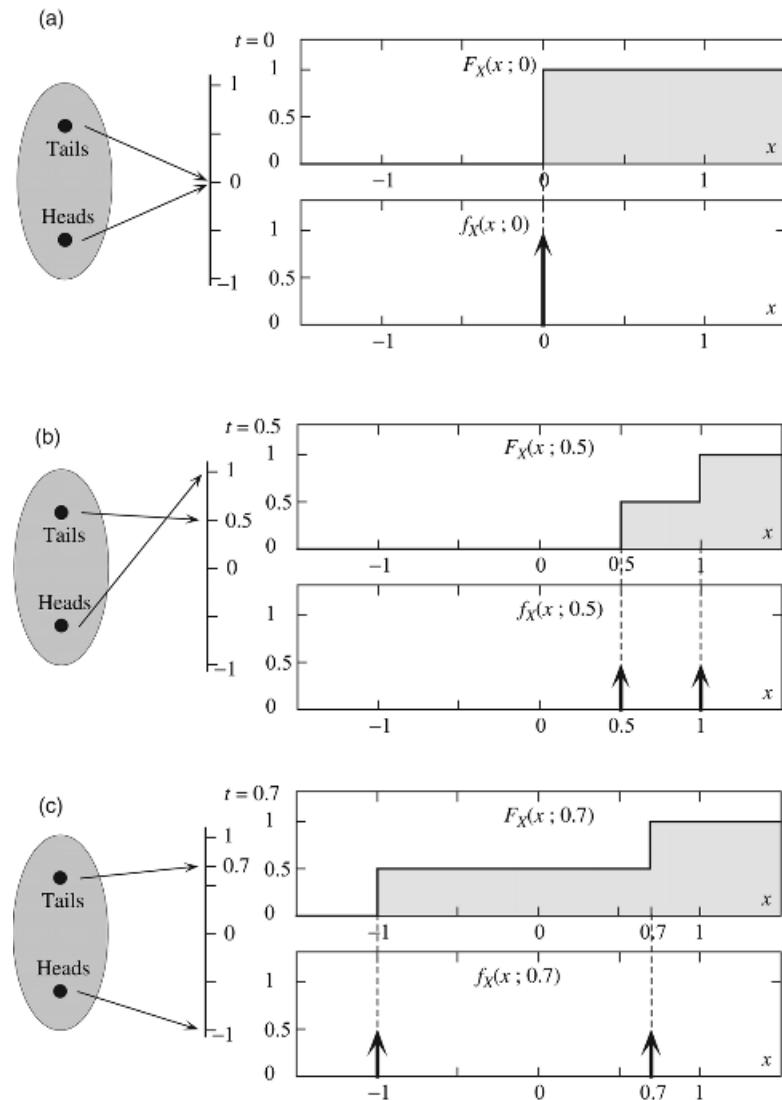


FIGURE 19.1.4

The variance is given by  $\sigma_X^2\left(\frac{1}{2}\right) = \left(\frac{1}{2} - \frac{3}{4}\right)^2 \frac{1}{2} + \left(1 - \frac{3}{4}\right)^2 \frac{1}{2} = \frac{1}{16} = 0.0625$ :

$$t = \frac{7}{10}, \quad x_1\left(\frac{7}{10}\right) = 1, \quad x_2\left(\frac{7}{10}\right) = \frac{7}{10}$$

At  $t = \frac{7}{10}$  the mapping diagram from the sample space to the real line is shown in Fig. 19.1.4c along with the corresponding distribution and density functions.

The mean value is given by  $\mu_X\left(\frac{7}{10}\right) = \frac{7}{10} \cdot \frac{1}{2} - 1 \cdot \frac{1}{2} = -\frac{3}{20} = -0.15$

The variance is given by  $\sigma_X^2\left(\frac{7}{10}\right) = \left(\frac{7}{10} + \frac{3}{20}\right)^2 \frac{1}{2} + \left(-1 + \frac{3}{20}\right)^2 \frac{1}{2} = \frac{289}{400} = 0.7225$ .

**Example 19.1.4** A random process, given by  $X(t) = A \sin(\omega t)$ , is shown in Fig. 19.1.5, where  $A$  is a random variable uniformly distributed in the interval  $(0,1]$ .

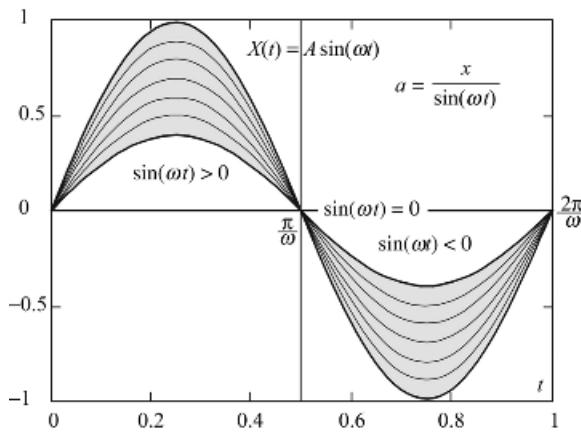


FIGURE 19.1.5

The density and distribution functions of  $A$  are

$$f_A(a) = \begin{cases} 1, & 0 < a \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad F_A(a) = \begin{cases} 0, & a \leq 0 \\ a, & 0 < a \leq 1 \\ 1, & a > 1 \end{cases}$$

We have to find the distribution function  $F_X(x; t)$ . For any given  $t$ ,  $x = a \sin(\omega t)$  is an equation to a straight line with slope  $\sin(\omega t)$ , and hence we can use the results of Examples 12.2.1 and 12.2.2 to solve for  $F_X(x; t)$ . The cases of  $\sin(\omega t) > 0$  and  $\sin(\omega t) < 0$  are shown in Fig. 19.1.6.

*Case 1:*  $\sin(\omega t) > 0$ . There are no points of intersection on the  $a$  axis for  $x \leq 0$ , and hence  $F_X(x; t) = 0$ . For  $0 < x \leq \sin(\omega t)$  we solve  $x = a \sin(\omega t)$  and obtain  $a = x/[\sin(\omega t)]$ . The region  $I_a$  for which  $a \sin(\omega t) \leq x$  is given by  $I_a = \{0 < a \leq$

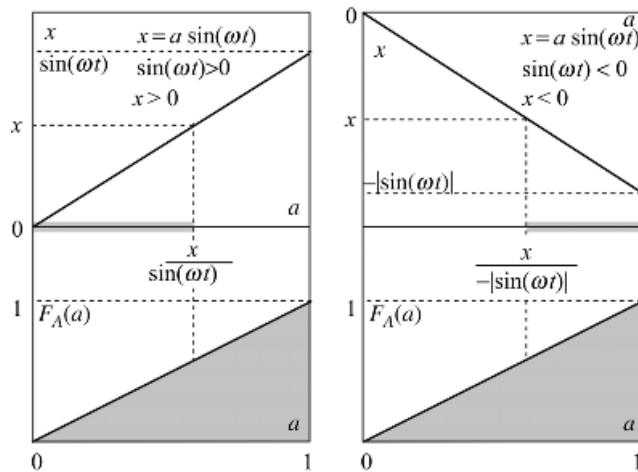


FIGURE 19.1.6

$x/[\sin(\omega t)]\}$  (Fig. 19.1.6). Thus

$$F_X(x; t) = F_A\left(\frac{x}{\sin(\omega t)}\right) - F_A(0) = F_A\left(\frac{x}{\sin(\omega t)}\right) = \frac{x}{\sin(\omega t)}$$

Finally for  $x > \sin(\omega t)$ , the region  $I_a$  for which  $a\sin(\omega t) \leq x$ , is given by  $I_a = \{0 < a \leq 1\}$  and  $F_X(x; t) = 1$ . Thus, for  $\sin(\omega t) > 0$ , we have

$$F_X(x; t) = \begin{cases} 0, & x \leq 0 \\ \frac{x}{\sin(\omega t)}, & 0 < x \leq \sin(\omega t) \\ 1, & x > \sin(\omega t) \end{cases}$$

*Case 2:*  $\sin(\omega t) < 0$ . The region  $I_a$  for which  $x > 0$  is given by  $I_a = \{0 < a \leq 1\}$ , and hence  $F_X(x; t) = 1$ . For  $-|\sin(\omega t)| < x \leq 0$ , we solve  $x = -a|\sin(\omega t)|$  and obtain  $a = [x/(-|\sin(\omega t)|)]$ . The region  $I_a$  for which  $-a|\sin(\omega t)| \leq x$  is given by

$$I_a = \left\{ \frac{x}{-|\sin(\omega t)|} < a \leq 1 \right\}$$

(Fig. 19.1.6). Thus,

$$F_X(x; t) = F_A(1) - F_A\left(\frac{x}{-|\sin(\omega t)|}\right) = 1 - \frac{x}{-|\sin(\omega t)|}$$

Finally, for  $x \leq -|\sin(\omega t)|$ , the region  $I_a$  for which  $-a|\sin(\omega t)| \leq x$  is given by  $I_a = \{1 < a \leq \infty\}$  and  $F_X(x; t) = 0$ . Thus, for  $\sin(\omega t) < 0$ , we have

$$F_X(x; t) = \begin{cases} 0, & x \leq -|\sin(\omega t)| \\ 1 - \frac{x}{-|\sin(\omega t)|}, & -|\sin(\omega t)| < x \leq 0 \\ 1, & x > 0 \end{cases}$$

*Case 3:*  $\sin(\omega t) = 0$ . The region  $I_a$  for which  $x > 0$  is given by  $I_a = \{0 < a \leq 1\}$  and  $F_X(x; t) = 1$ . For  $x \leq 0$ ,  $I_a = \emptyset$  and  $F_X(x; t) = 0$ . Thus, for  $\sin(\omega t) = 0$ , we have

$$F_X(x; t) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$$

**Example 19.1.5** We shall now find the distribution and density functions along with the mean and variance for the random process of Example 19.1.2 and see how this process differs from the previous ones.

We are given that  $X(t) = A \sin(\omega t + \Phi)$ , where  $A$  is a constant and  $f_\Phi(\phi) = 1/2\pi$  in the interval  $(0, 2\pi)$  and we have to find  $f_X(x; t)$ . We will solve this problem by (1) finding the distribution  $F_X(x; t)$  and differentiating it, and (2) by direct determination of the density function.

1. *Determination of Distribution Function  $F_X(x; t)$ .* The distribution function for  $\Phi$  is given by  $F_\Phi(\phi) = (\phi/2\pi)$ ,  $0 < \phi \leq 2\pi$ . The two solutions for  $x = A \sin(\omega t + \Phi)$  are obtained from the two equations:

$$\sin(\omega t + \phi_1) = \frac{x}{A} \quad \text{and} \quad \sin(\pi - \omega t - \phi_2) = \frac{x}{A}$$

Hence the solutions are given by

$$\phi_1 = \sin^{-1}\left(\frac{x}{A}\right) - \omega t \quad \text{and} \quad \phi_2 = \pi - \sin^{-1}\left(\frac{x}{A}\right)$$

and are shown in Fig. 19.1.7.

For  $x \leq -A$ , there are no points of intersection and hence  $F_X(x; t) = 0$ . For  $-A < x \leq A$ , the set of points along the  $\phi$  axis such that  $A \sin(\omega t + \phi) \leq x$  is  $(0, \phi_1] \cup (\phi_2, 2\pi]$ . Hence  $F_X(x; t)$  is given by

$$\begin{aligned} F_X(x; t) &= F_\Phi(\phi_1) - F_\Phi(0) + F_\Phi(2\pi) - F_\Phi(\phi_2) \\ &= \frac{1}{2\pi} \left\{ \sin^{-1}\left(\frac{x}{A}\right) - \omega t - 0 + 2\pi - \left[ \pi - \sin^{-1}\left(\frac{x}{A}\right) - \omega t \right] \right\} \\ &= \frac{1}{2\pi} \left\{ 2 \sin^{-1}\left(\frac{x}{A}\right) + \pi \right\} = \frac{1}{\pi} \sin^{-1}\left(\frac{x}{A}\right) + \frac{1}{2}, \quad -A < x \leq A \end{aligned}$$

Finally, for  $x > A$ , the entire curve  $A \sin(\omega t + \phi)$  is below  $x$ , and  $F_X(x; t) = 1$ .

## 2. Determination of Density Function $f_X(x; t)$

(a) The two solutions to  $x = A \sin(\omega t + \phi)$  have been found earlier.

(b) The absolute derivatives  $|\partial x / \partial \phi||_{\phi_1}$  and  $|\partial x / \partial \phi||_{\phi_2}$  are given by

$$\begin{aligned} \left| \frac{\partial x}{\partial \phi} \right|_{\phi_1} &= A \cos(\omega t + \phi_1) = A \cos\left[\omega t + \sin^{-1}\left(\frac{x}{A}\right) - \omega t\right] \\ &= A \cos\left[\sin^{-1}\left(\frac{x}{A}\right)\right] = A \frac{\sqrt{A^2 - x^2}}{A} = \sqrt{A^2 - x^2} \end{aligned}$$

and

$$\begin{aligned} \left| \frac{\partial x}{\partial \phi} \right|_{\phi_2} &= A \cos(\omega t + \phi_2) = \left| A \cos\left[\omega t + \pi - \sin^{-1}\left(\frac{x}{A}\right) - \omega t\right] \right| \\ &= \left| A \cos\left[\pi - \sin^{-1}\left(\frac{x}{A}\right)\right] \right| = A \frac{\sqrt{A^2 - x^2}}{A} = \sqrt{A^2 - x^2} \end{aligned}$$

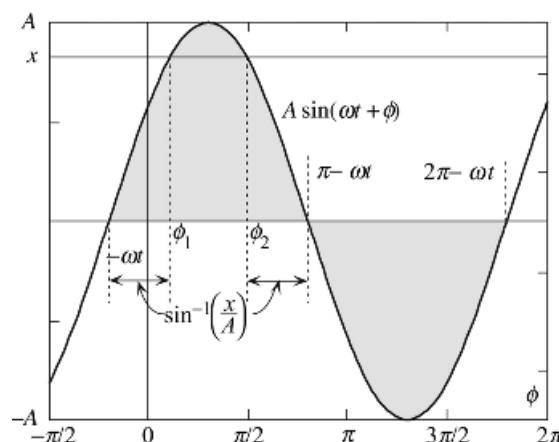


FIGURE 19.1.7

- (c) With the two solutions for  $x$ , the density function  $f_X(x; t)$  is given by Eq. (12.3.6):

$$f_X(x; t) = \frac{1}{\sqrt{A^2 - x^2}} \left( \frac{1}{2\pi} + \frac{1}{2\pi} \right) = \frac{1}{\pi\sqrt{A^2 - x^2}}, \quad -A < x \leq A$$

Integration of  $f_X(x; t)$  gives the distribution function  $F_X(x; t)$

$$F_X(x; t) = \begin{cases} 0, & x \leq -A \\ \frac{1}{\pi} \sin^{-1} \left( \frac{x}{A} \right) + \frac{1}{2}, & -A < x \leq A \\ 1, & x > A \end{cases}$$

and these two solutions are exactly the same as before. For this random process, we find that the density and the distribution functions are both independent of time. For  $A = 1$ , they become

$$f_X(x; t) = \frac{1}{\pi\sqrt{1 - x^2}}, \quad -1 < x \leq 1$$

$$F_X(x; t) = \begin{cases} 0, & x \leq -1 \\ \frac{\sin^{-1}(x)}{\pi} + \frac{1}{2}, & -1 < x \leq 1 \\ 1, & x > 1 \end{cases}$$

and these functions are shown in Fig. 19.1.8. Since  $f_X(x)$  has even symmetry, the mean value is 0 and the variance is obtained from

$$\sigma_X^2 = \int_{-1}^1 \frac{x^2}{\pi\sqrt{1 - x^2}} dx = \frac{1}{2}$$

The value of  $\sigma = \pm 1/\sqrt{2}$  is also shown in Fig. 19.1.8.

Later we will discuss random processes whose density and distribution functions are independent of time.

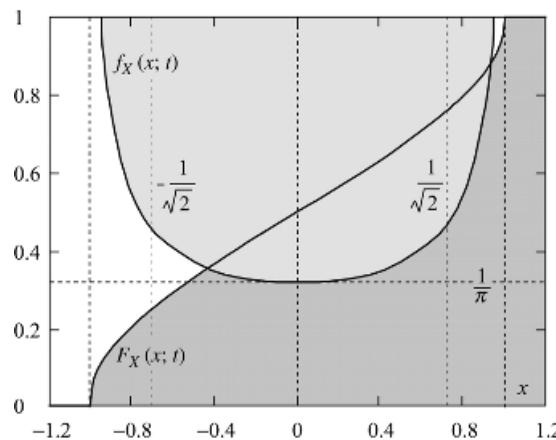


FIGURE 19.1.8

### Higher-Order Distribution Functions

If  $t_1$  and  $t_2$  are different times, then  $X_1 = X(t_1)$  and  $X_2 = X(t_2)$  are two different random variables as shown in Fig. 19.1.9.

A second-order distribution function  $F_X(x_1, x_2; t_1, t_2)$  for  $X_1$  and  $X_2$  can be defined as

$$F_X(x_1, x_2; t_1, t_2) = P[X(t_1) \leq x_1, X(t_2) \leq x_2] \quad (19.1.6)$$

and the second-order density function  $f_X(x_1, x_2; t_1, t_2)$  as

$$f_X(x_1, x_2; t_1, t_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_X(x_1, x_2; t_1, t_2) \quad (19.1.7)$$

Similarly, if  $t_1, \dots, t_n$  are different times, then an  $n$ th-order distribution function is defined as

$$F_X(x_1, \dots, x_n; t_1, \dots, t_n) = P[X(t_1) \leq x_1, \dots, X(t_n) \leq x_n] \quad (19.1.8)$$

and the  $n$ th-order density function as

$$f_X(x_1, \dots, x_n; t_1, \dots, t_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F_X(x_1, \dots, x_n; t_1, \dots, t_n) \quad (19.1.9)$$

A random variable is completely defined if its distribution function is known. Similarly, a random process  $X(t)$  is completely defined if its  $n$ th-order distribution is known for all  $n$ . Since this is not feasible for all  $n$ , a random process in general cannot be completely defined. Hence, we usually restrict the definition to second-order distribution function, in which case it is called a *second-order process*.

In a similar manner, we can define a joint distribution between two different random processes  $X(t)$  and  $Y(t)$  (Fig. 19.1.10) as given below, where the joint second-order distribution function  $F_{XY}(x_1, y_2; t_1, t_2)$  for  $X(t_1)$  and  $Y(t_2)$  is defined by

$$F_{XY}(x_1, y_2; t_1, t_2) = P[X(t_1) \leq x_1, Y(t_2) \leq y_2] \quad (19.1.10)$$

and the joint density function  $f_{XY}(x_1, y_2; t_1, t_2)$  as

$$f_{XY}(x_1, y_2; t_1, t_2) = \frac{\partial^2}{\partial x_1 \partial y_2} F_{XY}(x_1, y_2; t_1, t_2) \quad (19.1.11)$$

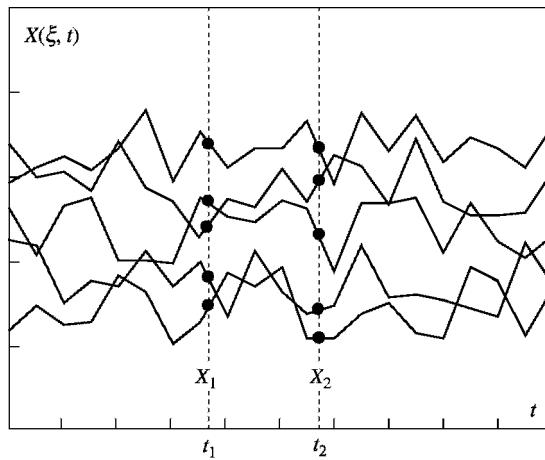


FIGURE 19.1.9

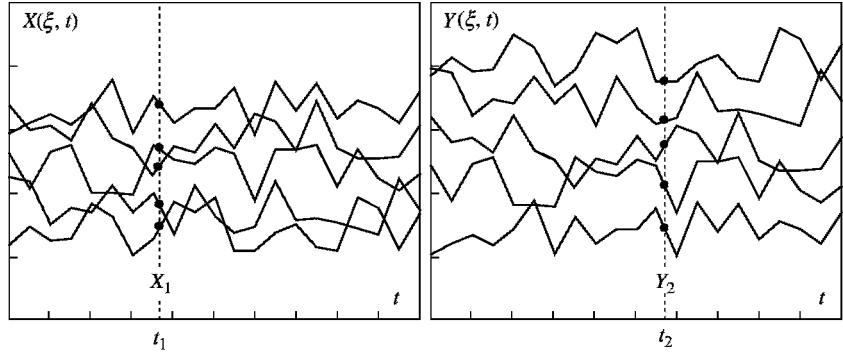


FIGURE 19.1.10

The joint  $n$ th-order distribution function  $F_{XY}(x_1, y_2; t_1, t_2)$  for  $X(t)$  and  $Y(t)$  is defined by

$$\begin{aligned} F_{XY}(x_1, \dots, x_n; y_1, \dots, y_n; t_1, \dots, t_n) \\ = P[X(t_1) \leq x_1, \dots, X(t_n) \leq x_n; Y(t_1) \leq y_1, \dots, Y(t_n) \leq y_n] \end{aligned} \quad (19.1.12)$$

and the corresponding  $n$ th-order density function, by

$$\begin{aligned} f_{XY}(x_1, \dots, x_n; y_1, \dots, y_n; t_1, \dots, t_n) \\ = \frac{\partial^{2n}}{\partial x_1 \dots \partial x_n \partial y_1 \dots \partial y_n} F_{XY}(x_1, \dots, x_n; y_1, \dots, y_n; t_1, \dots, t_n) \end{aligned} \quad (19.1.13)$$

### Second Order Moments

In Eqs. (19.1.3) and (19.1.4) defined mean and variance for a random process. The second moment of a random process has also been defined in Eq. (19.1.5). Since  $X(t_1)$  and  $X(t_2)$  are random variables, various types of joint moments can be defined.

*Autocorrelation.* The autocorrelation function (AC)  $R_X(t_1, t_2)$  is defined as the expected value of the product  $X(t_1)$  and  $X(t_2)$ :

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_X(x_1, x_2; t_1, t_2) dx_1 dx_2 \quad (19.1.14)$$

By substituting  $t_1 = t_2 = t$  in Eq. (19.1.14), we can obtain the second moment or the *average power* of the random process:

$$R_X(t) = E[X^2(t)] = \int_{-\infty}^{\infty} x^2 f_X(x; t) dx \quad (19.1.15)$$

*Autocovariance.* The autocovariance function (ACF)  $C_X(t_1, t_2)$  is defined as the covariance between  $X(t_1)$  and  $X(t_2)$ :

$$\begin{aligned} C_X(t_1, t_2) &= E[(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu_X(t_1))(x_2 - \mu_X(t_2)) f_X(x_1, x_2; t_1, t_2) dx_2 dx_1 \\ &= E[X(t_1)X(t_2)] - \mu_X(t_1)\mu_X(t_2) \end{aligned} \quad (19.1.16)$$

Thus, from Eqs. (19.1.15) and (19.1.16) the interrelationships between AC and ACF are

$$\begin{aligned} C_X(t_1, t_2) &= R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2) \\ R_X(t_1, t_2) &= C_X(t_1, t_2) + \mu_X(t_1)\mu_X(t_2) \end{aligned} \quad (19.1.17)$$

*Normalized Autocovariance.* The normalized autocovariance function (NACF)  $\rho_X(t_1, t_2)$  is the ACF normalized by the variance and is defined by

$$\rho_X(t_1, t_2) = \frac{C_X(t_1, t_2)}{\sqrt{\sigma_X^2(t_1)\sigma_X^2(t_2)}} = \frac{C_X(t_1, t_2)}{\sigma_X(t_1)\sigma_X(t_2)} \quad (19.1.18)$$

The NACF finds wide applicability in many problems in random processes, particularly in time-series analysis.

The following definitions pertain to two different random processes  $X(t)$  and  $Y(t)$ :

*Cross-Correlation.* The cross-correlation function (CC)  $R_{XY}(t_1, t_2)$  is defined as the expected value of the product  $X(t_1)$  and  $Y(t_2)$ :

$$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 y_2 f_{XY}(x_1, y_2; t_1, t_2) dy_2 dx_1 \quad (19.1.19)$$

By substituting  $t_1 = t_2 = t$  in Eq. (19.1.19), we can obtain the joint moment between the random processes  $X(t)$  and  $Y(t)$  as

$$R_{XY}(t) = E[X(t)Y(t)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x, y; t) dy dx \quad (19.1.20)$$

*Cross-Covariance.* The cross-covariance function (CCF)  $C_{XY}(t_1, t_2)$  is defined as the covariance between  $X(t_1)$  and  $Y(t_2)$ :

$$\begin{aligned} C_{XY}(t_1, t_2) &= E[(X(t_1) - \mu_X(t_1))(Y(t_2) - \mu_Y(t_2))] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu_X(t_1))(y_2 - \mu_Y(t_2)) f_{XY}(x_1, y_2; t_1, t_2) dy_2 dx_1 \\ &= E[X(t_1)Y(t_2)] - \mu_X(t_1)\mu_Y(t_2) \end{aligned} \quad (19.1.21)$$

Thus, from Eqs. (19.1.19) and (19.1.21) the interrelationships between CC and CCF are

$$\begin{aligned} C_{XY}(t_1, t_2) &= R_{XY}(t_1, t_2) - \mu_X(t_1)\mu_Y(t_2) \\ R_{XY}(t_1, t_2) &= C_{XY}(t_1, t_2) + \mu_X(t_1)\mu_Y(t_2) \end{aligned} \quad (19.1.22)$$

*Normalized Cross-Covariance.* The normalized cross-covariance function (NCCF)  $\rho_{XY}(t_1, t_2)$  is the CCF normalized by the variances and is defined by

$$\rho_{XY}(t_1, t_2) = \frac{C_{XY}(t_1, t_2)}{\sqrt{\sigma_X^2(t_1)\sigma_Y^2(t_2)}} = \frac{C_{XY}(t_1, t_2)}{\sigma_X(t_1)\sigma_Y(t_2)} \quad (19.1.23)$$

### Some Properties of $X(t)$ and $Y(t)$

Two random processes  $X(t)$  and  $Y(t)$  are *independent* if for all  $t_1$  and  $t_2$

$$F_{XY}(x, y; t_1, t_2) = F_X(x; t_1)F_Y(y; t_2) \quad (19.1.24)$$

They are *uncorrelated* if

$$C_{XY}(t_1, t_2) = R_{XY}(t_1, t_2) - \mu_X(t_1)\mu_Y(t_2) = 0$$

or

$$R_{XY}(t_1, t_2) = \mu_X(t_1)\mu_Y(t_2) \quad \text{for all } t_1 \text{ and } t_2 \quad (19.1.25)$$

They are *orthogonal* if for all  $t_1$  and  $t_2$

$$R_{XY}(t_1, t_2) = 0 \quad (19.1.26)$$

**Example 19.1.6** This example is slightly different from Example 19.1.4. A random process  $X(t)$  is given by  $X(t) = A \sin(\omega t + \phi)$  as shown in Fig. 19.1.11, where  $A$  is a uniformly distributed random variable with mean  $\mu_A$  and variance  $\sigma_A^2$ . We will find the mean, variance, autocorrelation, autocovariance, and normalized autocovariance of  $X(t)$ .

*Mean:*

$$E[X(t)] = \mu_X(t) = E[A \sin(\omega t + \phi)] = \mu_A \sin(\omega t + \phi)$$

*Variance:*

$$\begin{aligned} \text{var}[X(t)] &= \sigma_X^2(t) = E[A^2 \sin^2(\omega t + \phi)] - \mu_A^2 \sin^2(\omega t + \phi) \\ &= \{E[A^2] - \mu_A^2\} \sin^2(\omega t + \phi) = \sigma_A^2 \sin^2(\omega t + \phi) \end{aligned}$$

*Autocorrelation.* From Eq. (19.1.14) we have

$$\begin{aligned} R_X(t_1, t_2) &= E[X(t_1)X(t_2)] = E[A^2] \sin(\omega t_1 + \phi) \sin(\omega t_2 + \phi) \\ &= \frac{1}{2} E[A^2] \{\cos[\omega(t_1 - t_2)] - \cos[\omega(t_1 + t_2) + 2\phi]\} \end{aligned}$$

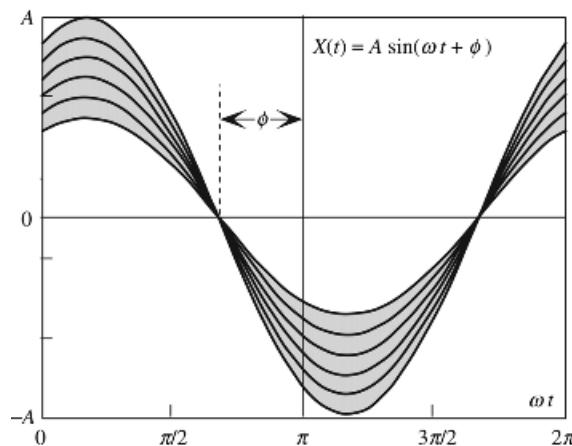


FIGURE 19.1.11

*Autocovariance.* From Eq. (19.1.16) we have

$$\begin{aligned}
 C_X(t_1, t_2) &= R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2) \\
 &= E[A^2] \sin(\omega t_1 + \phi) \sin(\omega t_2 + \phi) - \mu_A^2 \sin(\omega t_1 + \phi) \sin(\omega t_2 + \phi) \\
 &= \sigma_A^2 \sin(\omega t_1 + \phi) \sin(\omega t_2 + \phi) \\
 &= \frac{1}{2} \sigma_A^2 \{ \cos[\omega(t_1 - t_2)] - \cos[\omega(t_1 + t_2) + 2\phi] \}
 \end{aligned}$$

*Normalized Autocovariance.* From Eq. (19.1.18) we have

$$\rho_X(t_1, t_2) = \frac{C_X(t_1, t_2)}{\sigma_X(t_1)\sigma_X(t_2)} = \frac{\sigma_A^2 \sin(\omega t_1 + \phi) \sin(\omega t_2 + \phi)}{\sigma_A^2 \sin(\omega t_1 + \phi) \sin(\omega t_2 + \phi)} = 1$$

**Example 19.1.7** A random process  $X(t)$  with  $k$  changes in a time interval  $t$ , and its probability mass function is given by  $p(k; \lambda) = e^{-\lambda t}[(\lambda t)^k/k!]$ . It is also known that the joint probability  $P\{k_1$  changes in  $t_1, k_2$  changes in  $t_2\}$  is given by

$$\begin{aligned}
 P\{k_1 \text{ changes in } t_1, k_2 \text{ changes in } t_2\} &= P\{k_1 \text{ changes in } t_1, (k_2 - k_1) \text{ changes in } (t_2 - t_1)\} \\
 &= P\{k_1 \text{ changes in } t_1\} P\{(k_2 - k_1) \text{ changes in } (t_2 - t_1)\} \\
 &= e^{-\lambda t_1} \frac{(\lambda t_1)^{k_1}}{k_1!} e^{-\lambda(t_2 - t_1)} \frac{[\lambda(t_2 - t_1)]^{(k_2 - k_1)}}{(k_2 - k_1)!}
 \end{aligned}$$

We have to find the mean, variance, autocorrelation, autocovariance, and normalized autocovariance of  $X(t)$ :

*Mean:*

$$E[X(t)] = \mu_X(t) = \sum_{k=0}^{\infty} k e^{-\lambda t} \frac{(\lambda t)^k}{k!} = \lambda t$$

*Variance:*

$$\text{var}[X(t)] = \sigma_X^2(t) = \sum_{k=0}^{\infty} k^2 e^{-\lambda t} \frac{(\lambda t)^k}{k!} - (\lambda t)^2 = \lambda t$$

*Autocorrelation.* From Eq. (19.1.18) we have

$$\begin{aligned}
 R_X(t_1, t_2) &= E[X(t_1)X(t_2)] = E\{X(t_1)[X(t_2) - X(t_1) + X(t_1)]\} \\
 &= E[X^2(t_1)] + E[X(t_1)]E[X(t_2) - X(t_1)] \quad (\text{from condition given}) \\
 &= (\lambda t_1)^2 + \lambda t_1 + \lambda t_1 \lambda (t_2 - t_1) \\
 &= \lambda^2 t_1 t_2 + \lambda t_1 \text{ if } t_2 > t_1
 \end{aligned}$$

and we have a similar result if  $t_1 > t_2$ :

$$R_X(t_1, t_2) = \lambda^2 t_1 t_2 + \lambda t_2 \quad \text{if } t_1 > t_2$$

Combining these two results, we have

$$R_X(t_1, t_2) = \lambda^2 t_1 t_2 + \lambda \min(t_1, t_2)$$

*Autocovariance.* From Eq. (19.1.19) we have

$$\begin{aligned} C_X(t_1, t_2) &= R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2) \\ &= \lambda^2 t_1 t_2 + \lambda \min(t_1, t_2) - \lambda^2 t_1 t_2 = \lambda \min(t_1, t_2) \end{aligned}$$

*Normalized Autocovariance.* From Eq. (19.1.18) we have

$$\rho_X(t_1, t_2) = \frac{C_X(t_1, t_2)}{\sigma_X(t_1)\sigma_X(t_2)} = \frac{\min(t_1, t_2)}{\lambda\sqrt{t_1 t_2}} = \frac{1}{\lambda} \min\left(\sqrt{\frac{t_1}{t_2}}, \sqrt{\frac{t_2}{t_1}}\right)$$

**Example 19.1.8** Two random processes  $X(t)$  and  $Y(t)$  are given by

$$X(t) = A \sin(\omega t + \phi_1); \quad Y(t) = B \sin(\omega t + \phi_2)$$

where  $A$  and  $B$  are two random variables with parameters  $E[A] = \mu_A$ ,  $E[B] = \mu_B$ ,  $\text{var}[A] = \sigma_A^2$ ,  $\text{var}[B] = \sigma_B^2$ ,  $\text{cov}[AB] = \sigma_{AB}$ , and correlation coefficient  $\rho_{AB} = \sigma_{AB}/(\sigma_A \sigma_B)$ . We have to find the means and variances of  $X(t)$  and  $Y(t)$  and their cross-correlation, cross-covariance, and normalized cross-covariance. The means and variances can be obtained directly from Example 19.1.5.

*Means:*

$$\begin{aligned} \mu_X(t) &= E[A \sin(\omega t + \phi_1)] = \mu_A \sin(\omega t + \phi_1) \\ \mu_Y(t) &= E[B \cos(\omega t + \phi_2)] = \mu_B \cos(\omega t + \phi_2) \end{aligned}$$

*Variances:*

$$\begin{aligned} \sigma_X^2(t) &= \sigma_A^2 \sin^2(\omega t + \phi_1) \\ \sigma_Y^2(t) &= \sigma_B^2 \cos^2(\omega t + \phi_2) \end{aligned}$$

*Cross-Correlation:*

$$\begin{aligned} R_{XY}(t_1, t_2) &= E[X(t_1)Y(t_2)] = E[AB] \sin(\omega t_1 + \phi_1) \cos(\omega t_2 + \phi_2) \\ &= \frac{1}{2} E[AB] \{ \sin[\omega(t_1 + t_2) + \phi_1 + \phi_2] + \sin[\omega(t_1 - t_2) + \phi_1 - \phi_2] \} \end{aligned}$$

*Cross-Covariance:*

$$\begin{aligned} C_{XY}(t_1, t_2) &= R_{XY}(t_1, t_2) - \mu_X(t_1)\mu_Y(t_2) \\ &= E[AB] \sin(\omega t_1 + \phi_1) \cos(\omega t_2 + \phi_2) - \mu_A \mu_B \sin(\omega t_1 + \phi_1) \cos(\omega t_2 + \phi_2) \\ &= \sigma_{AB} \sin(\omega t_1 + \phi_1) \cos(\omega t_2 + \phi_2) \\ &= \frac{1}{2} \sigma_{AB} \{ \sin[\omega(t_1 + t_2) + \phi_1 + \phi_2] + \sin[\omega(t_1 - t_2) + \phi_1 - \phi_2] \} \end{aligned}$$

*Normalized Cross-Covariance:*

$$\rho_{XY}(t_1, t_2) = \frac{C_{XY}(t_1, t_2)}{\sigma_X(t_1)\sigma_Y(t_2)} = \frac{\sigma_{AB} \sin(\omega t_1 + \phi_1) \cos(\omega t_2 + \phi_2)}{\sigma_A \sigma_B \sin(\omega t_1 + \phi_1) \cos(\omega t_2 + \phi_2)} = \frac{\sigma_{AB}}{\sigma_A \sigma_B} = \rho_{AB}$$

## 19.2 STATIONARY RANDOM PROCESSES

The distribution functions of all the random processes except that in Example 19.1.4 were dependent on time. However, many of the random processes such as Example 19.1.4 have the important property that their statistics do not change with time, which is an important step toward obtaining the statistics from a single sample function. The statistics in the time interval  $(t_1, t_2)$  is the same as in the time interval  $(t_1 + \tau, t_2 + \tau)$ . In other words, the probabilities of the samples of a random process  $X(t)$  at times  $t_1, \dots, t_n$  will not differ from those at times  $t_1 + \tau, \dots, t_n + \tau$ . This means that the joint distribution function of  $X(t_1), \dots, X(t_n)$  is the same as  $X(t_1 + \tau), \dots, X(t_n + \tau)$ , or

$$F_X(x_1, \dots, x_n; t_1, \dots, t_n) = F_X(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau) \quad (19.2.1)$$

and the corresponding density function may be written as

$$f_X(x_1, \dots, x_n; t_1, \dots, t_n) = f_X(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau) \quad (19.2.2)$$

Random processes with the property of Eq. (19.2.1) or (19.2.2) are called *n*th-order *stationary processes*. A *strict-sense* or *strongly* stationary process is a random process that satisfies Eqs. (19.2.1) and (19.2.2) for all *n*. Analogously, we can also define lower orders of stationarity.

A random process is *first-order stationary* if

$$\begin{aligned} F_X(x; t) &= F_X(x; t + \tau) = F_X(x) \\ f_X(x; t) &= f_X(x; t + \tau) = f_X(x) \end{aligned} \quad (19.2.3)$$

and the distribution and density functions are *independent* of time. The random process in Examples 19.1.2 and 19.1.5 is an example of a first-order stationary process.

A random process is *second-order stationary* if

$$\begin{aligned} F_X(x_1, x_2; t_1, t_2) &= F_X(x_1, x_2; t_1 + \tau, t_2 + \tau) = F_X(x_1, x_2; \tau) \\ f_X(x_1, x_2; t_1, t_2) &= f_X(x_1, x_2; t_1 + \tau, t_2 + \tau) = f_X(x_1, x_2; \tau) \end{aligned} \quad (19.2.4)$$

The distribution and density functions are dependent not on two time instants  $t_1$  and  $t_2$  but on the time difference  $\tau = t_1 - t_2$  only. Second-order stationary processes are also called *wide-sense stationary* or *weakly stationary*. Hereafter, *stationary* means wide-sense stationary, and strict-sense stationary will be specifically mentioned.

In a similar manner, two processes  $X(t)$  and  $Y(t)$  are jointly stationary if for all *n*

$$\begin{aligned} F_{XY}(x_1, \dots, x_n; y_1, \dots, y_n; t_1, \dots, t_n) \\ = F_{XY}(x_1, \dots, x_n; y_1, \dots, y_n; t_1 + \tau, \dots, t_n + \tau) \end{aligned} \quad (19.2.5)$$

or

$$\begin{aligned} f_{XY}(x_1, \dots, x_n; y_1, \dots, y_n; t_1, \dots, t_n) \\ = f_{XY}(x_1, \dots, x_n; y_1, \dots, y_n; t_1 + \tau, \dots, t_n + \tau) \end{aligned} \quad (19.2.6)$$

and they are jointly wide-sense stationary if

$$\begin{aligned} F_{XY}(x_1, y_2; t_1, t_2) \\ = F_{XY}(x_1, y_2; t_1 + \tau, t_2 + \tau) = F_{XY}(x_1, y_2; \tau) \end{aligned} \quad (19.2.7)$$

and the joint density function

$$f_{XY}(x_1, y_2; t_1, t_2) = f_{XY}(x_1, y_2; t_1 + \tau, t_2 + \tau) = f_{XY}(x_1, y_2; \tau) \quad (19.2.8)$$

*n*th-order stationarity implies lower-order stationarities. Strict-sense stationarity implies wide-sense stationarity.

Similar to Eqs. (19.1.24)–(19.1.26), we can enumerate the following properties for stationary random processes  $X(t)$  and  $Y(t)$ . Two random processes  $X(t)$  and  $Y(t)$  are *independent* if for all  $x$  and  $y$

$$F_{XY}(x, y) = F_X(x)F_Y(y) \quad (19.2.9)$$

They are *uncorrelated* if for all  $\tau$

$$C_{XY}(\tau) = R_{XY}(\tau) - \mu_X\mu_Y = 0$$

or

$$R_{XY}(\tau) = \mu_X\mu_Y \quad (19.2.10)$$

They are *orthogonal* if for all  $\tau$

$$R_{XY}(\tau) = 0 \quad (19.2.11)$$

### Moments of Continuous-Time Stationary Processes

We can now define the various moments for a stationary random process  $X(t)$ .

*Mean:*

$$E[X(t)] = \mu_X = \int_{-\infty}^{\infty} xf(x)dx \quad (19.2.12)$$

*Variance:*

$$E[X(t) - \mu_X]^2 = \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)dx = E[X^2(t)] - \mu_X^2 \quad (19.2.13)$$

*Autocorrelation:*

$$R_X(\tau) = E[X(t)X(t + \tau)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; \tau) dx_1 dx_2 \quad (19.2.14)$$

*Autocovariance:*

$$\begin{aligned} C_X(\tau) &= E\{[X(t) - \mu_X][X(t + \tau) - \mu_X]\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu_X)(x_2 - \mu_X) f(x_1, x_2; \tau) dx_1 dx_2 \\ &= R_X(\tau) - \mu_X^2 \end{aligned} \quad (19.2.15)$$

*Normalized Autocovariance (NACF):*

$$\rho_X(\tau) = \frac{C_X(\tau)}{\sigma_X^2} \quad (19.2.16)$$

### White-Noise Process

A zero mean stationary random process  $X(t)$  whose autocovariance or autocorrelation is given by

$$C_X(\tau) = R_X(\tau) = \sigma_X^2 \delta(\tau) \quad (19.2.17)$$

where  $\delta(\tau)$  is the Dirac delta function, is called a *white-noise process*. The energy of a white-noise process is infinite since  $C_X(0) = R_X(0) = E[X^2(t)] = \infty$ . Hence it is an idealization. White-noise processes find extensive use in modeling communication systems.

The cross-moments of two jointly stationary processes  $X(t)$  and  $Y(t)$  are defined below:

*Cross-Correlation:*

$$R_{XY}(\tau) = E[X(t)Y(t + \tau)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 y_2 f(x_1, y_2; \tau) dy_2 dx_1 \quad (19.2.18)$$

*Cross-Covariance:*

$$\begin{aligned} C_{XY}(\tau) &= E\{[X(t) - \mu_X][Y(t + \tau) - \mu_Y]\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu_X)(y_2 - \mu_Y) f(x_1, y_2; \tau) dy_2 dx_1 \\ &= R_{XY}(\tau) - \mu_X \mu_Y \end{aligned} \quad (19.2.19)$$

*Normalized Cross-Covariance:*

$$\rho_{XY}(\tau) = \frac{C_{XY}(\tau)}{\sigma_X \sigma_Y} \quad (19.2.20)$$

A stationary random process  $X(t)$  is passed through a linear system with impulse response  $h(t)$ . The input–output relationship will be given by the convolution integral:

$$Y(t) = \int_{-\infty}^{\infty} X(t - \alpha)h(\alpha)d\alpha = \int_{-\infty}^{\infty} X(t)h(t - \alpha)d\alpha \quad (19.2.21)$$

The cross-correlation function  $R_{XY}(\tau)$  between the input and the output can be found as follows:

$$E[X(t)Y(t + \tau)] = E \int_{-\infty}^{\infty} X(t)X(t + \tau - \alpha)h(\alpha)d\alpha \quad (19.2.22)$$

or

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} R_X(\tau - \alpha)h(\alpha)d\alpha \quad (19.2.23)$$

Since the cross-correlation function depends only on  $\tau$ , the output  $Y(t)$  will also be a stationary random process.

### Properties of Correlation Functions of Stationary Processes

#### Autocorrelation Functions

1.  $R_X(0) = E[X^2(t)] = \text{average power} \geq 0$
2.  $R_X(\tau) = R_X(-\tau)$ . Or,  $R_X(\tau)$  is an even function.

$$R_X(\tau) = E[X(t)X(t + \tau)] = E[X(t + \tau)X(t)] = R_X(-\tau) \quad (19.2.24)$$

3.  $|R_X(\tau)| \leq R_X(0)$ . Or, the maximum value of  $|R_X(\tau)|$  occurs at  $\tau = 0$  and  $|R_X(\tau)|$  for any  $\tau$  cannot exceed the value  $R_X(0)$ . From

$$E[X(t) \pm X(t + \tau)]^2 \geq 0$$

we have

$$E[X^2(t)] + E[X^2(t + \tau)] \pm 2E[X(t)X(t + \tau)] \geq 0$$

or

$$R_X(0) \geq |R_X(\tau)| \quad (19.2.25)$$

4. If a constant  $T > 0$  exists such that  $R_X(T) = R_X(0)$ , then the  $R_X(\tau)$  is periodic and  $X(t)$  is called a *periodic stationary process*. From Schwartz' inequality [Eq. (14.5.13)] we have

$$\{E[g(X)h(X)]\}^2 \leq E[g^2(X)]E[h^2(X)]$$

Substituting  $g(X) = X(t)$  and  $h(X) = [X(t + \tau + T) - X(t + \tau)]$ , we can write

$$\{E[X(t)[X(t + \tau + T) - X(t + \tau)]]\}^2 \leq E[X^2(t)]E[X(t + \tau + T) - X(t + \tau)]^2$$

or

$$\{R_X(\tau + T) - R_X(\tau)\}^2 \leq 2R_X(0)[R_X(0) - R_X(T)]$$

Hence, if  $R_X(T) = R_X(0)$ , then, since  $\{R_X(\tau + T) - R_X(\tau)\}^2 \geq 0$ , the result  $R_X(\tau + T) = R_X(\tau)$  follows.

5.  $E[X(t + \phi)X(t + \tau + \phi)] = R_X(\tau) = E[X(t)X(t + \tau)]$   
 $E[X(t + \phi)Y(t + \tau + \phi)] = R_{XY}(\tau) = E[X(t)Y(t + \tau)]$

In the formulation of correlation functions the phase information is lost.

6. If  $E[X(t)] = \mu_X$  and  $Y(t) = a + X(t)$ , where  $a$  is constant, then  $E[Y(t)] = a + \mu_X$  and,

$$\begin{aligned} R_Y(\tau) &= E\{[a + X(t)][a + X(t + \tau)]\} \\ &= a^2 + aE[X(t + \tau)] + aE[X(t)] + E[X(t)X(t + \tau)] \\ &= a^2 + 2a\mu_X + R_X(\tau) = a^2 + 2a\mu_X + \mu_X^2 + C_X(\tau) \\ &= (a + \mu_X)^2 + C_X(\tau) \end{aligned} \quad (19.2.26)$$

If  $E[X(t)] = 0$ , then  $E[Y(t)] = a$ , and we can obtain the mean value of  $Y(t)$  from a knowledge of its autocorrelation function  $R_Y(\tau)$ .

### Cross-Correlation Functions

7.  $R_{XY}(\tau) = R_{YX}(-\tau)$ : This is not an even function. (19.2.27)

The result follows from the definition of cross-correlation:

$$R_{XY}(\tau) = E[X(t)Y(t+\tau)] = E[Y(t+\tau)X(t)] = R_{YX}(-\tau) \quad \text{but} \quad R_{XY}(0) = R_{YX}(0) \quad (19.2.28)$$

**8.**  $R_{XY}^2(\tau) \leq R_X(0)R_Y(0)$ : This result follows from Schwartz' inequality:

$$\{E[X(t)Y(t+\tau)]\}^2 \leq E[X^2(t)]E[Y^2(t+\tau)]$$

**9.**  $2|R_{XY}(\tau)| \leq R_X(0) + R_Y(0)$ : The result follows from

$$E[X(t) \pm Y(t+\tau)]^2 = R_X(0) + R_Y(0) \pm 2R_{XY}(\tau) \geq 0 \quad (19.2.29)$$

**10.** If  $Z(t) = X(t) + Y(t)$ , then

$$\begin{aligned} R_Z(\tau) &= E[Z(t)Z(t+\tau)] = E\{[X(t) + Y(t)][X(t+\tau) + Y(t+\tau)]\} \\ &= R_X(\tau) + R_Y(\tau) + R_{XY}(\tau) + R_{YX}(\tau) \end{aligned} \quad (19.2.30)$$

and if  $X(t)$  and  $Y(t)$  are orthogonal, then  $R_Z(\tau) = R_X(\tau) + R_Y(\tau)$ .

**11.** If  $\dot{X}(t)$  is the derivative of  $X(t)$ , then the cross-correlation between  $X(t)$  and  $\dot{X}(t)$  is

$$R_{X\dot{X}}(\tau) = \frac{dR_X(\tau)}{d\tau} \quad (19.2.31a)$$

This result can be shown from the formal definition of the derivative of  $X(t)$ . Substituting for  $\dot{X} = \lim_{\varepsilon \rightarrow 0} \{[X(t+\varepsilon) - X(t)]/\varepsilon\}$ , we obtain

$$\begin{aligned} R_{X\dot{X}}(\tau) &= E[X(t)X(t+\tau)] \\ &= \lim_{\varepsilon \rightarrow 0} E\left\{X(t)\left[\frac{X(t+\tau+\varepsilon) - X(t+\tau)}{\varepsilon}\right]\right\} \\ &= \lim_{\varepsilon \rightarrow 0} E\left\{\frac{X(t)X(t+\tau+\varepsilon) - X(t)X(t+\tau)}{\varepsilon}\right\} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{R_X(\tau+\varepsilon) - R_X(\tau)}{\varepsilon} = \frac{dR_X(\tau)}{d\tau} \end{aligned}$$

**12.** The autocorrelation of  $\dot{X}(t)$  is

$$R_{\dot{X}}(\tau) = -\frac{d^2R_X(\tau)}{d\tau^2} \quad (19.2.31b)$$

This result can be shown following a procedure similar to that described as in (11) above:

$$\begin{aligned}
R_{\dot{X}}(\tau) &= E[\dot{X}(t)\dot{X}(t+\tau)] \\
&= \lim_{\varepsilon \rightarrow 0} E\left[X(t) \frac{X(t+\tau+\varepsilon) - X(t+\tau)}{\varepsilon}\right] \\
&= \lim_{\delta \rightarrow 0} \lim_{\varepsilon \rightarrow 0} E\left[\left(\frac{X(t+\delta) - X(t)}{\delta}\right)\left(\frac{X(t+\tau+\varepsilon) - X(t+\tau)}{\varepsilon}\right)\right] \\
&= \lim_{\delta \rightarrow 0} \frac{1}{\delta} \lim_{\varepsilon \rightarrow 0} E\left[\frac{X(t+\delta) \frac{X(t+\tau+\varepsilon) - X(t+\tau)}{\varepsilon}}{-X(t) \frac{X(t+\tau+\varepsilon) - X(t+\tau)}{\varepsilon}}\right] \\
&= \lim_{\delta \rightarrow 0} \frac{1}{\delta} \lim_{\varepsilon \rightarrow 0} \left[ \frac{R_X(\tau-\delta+\varepsilon) - R_X(\tau-\delta)}{\varepsilon} - \frac{R_X(\tau+\varepsilon) - R_X(\tau)}{\varepsilon} \right] \\
&= \lim_{\delta \rightarrow 0} \frac{1}{\delta} \left[ \frac{dR_X(\tau-\delta)}{d\tau} - \frac{dR_X(\tau)}{d\tau} \right] = -\frac{d^2R_X(\tau)}{d\tau^2}
\end{aligned}$$

It can be shown that if a random process is wide-sense stationary, then it is necessary and sufficient that the following two conditions be satisfied:

1. The expected value is a constant,  $E[X(t)] = \mu_X$ .
2. The autocorrelation function  $R_X$  is a function of the time difference  $t_2 - t_1 = \tau$  and not individual times,  $R_X(t_1, t_2) = R_X(t_2 - t_1) = R_X(\tau)$ .

We will now give several examples to establish conditions for stationarity. The first few examples will be running examples that become progressively more difficult.

**Example 19.2.1** We will revisit Example 19.1.5 and find the conditions necessary for the random process  $X(t) = A \sin(\omega t + \Phi)$  to be stationary where  $A, \omega$  are constants and  $\Phi$  is a random variable:

1. *Mean:*

$$E[X(t)] = AE[\sin(\omega t + \Phi)] = A \int_{-a}^b \sin(\omega t + \phi) f_\Phi(\phi) d\phi$$

One of the ways the integral will be independent of  $t$  is for  $\Phi$  to be uniformly distributed in  $(0, 2\pi)$ , in which case we have

$$\frac{1}{2\pi} \int_0^{2\pi} \sin(\omega t + \phi) d\phi = \frac{-1}{2\pi} \cos(\omega t + \phi) \Big|_0^{2\pi} = 0$$

and  $E[X(t)] = 0$ .

**2. Autocorrelation:**

$$\begin{aligned}
E[X(t_1)X(t_2)] &= A^2 E[\sin(\omega t_1 + \Phi) \sin(\omega t_2 + \Phi)] \\
&= A^2 E[\cos[\omega(t_2 - t_1)] - \cos[\omega(t_2 + t_1) + 2\Phi]] \\
&= A^2 \cos[\omega(t_2 - t_1)] - A^2 E[\cos[\omega(t_2 + t_1) + 2\Phi]] \\
E\{\cos[\omega(t_2 + t_1) + 2\Phi]\} &= \frac{1}{2\pi} \int_0^{2\pi} \cos[\omega(t_2 + t_1) + 2\phi] d\phi \\
&= \frac{1}{2\pi} \frac{1}{2} \sin[\omega(t_2 + t_1) + 2\phi] \Big|_0^{2\pi} = 0
\end{aligned}$$

Hence,  $R_X(t_1, t_2) = A^2 \cos[\omega(t_2 - t_1)] = A^2 \cos[\omega\tau]$ , and  $X(t)$  is stationary if  $\Phi$  is uniformly distributed in  $(0, 2\pi)$ . Since  $R_X(\tau)$  is periodic,  $X(t)$  is a periodic stationary process.

**Example 19.2.2** We will modify Example 19.2.1 with both  $A$  and  $\Phi$  as random variables with density functions  $f_A(a)$  and  $f_\Phi(\phi)$ . We will now find the conditions under which  $X(t) = A \sin(\omega t + \Phi)$  will be stationary.

**1. Mean:**

$$E[X(t)] = E[A \sin(\omega t + \Phi)] = \iint a \sin(\omega t + \phi) f_{A\Phi}(a, \phi) d\phi da$$

The first condition for the double integral to be independent of  $t$  is for  $A$  and  $\Phi$  to be statistically independent, in which case we have

$$E[A \sin(\omega t + \Phi)] = \iint a \sin(\omega t + \phi) f_A(a) f_\Phi(\phi) d\phi da$$

and the second condition is for  $\Phi$  to be uniformly distributed in  $(0, 2\pi)$ , in which case we have  $(1/2\pi) \int_0^{2\pi} \sin(\omega t + \phi) d\phi = 0$  and  $E[X(t)] = 0$ .

**2. Autocorrelation:**

$$E[X(t_1)X(t_2)] = E[A^2 \sin(\omega t_1 + \Phi) \sin(\omega t_2 + \Phi)]$$

Since  $A$  and  $\Phi$  are independent, we have

$$\begin{aligned}
E[X(t_1)X(t_2)] &= E[A^2] E[\cos[\omega(t_2 - t_1)] - \cos[\omega(t_2 + t_1) + 2\Phi]] \\
&= E[A^2] \cos[\omega(t_2 - t_1)] - E[A^2] E[\cos[\omega(t_2 + t_1) + 2\Phi]]
\end{aligned}$$

and from the previous example  $E[\cos[\omega(t_2 + t_1) + 2\Phi]] = 0$ . Hence,  $R_X(t_1, t_2) = E[A^2] \cos[\omega(t_2 - t_1)] = E[A^2] \cos[\omega\tau]$  and  $X(t)$  is stationary if  $A$  and  $\Phi$  are independent and if  $\Phi$  is uniformly distributed in  $(0, 2\pi)$ . Since  $R_X(\tau)$  is periodic,  $X(t)$  a periodic stationary process.

**Example 19.2.3** We will now examine the conditions for stationarity for  $X(t) = A \sin(\Omega t + \Phi)$  when  $A$ ,  $\Omega$  and  $\Phi$  are all random variables with density functions  $f_A(a)$ ,  $f_\Omega(\omega)$ , and  $f_\Phi(\phi)$  respectively.

**1. Mean:**

$$E[X(t)] = E[A \sin(\Omega t + \Phi)] = \iiint a \sin(\omega t + \phi) f_{A\Omega\Phi}(a, \omega, \phi) d\phi d\omega da$$

This triple integral is a difficult one to evaluate. Hence we resort to conditional expectations by fixing the variable  $\Omega = \omega$ . Under this condition,  $E[X(t)] = \int E[X(t)|\Omega = \omega] f_\Omega(\omega) d\omega$  and from the previous example, if  $A$  and  $\Phi$  are independent and  $\Phi$  is uniformly distributed in  $(0, 2\pi)$ , we have  $E[X(t)|\Omega = \omega] = 0$ , and

$$\int E[X(t)|\Omega = \omega] f_\Omega(\omega) d\omega = \int E[A \sin(\omega t + \Phi)] f_\Omega(\omega) d\omega = 0$$

**2. Autocorrelation:**

$$\begin{aligned} R_X(t_1, t_2 | \Omega = \omega) &= E[X(t_1 | \Omega = \omega) X(t_2 | \Omega = \omega)] \\ &= E[A^2 \sin(\omega t_1 + \Phi) \sin(\omega t_2 + \Phi)] \end{aligned}$$

and from the previous example

$$E[A^2 \sin(\omega t_1 + \Phi) \sin(\omega t_2 + \Phi)] = \frac{1}{2} E[A^2] \cos[\omega \tau]$$

where  $\tau = t_2 - t_1$ . Hence

$$R_X(\tau) = \int R_X(\tau | \Omega = \omega) f_\Omega(\omega) d\omega = \int \frac{1}{2} E[A^2] \cos[\omega \tau] f_\Omega(\omega) d\omega$$

If  $\Omega$  is uniformly distributed in  $(0, \pi)$ , then

$$R_X(\tau) = \frac{1}{\pi} \int_0^\pi \frac{1}{2} E[A^2] \cos[\omega \tau] d\omega = \frac{E[A^2] \sin(\pi)}{2\pi \tau} = \frac{E[A^2]}{2} \text{Sa}(\pi)$$

**Example 19.2.4** In this example  $X(t) = A \cos(\omega t) + B \sin(\omega t)$ , where  $A$  and  $B$  are random variables with density functions  $f_A(a)$  and  $f_B(b)$ . We have to find the conditions under which  $X(t)$  will be stationary:

**1. Mean:**

$$E[X(t)] = E[A \cos(\omega t) + B \sin(\omega t)] = \cos(\omega t) E[A] + \sin(\omega t) E[B]$$

If  $E[X(t)]$  is to be independent of  $t$ , then  $E[A] = E[B] = 0$ , in which case  $E[X(t)] = 0$ .

## 2. Autocorrelation:

$$\begin{aligned}
R_X(t_1, t_2) &= E[X(t_1)X(t_2)] = E\{[A \cos(\omega t_1) + B \sin(\omega t_1)][A \cos(\omega t_2) + B \sin(\omega t_2)]\} \\
&= E\{A^2 \cos(\omega t_1) \cos(\omega t_2) + B^2 \sin(\omega t_1) \sin(\omega t_2) \\
&\quad + AB[\sin(\omega t_1) \cos(\omega t_2) + \cos(\omega t_1) \sin(\omega t_2)]\} \\
&= \frac{1}{2}E[A^2][\cos(\omega(t_2 - t_1)) + \cos(\omega(t_2 + t_1))] \\
&\quad + \frac{1}{2}E[B^2][\cos(\omega(t_2 - t_1)) - \cos(\omega(t_2 + t_1))] \\
&\quad + E[AB]\sin(\omega(t_2 + t_1))
\end{aligned}$$

The conditions under which this equation will be dependent only on  $(t_2 - t_1)$  are  $E[A^2] = E[B^2]$  and  $E[AB] = 0$ . In this case

$$R_X(t_1, t_2) = E[A^2][\cos(\omega(t_2 - t_1))] = E[A^2][\cos(\omega\tau)]$$

We can now summarize the conditions for stationarity:

- (a)  $E[A] = E[B] = 0$
- (b)  $E[A^2] = E[B^2]$
- (c)  $E[AB] = 0$

**Example 19.2.5** A die is tossed, and corresponding to the dots  $S = \{1, 2, 3, 4, 5, 6\}$ , a random process  $X(t)$  is formed with the following time functions as shown in Fig. 19.2.1:

$$\begin{aligned}
X(2; t) &= 3, X(4; t) = (2 - t), X(6; t) = (1 + t) \\
X(1; t) &= -3, X(3; t) = -(2 - t), X(5; t) = -(1 + t)
\end{aligned}$$

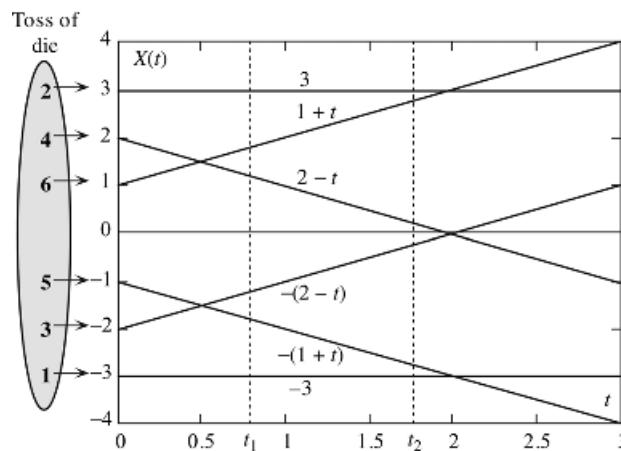


FIGURE 19.2.1

We have to find  $\mu_X(t)$ ,  $\sigma_X^2(t)$ ,  $R_X(t_1, t_2)$ ,  $C_X(t_1, t_2)$ , and  $\rho_X(t_1, t_2)$  and check whether  $X(t)$  is stationary:

*Mean:*

$$\mu_X(t) = \frac{1}{6} \sum_{i=1}^6 X_i(t) = 3 - 3 + (2 - t) - (2 - t) + (1 + t) - (1 + t) = 0$$

The mean value is a constant.

*Variance:*

$$\sigma_X^2(t) = \frac{1}{6} \sum_{i=1}^6 X_i^2(t) = \frac{1}{3} \cdot [3^2 + (2 - t)^2 + (1 + t)^2] = \frac{2}{3}[t^2 - t + 7]$$

*Autocorrelation:*

$$R_X(t_1, t_2) = \frac{1}{3}[9 + (1 + t_1)(1 + t_2) + (2 - t_1)(2 - t_2)] = \frac{1}{3}[14 - t_2 - t_1 + 2t_1 t_2]$$

*Autocovariance.* Since the mean value is zero,  $C_X(t_1, t_2) = R_X(t_1, t_2)$ .

*Normalized Autocovariance:*

$$\rho_X(t_1, t_2) = \frac{C_X(t_1, t_2)}{\sqrt{C_X(t_1)C_X(t_2)}} = \frac{[14 - t_2 - t_1 + 2t_1 t_2]}{\sqrt{(t_1^2 - t_1 + 7)(t_2^2 - t_2 + 7)}}$$

This process is not stationary.

**Example 19.2.6 (Random Binary Wave)** A sample function of a random binary wave  $X(t)$  consisting of independent rectangular pulses  $p(t)$ , each of which is of duration  $T$ , is shown in Fig. 19.2.2. The height  $H$  of the pulses is a random variable with constant amplitudes, which are equally likely to be  $\pm A$ . The time of occurrence of  $X(t)$  after  $t = 0$  is another random variable  $Z$ , which is uniformly distributed in  $(0, T)$ . We have to find the mean, variance, autocorrelation, autocovariance, and the normalized autocovariance of  $X(t)$ .

The random process  $X(t)$  is given by

$$X(t) = \sum_{k=-\infty}^{\infty} H p(t - kT - Z)$$

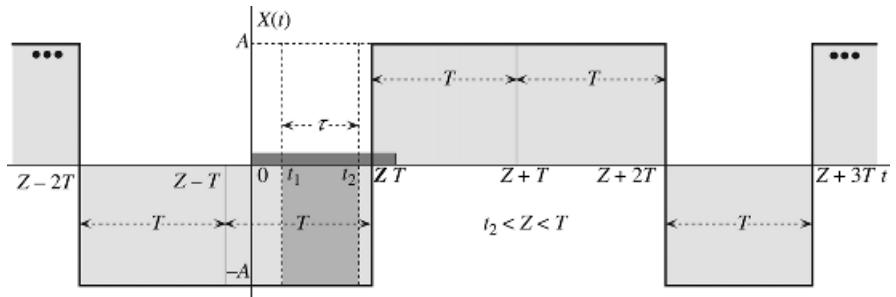


FIGURE 19.2.2

The probabilities of the random variable  $H$  are  $P(H = A) = \frac{1}{2}$  and  $P(H = -A) = \frac{1}{2}$ . Since  $Z$  is uniformly distributed in  $(0, T)$ , the following probabilities can be formulated for  $0 < t_1 < t_2 < T$ :

$$P(t_1 < Z < t_2) = \frac{t_2 - t_1}{T}, \quad P(0 < Z < t_1) = \frac{t_1}{T}, \quad P(t_2 < Z < T) = 1 - \frac{t_2}{T}$$

*Mean.* Since  $X(t)$  assumes only  $+A$  or  $-A$  with equal probability,

$$E[X(t)] = \left[ \frac{1}{2} \cdot A + \frac{1}{2} \cdot (-A) \right] = 0$$

*Variance.* The variance is given by

$$\text{var}[X(t)] = \sigma_X^2 = E[X^2(t)] = \left[ \frac{1}{2} \cdot A^2 + \frac{1}{2} \cdot (-A)^2 \right] = A^2$$

*Autocorrelation.* Determining  $R_X(t_1, t_2)$  is a bit more involved. The product  $X(t_1)X(t_2)$  in various intervals of time  $(t_1, t_2)$  can be found:

$$t_1 < Z < t_2 \leq T$$

Here  $t_1$  and  $t_2$  lie in adjacent pulse intervals as shown in Fig. 19.2.3. In Fig. 19.2.3a  $X(t_1) = -A$  and  $X(t_2) = +A$  and  $X(t_1)X(t_2) = -A^2$ . In Fig. 19.2.3b  $X(t_1) = +A$  and  $X(t_2) = +A$  and  $X(t_1)X(t_2) = A^2$ . The values  $-A^2$  and  $A^2$  will occur with equal probability:

$0 < Z < t_1 \leq T$ : In this case,  $t_1$  and  $t_2$  lie in the same pulse interval as shown in Fig. 19.2.3c. Here,  $X(t_1) = +A$  and  $X(t_2) = +A$  and  $X(t_1)X(t_2) = A^2$ .

$t_2 < Z < T$ : Here also  $t_1$  and  $t_2$  lie in the same pulse interval as shown in Fig. 19.2.3d. However,  $X(t_1) = -A$  and  $X(t_2) = -A$  and  $X(t_1)X(t_2) = A^2$ .

$(t_1 - t_2) > T$ : In this case  $t_1$  and  $t_2$  lie in different pulse intervals and  $X(t_1)X(t_2) = \pm A^2$ .

We can now find the autocorrelation function  $R_X(t_1, t_2) = E[X(t_1)X(t_2)]$ . For  $(t_1 - t_2) > T$ ,  $X(t_1)$  and  $X(t_2)$  are in different pulse intervals and invoking the independence of the pulses  $E[X(t_1)X(t_2)] = E[X(t_1)]E[X(t_2)] = 0$  since  $E[X(t)] = 0$ .

For  $(t_1 - t_2) \leq T$ , we have the following, using conditional expectations:

$$\begin{aligned} E[X(t_1)X(t_2)] &= E[X(t_1)X(t_2)|t_1 < Z < t_2]P(t_1 < Z < t_2) \\ &\quad + E[X(t_1)X(t_2)|0 < Z < t_1]P(0 < Z < t_1) \\ &\quad + E[X(t_1)X(t_2)|t_1 < Z < T]P(t_1 < Z < T) \end{aligned}$$

The conditional expectations can be determined using the probabilities and the products  $X(t_1)X(t_2)$  found earlier for the various intervals. Since  $A^2$  and  $-A^2$  occur with equal probability in the interval  $t_1 < Z < t_2 \leq T$ , the first conditional expectation term becomes

$$E[X(t_1)X(t_2)|t_1 < Z < t_2]P(t_1 < Z < t_2) = A^2 \cdot \frac{1}{2} \cdot \frac{t_2 - t_1}{T} - A^2 \cdot \frac{1}{2} \cdot \frac{t_2 - t_1}{T} = 0$$

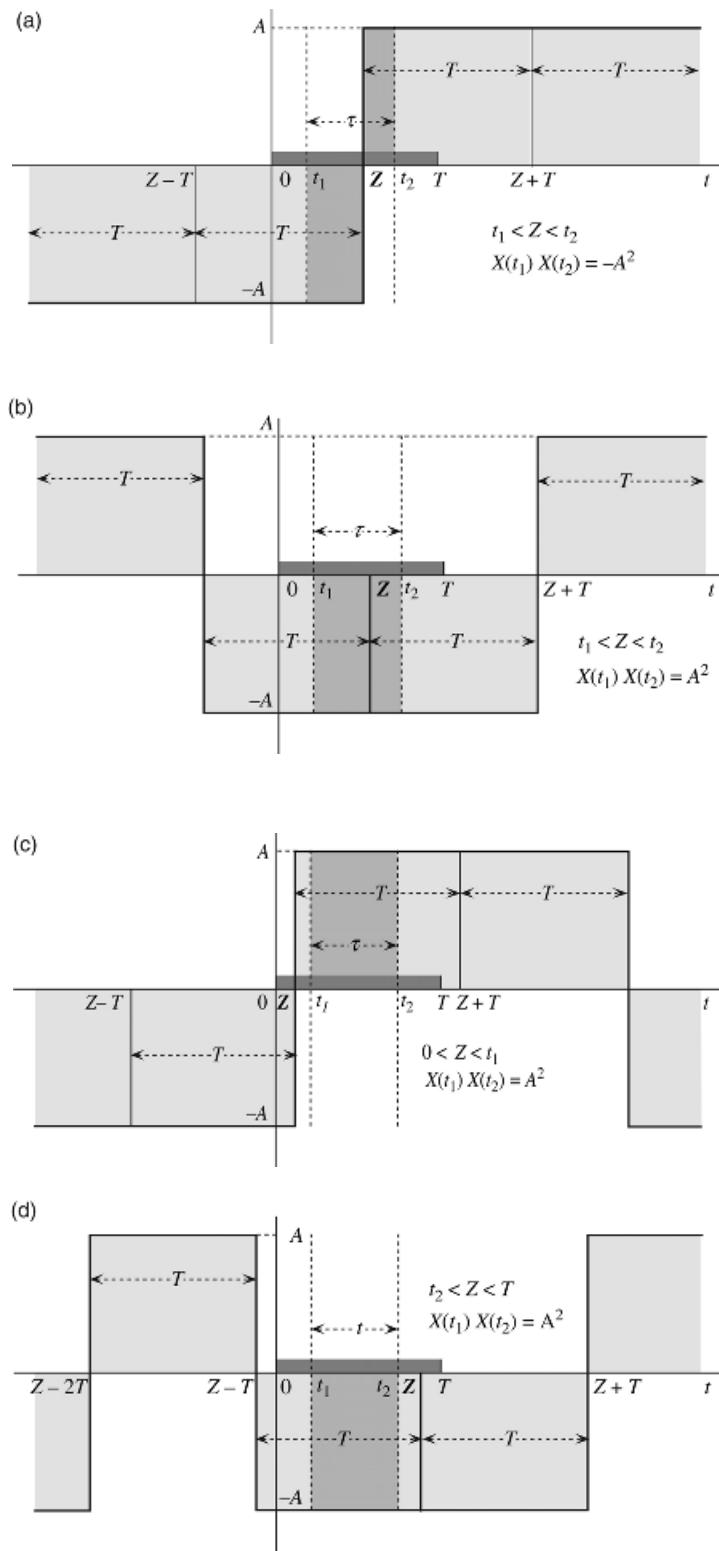


FIGURE 19.2.3

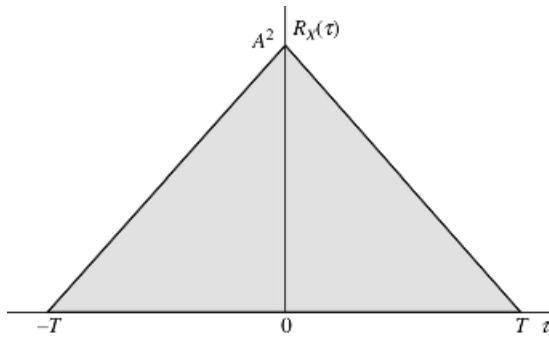


FIGURE 19.2.4

Hence

$$R_X(t_1, t_2) = \begin{cases} A^2 \left( \frac{t_1}{T} + 1 - \frac{t_2}{T} \right) = A^2 \left( 1 - \frac{t_2 - t_1}{T} \right), & (t_2 - t_1) \leq T \\ 0, & (t_2 - t_1) > T \end{cases}$$

This equation was derived under the assumption  $t_1 < t_2$ . For arbitrary values of  $t_1$  and  $t_2$  with  $\tau = (t_2 - t_1)$ , the autocorrelation function  $R_X(t_1, t_2)$  is given by

$$R_X(\tau) = \begin{cases} A^2 \left( 1 - \frac{|\tau|}{T} \right), & |\tau| < T \\ 0, & \text{otherwise} \end{cases}$$

The autocorrelation function for the process  $X(t)$  is shown in Fig. 19.2.4.

*Autocovariance.* Since the mean value is 0,  $C_X(\tau) = R_X(\tau)$ .

*Normalized Autocovariance.* The NACF is given by

$$\rho_X(\tau) = \frac{C_X(\tau)}{\sigma_X^2} = \begin{cases} 1 - \frac{|\tau|}{T}, & |\tau| < T \\ 0, & \text{otherwise} \end{cases}$$

**Example 19.2.7 (Random Telegraph Wave)** This example is a little different from the previous one. A sample function of a random telegraph wave is shown in Fig. 19.2.5. The wave assumes either of the values 1 or 0 at any instant of time.

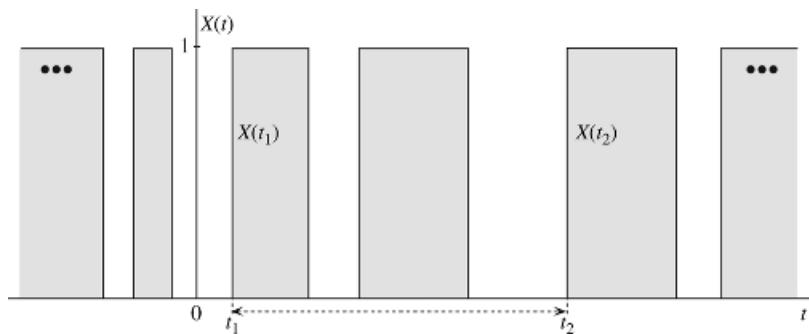


FIGURE 19.2.5

The probability of  $k$  changes from 0 to 1 in a time interval  $t$  is Poisson-distributed with probability mass function  $p(k; \lambda)$  given by

$$p(k; \lambda) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad t \geq 0$$

where  $\lambda$  is the average number of changes per unit time. We have to find the mean, variance, autocorrelation, autocovariance, and normalized autocovariance.

*Mean.* Since  $X(t)$  assumes only 1 or 0 with equal probability, we obtain

$$E[X(t)] = \mu_X = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = \frac{1}{2}$$

*Variance.* The variance is given by

$$\text{var}[X(t)] = \sigma_X^2 = E[X^2(t)] - \mu_X^2 = \left[ \frac{1}{2} \cdot 1^2 + \frac{1}{2} \cdot 0^2 \right] - \frac{1}{4} = \frac{1}{4}$$

*Autocorrelation.* The autocorrelation function can be given in terms of the joint density functions with  $t_2 > t_1$ :

$$\begin{aligned} R_X(t_1, t_2) &= E[X(t_1) = 0 \cdot X(t_2) = 0] + E[X(t_1) = 0 \cdot X(t_2) = 1] \\ &\quad + E[X(t_1) = 1 \cdot X(t_2) = 0] + E[X(t_1) = 1 \cdot X(t_2) = 1] \\ &= (0 \cdot 0)P[X(t_1) = 0 \cdot X(t_2) = 0] + (0.1)P[X(t_1) = 0 \cdot X(t_2) = 1] \\ &\quad + (1 \cdot 0)P[X(t_1) = 1 \cdot X(t_2) = 0] + (1.1)P[X(t_1) = 1 \cdot X(t_2) = 1] \end{aligned}$$

or

$$R_X(t_1, t_2) = P[X(t_1) = 1 \cdot X(t_2) = 1]$$

Expressing the joint probability in terms of conditional probabilities, we have

$$R_X(t_1, t_2) = P[X(t_2) = 1 | X(t_1) = 1]P[X(t_1) = 1]$$

The conditional probability in this equation is the probability of even number of changes and using the Poisson distribution:

$$\begin{aligned} R_X(t_1, t_2) &= \frac{1}{2} \sum_{\substack{k=0 \\ k \text{ even}}}^{\infty} \frac{[\lambda(t_2 - t_1)]^k}{k!} e^{-\lambda(t_2 - t_1)} \\ &= \frac{e^{-\lambda(t_2 - t_1)}}{2} \left\{ \frac{1}{2} \sum_{k=0}^{\infty} \frac{[\lambda(t_2 - t_1)]^k}{k!} + \sum_{k=0}^{\infty} \frac{-[\lambda(t_2 - t_1)]^k}{k!} \right\} \\ &= \frac{e^{-\lambda(t_2 - t_1)}}{2} \left\{ \frac{e^{\lambda(t_2 - t_1)} + e^{-\lambda(t_2 - t_1)}}{2} \right\} \\ &= \frac{1}{4} \{1 + e^{-2\lambda(t_2 - t_1)}\}, \quad t_2 > t_1 \end{aligned}$$

A similar equation holds good for  $t_2 < t_1$ . Hence, substituting  $|t_2 - t_1| = |\tau|$  in the equation above,  $R_X(\tau)$  can be given by

$$R_X(\tau) = \frac{1}{4} \{1 + e^{-2\lambda|\tau|}\}$$

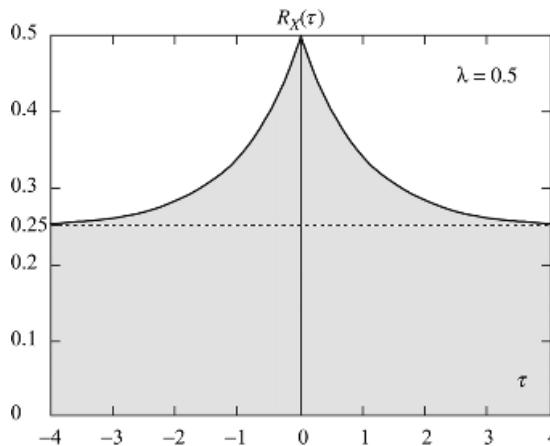


FIGURE 19.2.6

The autocorrelation function for the process  $X(t)$  is shown in Fig. 19.2.6 for  $\lambda = 0.5$ .

*Autocovariance:*

$$C_X(\tau) = R_X(\tau) - \mu_X^2 = \frac{e^{-2\lambda|\tau|}}{4}$$

*Normalized Autocovariance:*

$$\rho_X(\tau) = \frac{C_X(\tau)}{\sigma_X^2} = e^{-2\lambda|\tau|}$$

**Example 19.2.8 (Modulation)** A random process  $Y(t)$  is given by  $Y(t) = X(t) \cos(\omega t + \Phi)$ , where  $X(t)$ , a zero mean wide-sense stationary random process with autocorrelation function  $R_X(\tau) = 2e^{-2\lambda|\tau|}$  is modulating the carrier  $\cos(\omega t + \Phi)$ . The random variable  $\Phi$  is uniformly distributed in the interval  $(0, 2\pi)$ , and is independent of  $X(t)$ . We have to find the mean, variance, and autocorrelation of  $Y(t)$ :

*Mean.* The independence of  $X(t)$  and  $\Phi$  allows us to write

$$E[Y(t)] = E[X(t)]E[\cos(\omega t + \Phi)]$$

and with  $E[X(t)] = 0$  and  $E[\cos(\omega t + \Phi)] = 0$  from Example 19.2.1,  $E[Y(t)] = 0$ .

*Variance.* Since  $X(t)$  and  $\Phi$  are independent, the variance can be given by

$$\sigma_Y^2 = E[Y^2(t)] = E[X^2(t) \cos^2(\omega t + \Phi)] = \sigma_X^2 E[\cos^2(\omega t + \Phi)]$$

However

$$E[\cos^2(\omega t + \Phi)] = \frac{1}{2}E[1 + \cos(2\omega t + 2\Phi)] = \frac{1}{2} \quad \text{and} \quad \sigma_X^2 = C_X(0) = R_X(0) = 2$$

and hence  $\sigma_Y^2 = \sigma_X^2/2 = 1$ .

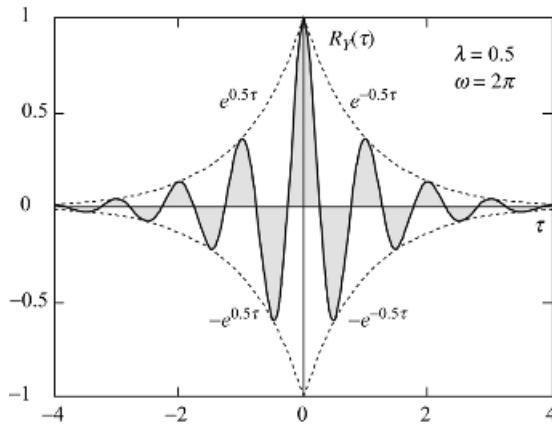


FIGURE 19.2.7

Autocorrelation:

$$\begin{aligned}
 R_Y(\tau) &= E[Y(t)Y(t+\tau)] = E[X(t)\cos(\omega t + \Phi)X(t+\tau)\cos(\omega t + \omega\tau + \Phi)] \\
 &= R_X(\tau) \frac{1}{2} E[\cos(\omega\tau) + \cos(2\omega t + \omega\tau + 2\Phi)] \\
 &= \frac{R_X(\tau)}{2} \cos(\omega\tau) + \frac{R_X(\tau)}{2} E[\cos(2\omega t + \omega\tau + 2\Phi)]
 \end{aligned}$$

From Example 19.2.1  $E[\cos(2\omega t + \omega\tau + 2\Phi)] = 0$ , and hence

$$R_Y(\tau) = \frac{R_X(\tau)}{2} \cos(\omega\tau) = e^{-2\lambda|\tau|} \cos(\omega\tau)$$

A graph of  $R_Y(\tau)$  is shown in Fig. 19.2.7 with  $\lambda = 0.5$  and  $\omega = 2\pi$ .

**Example 19.2.9 (Cross-Correlation)** Two random processes  $X(t)$  and  $Y(t)$  are given by  $X(t) = A \cos(\omega t) + B \sin(\omega t)$  and  $Y(t) = -A \sin(\omega t) + B \cos(\omega t)$ , where  $A$  and  $B$  are random variables with density functions  $f_A(a)$  and  $f_B(b)$ . We have to find the cross-correlation, cross-covariance, and the normalized cross-covariance between  $X(t)$  and  $Y(t)$ .

From Example 19.2.4 the processes  $X(t)$  and  $Y(t)$  are stationary if  $E[A] = E[B] = 0$ ,  $E[A^2] = E[B^2]$ , and  $E[AB] = 0$ . Hence the mean values  $\mu_X = \mu_Y = 0$ . The variances of these processes are  $E[A^2] = E[B^2] = \sigma^2$ .

*Cross-Correlation.* The cross-correlation function  $R_{XY}(t_1, t_2)$  can be written as

$$\begin{aligned}
 R_{XY}(t_1, t_2) &= E[X(t_1)Y(t_2)] \\
 &= E\{[A \cos(\omega t_1) + B \sin(\omega t_1)][-A \sin(\omega t_2) + B \cos(\omega t_2)]\} \\
 &= E\{-A^2 \cos(\omega t_1) \sin(\omega t_2) + B^2 \sin(\omega t_1) \cos(\omega t_2) \\
 &\quad + AB[\cos(\omega t_1) \cos(\omega t_2) - \sin(\omega t_1) \sin(\omega t_2)]\}
 \end{aligned}$$

Substituting the values  $E[A^2] = E[B^2] = \sigma^2$  and  $E[AB] = 0$  in this equation, we obtain

$$R_{XY}(t_1, t_2) = \sigma^2 \sin[\omega(t_1 - t_2)] = -\sigma^2 \sin[\omega\tau]$$

where we have substituted  $\tau = (t_2 - t_1)$ . The term  $R_{XY}(\tau)$  is shown in Fig. 19.2.8 for  $\sigma^2 = 4$  and  $\omega = 2\pi$ .

Here  $R_{XY}(\tau)$  is an odd function, unlike the autocorrelation function. Since  $R_{XY}(\tau) = 0$  for  $\tau = 0$ , we conclude that  $X(t)$  and  $Y(t)$  are orthogonal.

*Cross-Covariance.* Since  $\mu_X = \mu_Y = 0$ ,  $C_{XY}(\tau) = R_{XY}(\tau)$ .

*Normalized Cross-Covariance:*

$$\rho_{XY}(\tau) = \frac{C_{XY}(\tau)}{\sigma^2} = \frac{-\sigma^2 \sin[\omega\tau]}{\sigma^2} = -\sin[\omega\tau]$$

**Example 19.2.10** A random telegraph wave  $X(t)$  as in Example 19.2.7 is passed through a linear system with impulse response  $h(t) = e^{-\beta t}u(t)$ , where  $u(t)$  is a unit step function. The output of the system is  $Y(t)$ . It is desired to find the cross-correlation function  $R_{XY}(\tau)$ .

From Example 19.2.7 the autocorrelation function  $R_X(\tau) = \frac{1}{4}\{1 + e^{-2\lambda|\tau|}\}$ . Hence, from Eq. (19.2.23), we have

$$\begin{aligned} R_{XY}(\tau) &= \int_0^\infty \frac{1}{4}(1 + e^{-2\lambda|\tau-\alpha|})e^{-\beta\alpha} d\alpha \\ &= \begin{cases} \frac{1}{4\beta} + \frac{1}{4} \int_0^\tau e^{-2\lambda(\tau-\alpha)} e^{-\beta\alpha} d\alpha + \frac{1}{4} \int_\tau^\infty (e^{-2\lambda(\alpha-\tau)}) e^{-\beta\alpha} d\alpha & \tau > 0 \\ \frac{1}{4\beta} + \frac{1}{4} \int_0^\infty e^{-2\lambda(\alpha-\tau)} e^{-\beta\alpha} d\alpha & \tau \leq 0 \end{cases} \end{aligned}$$

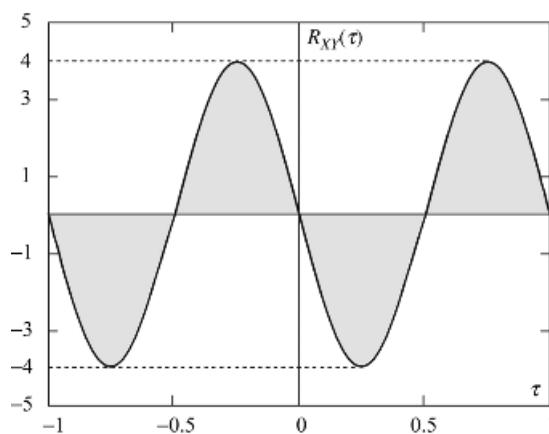


FIGURE 19.2.8

Evaluating the integrals and substituting  $\lambda = 1$  and  $\beta = 0.5$ , we obtain

$$R_{XY}(\tau) = \frac{1}{4\beta} + \begin{cases} \frac{\lambda e^{-\beta\tau}}{4\lambda^2 - \beta^2} - \frac{e^{-2\lambda\tau}}{4(2\lambda - \beta)}, & \tau > 0 \\ \frac{e^{2\lambda\tau}}{4(2\lambda + \beta)}, & \tau \leq 0 \end{cases}$$

or

$$R_{XY}(\tau) = \frac{1}{2} + \begin{cases} \frac{4}{15}e^{-(\tau/2)} - \frac{1}{6}e^{-2\tau}, & \tau > 0 \\ \frac{1}{10}e^{2\tau}, & \tau \leq 0 \end{cases}$$

The cross-correlation function  $R_{XY}(\tau)$  is shown in Fig. 19.2.9.

$R_{XY}(\tau)$  does not possess any symmetry, unlike  $R_X(\tau)$ .

### Moments of Discrete-Time Stationary Processes

In actual practice observations are made on the sample function of a stationary random process  $X(t)$  at equally spaced time intervals  $\{t_i, i = 0, \pm 1, \dots\}$  with corresponding sequence of random variables  $\{X_i, i = 0, \pm 1, \dots\}$ . These observation random variables will *not be independent*. Since  $\{X_i\}$  are samples of a stationary random process, the means and variances of these samples are the same as in the original process:

$$E[X_i] = \mu_X; \quad \text{var}[X_i] = \sigma_X^2, \quad i = 0, \pm 1, \dots \quad (19.2.32)$$

Analogous to the continuous case, we can define the various second moments:

*Autocovariance:*

$$C_X(h) = E[(X_i - \mu_X)(X_{i+h} - \mu_X)], \quad i = 0, \pm 1, \dots \quad (19.2.33)$$

*Autocorrelation:*

$$R_X(h) = E[X_i X_{i+h}], \quad i = 0, \pm 1, \dots \quad (19.2.34)$$

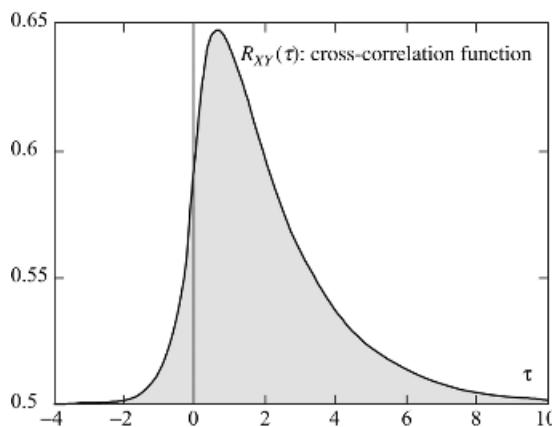


FIGURE 19.2.9

Normalized Autocovariance (NACF):

$$\rho_X(h) = \frac{C_X(h)}{C_X(0)} = \frac{C_X(h)}{\sigma_X^2} \quad (19.2.35)$$

If  $\{Y_i, i = 0, \pm 1, \dots\}$  is the sequence obtained from a second stationary random process  $Y(t)$  with mean  $\mu_Y$  and variance  $\sigma_Y^2$ , then we can define the cross-moments as follows:

Cross-Covariance:

$$C_{XY}(h) = E[(X_i - \mu_X)(Y_{i+h} - \mu_Y)] \quad i = 0, \pm 1, \dots \quad (19.2.36)$$

Cross-Correlation:

$$R_{XY}(h) = E[X_i Y_{i+h}] \quad i = 0, \pm 1, \dots \quad (19.2.37)$$

Normalized Cross-Covariance (NCCF):

$$\rho_{XY}(h) = \frac{C_{XY}(h)}{\sqrt{C_X(0)C_Y(0)}} = \frac{C_{XY}(h)}{\sigma_X\sigma_Y} \quad (19.2.38)$$

**Example 19.2.11** A discrete zero mean stationary random process  $X_i$  is given by

$$X_i = \phi X_{i-1} + v_i, \quad i = 1, \dots \quad (19.2.39)$$

where  $v_i$  is a zero mean Gaussian random process with variance  $\sigma_v^2$ . We want to find the variance of  $X_i$  and the NACF  $\rho_X(h)$ .

Multiplying both sides of Eq. (19.2.39) by  $X_i$  and taking expectations, we have

$$E[X_i^2] = \phi E[X_i X_{i-1}] + E[X_i v_i] \quad \text{or} \quad \sigma_X^2 = \phi C_X(1) + E[X_i v_i] \quad (19.2.40)$$

The cross-correlation  $E[X_i v_i]$  can be computed as follows:

$$E[X_i v_i] = E[(\phi X_{i-1} + v_i)v_i] = \sigma_v^2 \quad (19.2.41)$$

since  $v_i$  occurs after  $X_{i-1}$ . Hence, substituting Eq. (19.2.41) in Eq. (19.2.40) and dividing throughout by  $\sigma_X^2$ , we obtain

$$1 = \phi \rho_X(1) + \frac{\sigma_v^2}{\sigma_X^2} \quad \text{and} \quad \sigma_X^2 = \frac{\sigma_v^2}{1 - \phi \rho_X(1)} \quad (19.2.42)$$

Premultiplying both sides of Eq. (19.2.39) by  $X_{i-h}$  and taking expectations, we have

$$E[X_{i-h} X_i] = E[\phi X_{i-h} X_{i-1}] + E[X_{i-h} v_i] \quad (19.2.43)$$

In Eq. (19.2.43)  $E[X_{i-h} v_i] = 0$  since  $v_i$  occurs after  $X_{i-h}$  for  $h > 0$ ; hence

$$C_X(h) = \phi C_X(h-1), \quad h > 0 \quad (19.2.44)$$

Dividing Eq. (19.2.44) by  $C_X(0) = \sigma_X^2$ , we obtain an equation for the NACF  $\rho_X(h)$

$$\rho_X(h) = \phi \rho_X(h-1), \quad h > 0 \quad (19.2.45)$$

and solving for  $\rho_X(h)$  with initial condition  $\rho_X(0) = 1$ , we obtain

$$\rho_X(h) = \phi^h, \quad h > 0 \quad (19.2.46)$$

Substituting Eq. (19.2.46) in Eq. (19.2.42), we have

$$\sigma_X^2 = \frac{\sigma_v^2}{1 - \phi^2} \quad (19.2.47)$$

The process defined by Eq. (19.2.39) is called *an autoregressive process of order 1*.

## 19.3 ERGODIC PROCESSES

The ensemble average of a random process  $X(t)$  is the mean value  $\mu_X(t)$  defined by,

$$\mu_X(t) = \int_{-\infty}^{\infty} xf_X(x; t)dx \quad (19.3.1)$$

Finding the ensemble average  $\mu_X(t)$  requires storing a multiplicity of sample functions and finding the average. In many instances this process may be nontrivial. Given a sample function of any random process  $X(t)$ , the *time average*  $\hat{\mu}_X$  is defined by

$$\hat{\mu}_X = \frac{1}{2T} \int_{-T}^T X(t)dt \quad (19.3.2)$$

A reasonable question to ask is whether the ensemble average can be obtained from a much easier time average. We observe that the ensemble average is not a random variable but is a function of time, whereas the time average is a random variable that is not a function of time. If these averages are to be equal, then the first condition that we have to impose is that the random process  $X(t)$  be stationary, which removes the time factor in the mean value. Under these conditions, we can write

$$E\left[\frac{1}{2T} \int_{-T}^T X(t)dt\right] = E[\hat{\mu}_X] = \frac{1}{2T} \int_{-T}^T E[X(t)]dt = \mu_X \quad (19.3.3)$$

and  $\hat{\mu}_X$  is an unbiased estimator of  $\mu_X$ .

If a random variable is to be equal to a constant, then the second condition from Example 14.1.1 is that the variance of  $\hat{\mu}_X$  must tend to zero as  $T \rightarrow \infty$ . Such random processes are said to satisfy the *ergodic hypothesis*. We will now derive the conditions for a random process to be *mean-ergodic* and *correlation-ergodic*.

### Mean-Ergodic

A stationary random process  $X(t)$  is called *mean-ergodic* if the ensemble average is equal to the time average of the sample function  $x(t)$ . We will assume that the following conditions are satisfied by  $X(t)$ :

$X(t)$  is stationary, implying that it has a constant mean  $\mu_X$ , and the autocorrelation function  $R_X(t, t + \tau)$  is a function of  $\tau$  only, or  $R_X(t, t + \tau) = R_X(\tau)$ .

$R_X(0) = E[X^2(t)]$  is bounded, or  $R_X(0) < \infty$ . Hence  $C_X(0) = R_X(0) - \mu_X^2 < \infty$ .

We have to find the conditions under which the variance  $\sigma_{\hat{\mu}_X}^2$  of the time average  $\hat{\mu}_X$  goes to 0. The variance  $\sigma_{\hat{\mu}_X}^2$  is given by

$$\begin{aligned}
 \sigma_{\hat{\mu}_X}^2 &= E[(\hat{\mu}_X - \mu_X)^2] = E\left\{\left[\frac{1}{2T} \int_{-T}^T [X(t) - \mu_X] dt\right]^2\right\} \\
 &= E\left\{\left(\frac{1}{2T}\right)^2 \int_{-T}^T \int_{-T}^T [X(t) - \mu_X][X(s) - \mu_X] dt ds\right\} \\
 &= \left(\frac{1}{2T}\right)^2 \int_{-T}^T \int_{-T}^T C_X(t,s) dt ds \\
 &= \left(\frac{1}{2T}\right)^2 \int_{-T}^T \int_{-T}^T C_X(s,t) dt ds, \quad X(t) \text{ is stationary} \quad (19.3.4)
 \end{aligned}$$

Substitution of  $\tau = s - t$  and  $u = t$  in Eq. (19.3.4), transforms the coordinate system from  $[s,t]$  to  $[\tau,u]$ . The Jacobian  $\|J\|$  of the transformation is given by

$$\|J\| = \left| \begin{bmatrix} \frac{\partial \tau}{\partial t} & \frac{\partial \tau}{\partial s} \\ \frac{\partial s}{\partial t} & \frac{\partial s}{\partial s} \end{bmatrix} \right| = \left| \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \right| = 1 \quad (19.3.5)$$

Thus

$$\sigma_{\hat{\mu}_X}^2 = \left(\frac{1}{2T}\right)^2 \iint C_X(\tau) d\tau du \quad (19.3.6)$$

The integration has to be carried out over the parallelogram shown in Fig. 19.3.1 in regions I and II.

*Region I:*  $\tau \leq 0$ : Limits for  $\tau$  are from  $-2T$  to 0; limits for  $u$  are from  $-T$  to  $T + \tau$ .

*Region II:*  $\tau > 0$ : Limits for  $\tau$  are from 0 to  $2T$ ; limits for  $u$  are from  $-T + \tau$  to  $T$ .

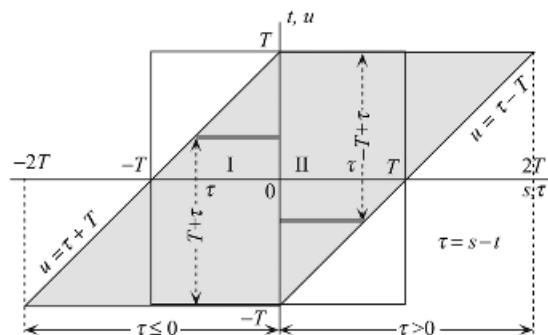


FIGURE 19.3.1

Substituting these limits in Eq. (19.3.6), we obtain the variance  $\sigma_{\hat{\mu}_X}^2$ :

$$\begin{aligned}\sigma_{\hat{\mu}_X}^2 &= \left(\frac{1}{2T}\right)^2 \left\{ \int_{-T+\tau}^T \left[ \int_0^{2T} C_X(\tau) d\tau \right] du + \int_{-T}^{T+\tau} \left[ \int_{-2T}^0 C_X(\tau) d\tau \right] du \right\} \\ &= \frac{1}{4T^2} \left\{ \int_0^{2T} C_X(\tau) [2T - \tau] d\tau + \int_{-2T}^0 C_X(\tau) [2T + \tau] d\tau \right\} \\ &= \frac{1}{4T^2} \left\{ \int_{-2T}^{2T} C_X(\tau) [2T - |\tau|] d\tau \right\} = \frac{1}{2T} \left\{ \int_{-2T}^{2T} C_X(\tau) \left[ 1 - \frac{|\tau|}{2T} \right] d\tau \right\} \quad (19.3.7)\end{aligned}$$

Finally, if the random process is to be mean-ergodic, the following condition has to be satisfied:

$$\lim_{T \rightarrow \infty} \sigma_{\hat{\mu}_X}^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \left\{ \int_{-2T}^{2T} C_X(\tau) \left[ 1 - \frac{|\tau|}{2T} \right] d\tau \right\} \rightarrow 0 \quad (19.3.8)$$

Since

$$\left| \int_{-2T}^{2T} C_X(\tau) \left[ 1 - \frac{|\tau|}{2T} \right] d\tau \right| \leq \int_{-2T}^{2T} |C_X(\tau)| \left| \left[ 1 - \frac{|\tau|}{2T} \right] \right| d\tau \leq \int_{-2T}^{2T} |C_X(\tau)| d\tau \quad (19.3.9)$$

hence, if  $\int_{-\infty}^{\infty} |C_X(\tau)| d\tau$  is bounded, then  $\lim_{T \rightarrow \infty} \sigma_{\hat{\mu}_X}^2$  does indeed go to zero. A necessary and sufficient condition for a random process to be mean-ergodic is

$$\int_{-\infty}^{\infty} |C_X(\tau)| d\tau < \infty \quad (19.3.10)$$

Even though we have derived necessary and sufficient condition for a mean-ergodic process, we cannot use it to test the ergodicity of any process since it involves prior knowledge of the autocovariance function  $C_X(\tau)$ . However, if a partial knowledge of the ACF such as  $|C_X(\tau)|$  goes to 0 as  $\tau \rightarrow \infty$ , then we can conclude that the process is ergodic. Although ergodic processes must be stationary, stationary processes need not be ergodic. Since it is rather difficult to show that any arbitrary random process is ergodic, we will assume, unless otherwise stated, that a process that is stationary is also ergodic.

### Correlation-Ergodic

The time autocorrelation function of a stationary random process  $X(t)$  is defined by

$$\hat{R}_X(\lambda) = \frac{1}{2T} \int_{-T}^T X(t) X(t + \lambda) dt \quad (19.3.11)$$

A stationary random process  $X(t)$  is correlation-ergodic if

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \hat{R}_X(\lambda) dt = \lim_{T \rightarrow \infty} \left[ \frac{1}{2T} \int_{-T}^T X(t) X(t + \lambda) dt \right] = R_X(\lambda) \quad (19.3.12)$$

for all  $\lambda$ . Clearly  $\hat{R}_X(\lambda)$  is a random variable, but

$$E[\hat{R}_X(\lambda)] = \frac{1}{2T} E \left[ \int_{-T}^T X(t) X(t + \lambda) dt \right] = \frac{1}{2T} \left[ \int_{-T}^T R_X(\lambda) dt \right] = R_X(\lambda)$$

and  $\hat{R}_X(\lambda)$  is an unbiased estimator of  $R_X(\lambda)$ . As in the case of mean-ergodic processes, here also the variance of  $\hat{R}_X(\lambda)$  must tend to 0 as  $T \rightarrow \infty$ . Hence we can form the

variance  $\sigma_{\hat{R}_X^\lambda}^2$  of  $\hat{R}_X(\lambda)$  for any fixed  $\lambda$ :

$$\begin{aligned}
\sigma_{\hat{R}_X^\lambda}^2 &= E\left\{\left[\hat{R}_X(\lambda) - R_X(\lambda)\right]^2\right\} = E\left\{\left[\frac{1}{2T} \int_{-T}^T [X(t)X(t+\lambda) - R_X(\lambda)]dt\right]^2\right\} \\
&= \left(\frac{1}{2T}\right)^2 \int_{-T}^T \int_{-T}^T E\{[X(t)X(t+\lambda) - R_X(\lambda)][X(s)X(s+\lambda) - R_X(\lambda)]\}dt ds \\
&= \left(\frac{1}{2T}\right)^2 \int_{-T}^T \int_{-T}^T C_{R_X}^\lambda(t,s)dt ds \\
&= \left(\frac{1}{2T}\right)^2 \int_{-T}^T \int_{-T}^T C_{R_X}^\lambda(s-t)dt ds
\end{aligned} \tag{19.3.13}$$

Substituting  $(s-t) = \tau$  in Eq. (19.3.13), we obtain a result similar to Eq. (19.3.8), namely

$$\lim_{T \rightarrow \infty} \sigma_{\hat{R}_X^\lambda}^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \left\{ \int_{-2T}^{2T} C_{R_X}^\lambda(\tau) \left[ 1 - \frac{|\tau|}{2T} \right] d\tau \right\} \rightarrow 0 \tag{19.3.14}$$

and the necessary and sufficient condition for  $X(t)$  to be correlation-ergodic is

$$\int_{-\infty}^{\infty} |C_{R_X}^\lambda(\tau)| d\tau < \infty \tag{19.3.15}$$

where

$$\begin{aligned}
C_{R_X}^\lambda(\tau) &= R_{R_X}^\lambda(\tau) - R_X^2(\lambda) \\
R_{R_X}^\lambda &= E[X(t)X(t+\lambda)X(t+\tau)X(t+\tau+\lambda)]
\end{aligned}$$

Thus the autocorrelation ergodicity involves the fourth-order stationary moments of  $X(t)$ . As a corollary to this result, we can define *mean-square* or *power ergodicity* as

$$\lim_{T \rightarrow \infty} \left[ \frac{1}{2T} \int_{-T}^T X^2(t) dt \right] = R_X(0) \tag{19.3.16}$$

and the necessary and sufficient conditions for mean-square ergodicity are

$$\lim_{T \rightarrow \infty} \sigma_{\hat{R}_X^0}^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \left\{ \int_{-2T}^{2T} C_{R_X}^0(\tau) \left[ 1 - \frac{|\tau|}{2T} \right] d\tau \right\} \rightarrow 0 \tag{19.3.17}$$

or

$$\int_{-\infty}^{\infty} |C_{R_X}^0(\tau)| d\tau < \infty \tag{19.3.18}$$

where

$$\begin{aligned}
C_{R_X}^0(\tau) &= R_{R_X}^0(\tau) - R_X^2(0) \\
R_{R_X}^0(\tau) &= E\{[X(t)X(t+\tau)]^2\}
\end{aligned}$$

**Example 19.3.1** We will revisit Example 19.2.1, where  $X(t) = A \sin(\omega_0 t + \Phi)$ , where  $A$  and  $\Phi$  are independent and  $A$  is uniformly distributed over  $(0,1)$  and  $\Phi$  is uniformly distributed over  $(0,2\pi)$  and find whether it is mean- and power-ergodic.

We have already seen in Example 19.2.1 that

$$E[X(t)] = E[A]E[\sin(\omega_0 t + \Phi)] = 0$$

$$\begin{aligned} R_X(\tau) &= E[X(t)X(t + \tau)] = E[A^2]E[\sin(\omega_0 t + \Phi)\sin(\omega_0 t + \omega_0 \tau + \Phi)] \\ &= \frac{1}{2}E[A^2]\cos(\omega_0 \tau) \end{aligned}$$

and  $X(t)$  is indeed stationary. In fact, it is also strict-sense stationary. We will check first whether it is mean-ergodic. Since the mean value is zero,  $C_X(\tau) = \frac{1}{2}E[A^2]\cos(\omega_0 \tau)$ . The variance of the time average  $\sigma_{\hat{\mu}_X}^2$  is given by

$$\sigma_{\hat{\mu}_X}^2 = \frac{E[A^2]}{2T} \left\{ \int_{-2T}^{2T} \left[ 1 - \frac{|\tau|}{2T} \right] \frac{\cos(\omega_0 \tau)}{2} d\tau \right\}$$

We can use the result from Fourier transform that time multiplication is frequency convolution and evaluate the integral shown above. Substituting the Fourier transforms of the triangular and the cosine functions

$$\left[ 1 - \frac{|\tau|}{2T} \right] \iff 2T \left[ \frac{\sin(\omega T)}{\omega T} \right]^2 \quad \text{and} \quad \frac{\cos(\omega_0 \tau)}{2} \iff \frac{\pi}{2} [\delta(\omega + \omega_0) + \delta(\omega - \omega_0)]$$

in the equation for the variance, we obtain

$$\begin{aligned} \sigma_{\hat{\mu}_X}^2 &= \frac{1}{2\pi} \frac{E[A^2]}{2T} \int_{-\infty}^{\infty} 2T \left[ \frac{\sin(\omega T)}{\omega T} \right]^2 \frac{\pi}{2} [\delta(\omega + \omega_0) + \delta(\omega - \omega_0)] d\omega \\ &= \frac{E[A^2]}{2\pi} \left[ \frac{\sin(\omega_0 T)}{\omega_0 T} \right]^2 \end{aligned}$$

and as  $T \rightarrow \infty$ ,  $\sigma_{\hat{\mu}_X}^2 \rightarrow 0$ , for  $\omega_0 \neq 0$ . Thus  $X(t)$  is mean-ergodic.

Substituting  $E[A^2] = \frac{1}{3}$  in the variance equation, we obtain

$$\sigma_{\hat{\mu}_X}^2 = \frac{1}{6\pi} \left[ \frac{\sin(\omega_0 T)}{\omega_0 T} \right]^2$$

To check for power ergodicity, it is simpler to use Eq. (19.3.16) rather than Eq. (19.3.17). Thus

$$\lim_{T \rightarrow \infty} \left[ \frac{1}{2T} \int_{-T}^T X^2(t) dt \right] = \lim_{T \rightarrow \infty} \left[ \frac{1}{2T} \int_{-T}^T A^2 \cos^2(\omega_0 t + \Phi) dt \right] = \frac{A^2}{2} \quad \text{for } \omega_0 \neq 0$$

Since this expression does not converge to  $R_X(0) = \frac{1}{6}$ , the process  $X(t)$  is neither power-ergodic nor correlation-ergodic.

**Example 19.3.2** A random process  $X(t) = A$ , where  $A$  is random variable uniformly distributed over  $(0,1]$ . Since  $E[A] = \frac{1}{2}$  and  $R_A(\tau) = E[A^2] = \frac{1}{3}$ , this process is stationary. To check for ergodicity in the mean, we apply Eq. (19.3.2) for the time average and check whether it tends to the ensemble average as  $T$  tends to  $\infty$ . Thus

$$\hat{\mu}_X = \frac{1}{2T} \int_{-T}^T X(t) dt = \frac{1}{2T} \int_{-T}^T A dt = A \quad \text{as } T \rightarrow \infty$$

is not equal to the ensemble average  $\mu_X = \frac{1}{2}$ . Hence  $X(t)$  is neither mean-ergodic nor correlation-ergodic.

**Example 19.3.3** We now consider a random process  $X(t) = B + A \sin(\omega_0 t + \Phi)$ , where  $B$ ,  $A$ , and  $\Phi$  are independent random variables.  $A$  and  $B$  are uniformly distributed over  $(0,1]$  and  $(0,2]$  respectively. We will check whether  $X(t)$  is mean-ergodic.

We first determine whether  $X(t)$  is stationary. The mean value  $\mu_X$  is given by

$$E[B + A \sin(\omega_0 t + \Phi)] = \mu_B + E[A \sin(\omega_0 t + \Phi)] = \mu_B$$

since  $E[A \sin(\omega_0 t + \Phi)] = 0$  from previous examples. The autocorrelation function is

$$\begin{aligned} R_X(\tau) &= E\{[B + A \sin(\omega_0 t + \Phi)][B + A \sin(\omega_0 t + \omega_0 \tau + \Phi)]\} \\ &= E[B^2] + 0 + 0 + E[A^2]E[\sin(\omega_0 t + \Phi)A \sin(\omega_0 t + \omega_0 \tau + \Phi)] \\ &= E[B^2] + E[A^2]\frac{1}{2}E[\cos(\omega_0 \tau) - \cos(2\omega_0 t + \omega_0 \tau + 2\Phi)] \\ &= E[B^2] + \frac{1}{2}E[A^2]\cos(\omega_0 \tau) \end{aligned}$$

Hence the process  $X(t)$  is stationary. The autocovariance function  $C_X(\tau)$  is given by

$$\begin{aligned} C_X(\tau) &= R_X(\tau) - \mu_X^2 = E[B^2] + \frac{1}{2}E[A^2]\cos(\omega_0 \tau) - \mu_B^2 \\ &= \sigma_B^2 + \frac{1}{2}E[A^2]\cos(\omega_0 \tau) \end{aligned}$$

and substituting this equation in Eq. (19.3.7), we obtain the variance of the time average as

$$\begin{aligned} \sigma_{\hat{\mu}_X}^2 &= \frac{1}{2T} \left\{ \int_{-2T}^{2T} \left[ \sigma_B^2 + \frac{1}{2}E[A^2]\cos(\omega_0 \tau) \right] \left[ 1 - \frac{|\tau|}{2T} \right] d\tau \right\} \\ &= \frac{1}{2T} \left[ 2T\sigma_B^2 + E[A^2] \frac{2T}{2\pi} \left[ \frac{\sin(\omega_0 T)}{\omega_0 T} \right]^2 \right] = \sigma_B^2 + \frac{E[A^2]}{2\pi} \left[ \frac{\sin(\omega_0 T)}{\omega_0 T} \right]^2 \end{aligned}$$

and

$$\lim_{T \rightarrow \infty} \sigma_{\hat{\mu}_X}^2 = \lim_{T \rightarrow \infty} \left\{ \sigma_B^2 + \frac{E[A^2]}{2\pi} \left[ \frac{\sin(\omega_0 T)}{\omega_0 T} \right]^2 \right\} = \sigma_B^2 \quad \text{if } \omega_0 \neq 0$$

We find that even though  $X(t)$  is stationary, it is not mean-ergodic. Hence the time average is not a good estimator of the ensemble average. We can now substitute  $E[A^2] = \frac{1}{3}$  and  $\sigma_B^2 = \frac{1}{3}$  in the variance equation and write

$$\sigma_{\hat{\mu}_X}^2 = \sigma_B^2 + \frac{E[A^2]}{2\pi} \left[ \frac{\sin(\omega_0 T)}{\omega_0 T} \right]^2 = \frac{1}{3} \left[ 1 + \frac{\sin(\omega_0 T)}{2\pi\omega_0 T} \right]$$

**Example 19.3.4** The autocovariance of a zero mean Gaussian random process  $X(t)$  is given by  $C_X(\tau) = \sigma_X^2 e^{-2\alpha|\tau|}$ , where  $\sigma_X^2$  is its variance. We will check whether this

process is mean-ergodic and correlation-ergodic. The variance of the time average from Eq. (19.3.7) is

$$\begin{aligned}\sigma_{\hat{\mu}_X}^2 &= \frac{1}{2T} \left\{ \int_{-2T}^{2T} C_X(\tau) \left[ 1 - \frac{|\tau|}{2T} \right] d\tau \right\} = \frac{1}{2T} \left\{ \int_{-2T}^{2T} \sigma_X^2 e^{-2\alpha|\tau|} \left[ 1 - \frac{|\tau|}{2T} \right] d\tau \right\} \\ &= \frac{\sigma_X^2}{2\alpha T} \left[ 1 - \frac{1 - e^{-4\alpha T}}{4\alpha T} \right]\end{aligned}$$

and  $\lim_{T \rightarrow \infty} \sigma_{\hat{\mu}_X}^2 = 0$ . Hence the process is mean-ergodic. To determine whether it is correlation-ergodic, the fourth moment  $R_{R_X}^\lambda(\tau) = E[X(t)X(t+\lambda)X(t+\tau)X(t+\tau+\lambda)]$  is calculated from the following result. If  $\{X_1, X_2, X_3, X_4\}$  are zero mean Gaussian-distributed random variables, then  $E[X_1X_2X_3X_4] = E[X_1X_2]E[X_3X_4] + E[X_1X_3]E[X_2X_4] + E[X_1X_4]E[X_2X_3]$ . Hence

$$\begin{aligned}E[X(t)X(t+\lambda)X(t+\tau)X(t+\tau+\lambda)] &= E[X(t)X(t+\lambda)E[X(t+\tau)X(t+\tau+\lambda)]] \\ &\quad + E[X(t)X(t+\tau)]E[X(t+\lambda)X(t+\tau+\lambda)] \\ &\quad + E[X(t)X(t+\tau+\lambda)]E[X(t+\lambda)X(t+\tau)] \\ &= R_X^2(\lambda) + R_X^2(\tau) + R_X(\tau+\lambda)R_X(\tau-\lambda)\end{aligned}$$

The covariance  $C_{R_X}^\lambda(\tau) = R_{R_X}^\lambda(\tau) - R_X^2(\lambda) = R_X^2(\tau) + R_X(\tau+\lambda)R_X(\tau-\lambda)$ , and substituting for the correlation functions, we obtain  $C_{R_X}^\lambda(\tau)$  as

$$C_{R_X}^\lambda(\tau) = \sigma_X^4 e^{-4\alpha|\tau|} + \sigma_X^4 e^{-2\alpha|\tau+\lambda|} e^{-2\alpha|\tau-\lambda|} = 2\sigma_X^4 e^{-4\alpha|\tau|}, \quad \tau > \lambda$$

Substituting for  $C_{R_X}^\lambda(\tau)$  in Eq. (19.3.14), we obtain

$$\begin{aligned}\lim_{T \rightarrow \infty} \sigma_{\hat{R}_X^\lambda}^2 &= \lim_{T \rightarrow \infty} \frac{1}{2T} \left\{ \int_{-2T}^{2T} 2\sigma_X^4 e^{-4\alpha|\tau|} \left[ 1 - \frac{|\tau|}{2T} \right] d\tau \right\} \\ &= \lim_{T \rightarrow \infty} \frac{\sigma_X^4}{\alpha T} \left[ 1 - \frac{1 - e^{-8\alpha T}}{8\alpha T} \right] \rightarrow 0\end{aligned}$$

A similar result holds good for  $\tau < \lambda$ . We conclude from the preceding result that the process  $X(t)$  is also correlation-ergodic and hence power-ergodic.

## 19.4 ESTIMATION OF PARAMETERS OF RANDOM PROCESSES

### 19.4.1 Continuous-Time Processes

The ergodic hypothesis detailed in the previous section enables us to define estimators for the parameters of a random process from its sample realization. If  $X(t)$  is a sample realization of a stationary ergodic random process available only for a time duration  $T$ , then the estimators for the parameters, mean, variance, autocorrelation, autocovariance, and normalized autocovariance can be defined.

Mean:

$$\hat{\mu}_X = \frac{1}{T} \int_0^T X(t) dt \quad (19.4.1)$$

We note that  $\hat{\mu}_X$  is an unbiased estimator of  $\mu_X$  since

$$E[\hat{\mu}_X] = E\left[\frac{1}{T} \int_0^T X(t) dt\right] = \frac{1}{T} \int_0^T E[X(t)] dt = \mu_X$$

The variance of  $\hat{\mu}_X$  is very similar to Eq. (19.3.7) and is given by

$$\begin{aligned} \text{var}[\hat{\mu}_X] &= \sigma_{\hat{\mu}_X}^2 = \frac{1}{T} \left\{ \int_{-T}^T C_X(\tau) \left[ 1 - \frac{|\tau|}{T} \right] d\tau \right\} \\ &= \frac{2}{T} \left\{ \int_0^T C_X(\tau) \left[ 1 - \frac{\tau}{T} \right] d\tau \right\}, \quad \text{even function} \end{aligned} \quad (19.4.2)$$

Variance:

$$\hat{\sigma}_X^2 = \frac{1}{T} \int_0^T [X(t) - \hat{\mu}_X]^2 dt \quad (19.4.3)$$

If the mean value  $\mu_X$  is known, then  $\hat{\sigma}_X^2$  will be an unbiased estimator. We will determine the conditions under which  $\hat{\sigma}_X^2$  given by Eq. (19.4.3) can be an unbiased estimator:

$$\begin{aligned} E[\hat{\sigma}_X^2] &= E\left\{ \frac{1}{T} \int_0^T [X(t) - \hat{\mu}_X]^2 dt \right\} \\ &= E\left\{ \frac{1}{T} \int_0^T [\hat{X}(t) - \mu_X - (\hat{\mu}_X - \mu_X)]^2 dt \right\} \\ &= E\left\{ \frac{1}{T} \int_0^T \left[ (X(t) - \mu_X)^2 + (\hat{\mu}_X - \mu_X)^2 - 2(X(t) - \mu_X)(\hat{\mu}_X - \mu_X) \right] dt \right\} \\ &= E\left\{ \frac{1}{T} \int_0^T \left[ \frac{(X(t) - \mu_X)^2}{\sigma_{\hat{\mu}_X}^2} + \frac{(\hat{\mu}_X - \mu_X)^2}{\sigma_{\hat{\mu}_X}^2} - 2 \frac{(X(t) - \mu_X)(\hat{\mu}_X - \mu_X)}{\sigma_{\hat{\mu}_X}^2} \right] dt \right\} + E\left\{ \frac{(\hat{\mu}_X - \mu_X)^2}{\sigma_{\hat{\mu}_X}^2} \right\} \\ &\quad - \frac{2}{T^2} E\left\{ \int_0^T \int_0^T [X(t) - \mu_X][X(s) - \mu_X] ds dt \right\} \end{aligned}$$

and from Eq. (19.4.2) we have

$$\begin{aligned} E[\hat{\sigma}_X^2] &= \sigma_X^2 + \sigma_{\hat{\mu}_X}^2 - \frac{2}{T} \left\{ \int_{-T}^T C_X(\tau) \left[ 1 - \frac{|\tau|}{T} \right] d\tau \right\} \sigma_{\hat{\mu}_X}^2 \\ &= \sigma_X^2 - \sigma_{\hat{\mu}_X}^2 \end{aligned} \quad (19.4.4)$$

Hence  $\hat{\sigma}_X^2$  cannot be an unbiased estimator for finite  $T$  with the bias equal to  $-\sigma_{\hat{\mu}_X}^2$ . However, it is asymptotically unbiased.

*Autocovariance.* Two estimators for the autocovariance function are given by

$$\bar{C}_X(\tau) = \frac{1}{T-\tau} \int_0^{T-\tau} [X(t) - \hat{\mu}_X][X(t+\tau) - \hat{\mu}_X]dt, \quad 0 < \tau \leq T \quad (19.4.5)$$

$$\hat{C}_X(\tau) = \frac{1}{T} \int_0^{T-\tau} [X(t) - \hat{\mu}_X][X(t+\tau) - \hat{\mu}_X]dt, \quad 0 < \tau \leq T \quad (19.4.6)$$

If the mean  $\mu_X$  is known, then  $\bar{C}_X(\tau)$  given by Eq. (19.4.5) is an unbiased estimator since

$$\begin{aligned} E[\bar{C}_X(\tau)] &= \frac{1}{T-\tau} \int_0^{T-\tau} E\{[X(t) - \mu_X][X(t+\tau) - \mu_X]\} dt \\ &= C_X(\tau) \frac{T-\tau}{T} = C_X(\tau), \quad 0 < \tau \leq T \end{aligned}$$

and  $\hat{C}_X(\tau)$  is a biased estimator as shown below:

$$\begin{aligned} \hat{C}_X(\tau) &= \frac{1}{T} \int_0^{T-\tau} [X(t) - \mu_X][X(t+\tau) - \mu_X]dt \\ &= C_X(\tau) \frac{T-\tau}{T} = C_X(\tau) \left[1 - \frac{\tau}{T}\right], \quad 0 < \tau \leq T \end{aligned}$$

with bias equal to  $-C_X(\tau)(\tau/T)$ .

In actual practice only the sample mean  $\hat{\mu}_X$  is known, and we will determine whether  $\bar{C}_X(\tau)$  can be an unbiased estimator. Equation (19.4.5) can be rewritten as

$$\begin{aligned} E[\bar{C}_X(\tau_1)] &= E\left\{ \frac{1}{T-\tau_1} \int_0^{T-\tau_1} [X(t) - \hat{\mu}_X][X(t+\tau_1) - \hat{\mu}_X]dt \right\} \\ &= E\left\{ \frac{1}{T-\tau_1} \int_0^{T-\tau_1} [(X(t) - \mu_X) - (\hat{\mu}_X - \mu_X)] \right. \\ &\quad \left. \times [(X(t+\tau_1) - \mu_X) - (\hat{\mu}_X - \mu_X)]dt \right\}, \quad 0 < \tau_1 \leq T \end{aligned}$$

and substituting for  $\hat{\mu}_X = (1/T) \int_0^T X(s)ds$  in the equation above, we have

$$\begin{aligned} E[\bar{C}_X(\tau_1)] &= E\left\{ \frac{1}{T-\tau_1} \int_0^{T-\tau_1} [(X(t) - \mu_X)(X(t+\tau_1) - \mu_X) + (\hat{\mu}_X - \mu_X)^2]dt \right\} \\ &\quad - E\left\{ \frac{1}{T-\tau_1} \int_0^{T-\tau_1} [(X(t) - \mu_X)(\hat{\mu}_X - \mu_X)] \right. \\ &\quad \left. + [(X(t+\tau_1) - \mu_X)(\hat{\mu}_X - \mu_X)]dt \right\} \\ &= C_X(\tau_1) + \sigma_{\hat{\mu}_X}^2 - \frac{1}{T(T-\tau_1)} \int_0^{T-\tau_1} \int_0^T C_X(s-t)ds dt \\ &\quad - \frac{1}{T(T-\tau_1)} \int_0^{T-\tau_1} \int_0^T C_X(s-t-\tau_1)ds dt, \quad 0 < \tau_1 \leq T \quad (19.4.7) \end{aligned}$$

where

$$E[(\hat{\mu}_X - \mu_X)^2] = \sigma_{\hat{\mu}_X}^2 = \frac{1}{T} \int_{-T}^T C_X(\tau) \left(1 - \frac{|\tau|}{T}\right) d\tau \quad [\text{from Eq. (19.4.2)}]$$

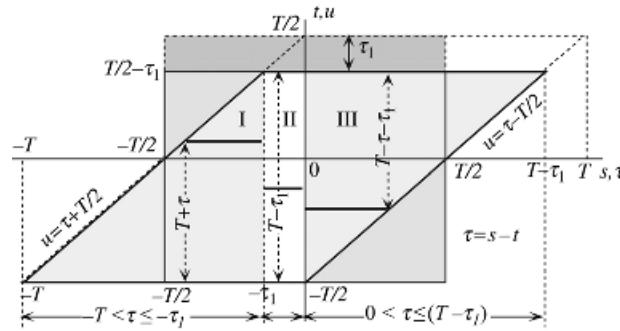


FIGURE 19.4.1

In Eq. (19.4.7) we make the transformation  $\tau = (s - t)$  and  $u = t$ . As the Jacobian of the transformation is 1, we can use techniques similar to those employed in Eqs. (19.3.5)–(19.3.7) to determine  $E[\bar{C}_X(\tau_1)]$ . This transformation amounts to finding the integral in the parallelogram shown in Fig. 19.4.1.

The integration along the  $\tau$  axis is from  $-T$  to  $(T - \tau_1)$ . Thus

$$E[\bar{C}_X(\tau_1)] = C_X(\tau_1) + \sigma_{\hat{\mu}_X}^2 - \frac{1}{T(T - \tau_1)} \left\{ \int du \int_{-T}^{T - \tau_1} [C_X(\tau) + C_X(\tau - \tau_1)] d\tau \right\} \quad (19.4.8)$$

The limits of integration for  $\int du$  are determined as follows in the three regions:

*Region I:*  $-T < \tau \leq -\tau_1$

$$\int_{-T/2}^{\tau + T/2} du = T + \tau = T \left( 1 - \frac{|\tau|}{T} \right) \quad (19.4.9a)$$

*Region II:*  $-\tau_1 < \tau \leq 0$

$$\int_{-T/2}^{T/2 - \tau_1} du = T - \tau_1 = T \left( 1 - \frac{\tau_1}{T} \right) \quad (19.4.9b)$$

*Region III:*  $0 < \tau \leq (T - \tau_1)$

$$\int_{\tau - T/2}^{T/2 - \tau_1} du = T - \tau - \tau_1 = T \left( 1 - \frac{\tau + \tau_1}{T} \right) \quad (19.4.9c)$$

Substituting Eqs. (19.4.9) in Eq. (19.4.8), we obtain  $E[\bar{C}_X(\tau_1)]$  as follows:

$$E[\bar{C}_X(\tau_1)] = C_X(\tau_1) + \sigma_{\hat{\mu}_X}^2 - \frac{1}{(T - \tau_1)} \left\{ \begin{aligned} & \int_{-T}^{-\tau_1} \left[ 1 - \frac{|\tau|}{T} \right] [C_X(\tau) + C_X(\tau - \tau_1)] d\tau \\ & + \int_{-\tau_1}^0 \left[ 1 - \frac{\tau_1}{T} \right] [C_X(\tau) + C_X(\tau - \tau_1)] d\tau \\ & + \int_0^{T - \tau_1} \left[ 1 - \frac{\tau + \tau_1}{T} \right] [C_X(\tau) + C_X(\tau - \tau_1)] d\tau \end{aligned} \right\}, \quad 0 < \tau_1 \leq T \quad (19.4.10)$$

For large values of  $T$  or for low values of the lag  $\tau_1$ , the terms within braces can be approximated as  $2\sigma_{\hat{\mu}_X}^2$ , and the approximation for  $E[\bar{C}_X(\tau_1)]$  is

$$E[\bar{C}_X(\tau_1)] \approx C_X(\tau_1) - \sigma_{\hat{\mu}_X}^2, \quad 0 < \tau_1 \leq T \quad (19.4.11)$$

Thus,  $\bar{C}_X(\tau)$  is a biased estimator with the bias  $B \approx -\sigma_{\hat{\mu}_X}^2$ .

In an analogous manner the expected value of  $\hat{C}_X(\tau_1)$  given by Eq. (19.4.6) can be evaluated as follows:

$$\begin{aligned} E[\hat{C}_X(\tau_1)] &= E\left\{\frac{1}{T} \int_0^{T-\tau_1} [(X(t) - \mu_X)(X(t + \tau) - \mu_X) + E(\hat{\mu}_X - \mu_X)^2] dt\right\} \\ &\quad - E\left\{\frac{1}{T} \int_0^{T-\tau_1} [(X(t) - \mu_X)(\hat{\mu}_X - \mu_X)] + [(X(t + \tau_1) - \mu_X)(\hat{\mu}_X - \mu_X)] dt\right\} \\ &= \frac{T - \tau_1}{T} [C_X(\tau_1) + \sigma_{\hat{\mu}_X}^2] - \frac{1}{T^2} \int_0^{T-\tau_1} \int_0^T C_X(s - t) ds dt \\ &\quad - \frac{1}{T^2} \int_0^{T-\tau_1} \int_0^T C_X(s - t - \tau_1) ds dt, \quad 0 < \tau_1 \leq T \end{aligned} \quad (19.4.12)$$

By substituting  $\tau = (s - t)$  and  $u = t$  and performing the integration in the parallelogram shown in Fig. 19.4.1, the final result similar to Eq. (19.4.10) is

$$\begin{aligned} E[\hat{C}_X(\tau_1)] &= \frac{T - \tau_1}{T} [C_X(\tau_1) + \sigma_{\hat{\mu}_X}^2] \\ &\quad - \frac{1}{T} \left\{ \begin{aligned} &\int_{-T}^{-\tau_1} \left[ 1 - \frac{|\tau|}{T} \right] [C_X(\tau) + C_X(\tau - \tau_1)] d\tau \\ &+ \int_{-\tau_1}^0 \left[ 1 - \frac{\tau_1}{T} \right] [C_X(\tau) + C_X(\tau - \tau_1)] d\tau \\ &+ \int_0^{T-\tau_1} \left[ 1 - \frac{\tau + \tau_1}{T} \right] [C_X(\tau) + C_X(\tau - \tau_1)] d\tau \end{aligned} \right\}, \quad 0 < \tau_1 \leq T \end{aligned} \quad (19.4.13)$$

For large values of  $T$  or for low values of the lag  $\tau_1$ , the terms within braces can be approximated as  $[(T - \tau_1)/T]2\sigma_{\hat{\mu}_X}^2$ , and the approximation for  $E[\hat{C}_X(\tau_1)]$  is

$$E[\hat{C}_X(\tau_1)] \approx [C_X(\tau_1) - \sigma_{\hat{\mu}_X}^2] \left[ 1 - \frac{\tau_1}{T} \right], \quad 0 < \tau_1 \leq T \quad (19.4.14)$$

Thus,  $\hat{C}_X(\tau_1)$  is also a biased estimator with a larger bias than  $\bar{C}_X(\tau)$  with the bias

$$B \approx -C_X(\tau_1) \frac{\tau_1}{T} - \sigma_{\hat{\mu}_X}^2 \left[ 1 - \frac{\tau_1}{T} \right], \quad 0 < \tau_1 \leq T$$

We will show later that  $\hat{C}_X(\tau)$  is the preferred estimator as it gives better minimum mean-square error.

**Example 19.4.1** The autocovariance function of a random process  $X(t)$  is  $C_X(\tau) = e^{-\alpha|\tau|}$ , with  $\alpha = 0.5$ . The data interval  $T = 50$ . We will find the variance of the estimated mean

value  $\hat{\mu}_X$  as a function of  $T$ , compare the true value  $C_X(\tau)$  to  $E[\bar{C}_X(\tau)]$  given by Eq. (19.4.10) and  $E[\hat{C}_X(\tau)]$  given by Eq. (19.4.13), and finally compare the approximations to  $E[\hat{C}_X(\tau)]$  given by Eq. (19.4.14).

*Variance*  $\sigma_{\hat{\mu}_X}^2(T)$ : From Eq. (19.4.2), we obtain

$$\begin{aligned}\sigma_{\hat{\mu}_X}^2(T) &= \frac{2}{T} \left\{ \int_0^T C_X(\tau) \left[ 1 - \frac{\tau}{T} \right] d\tau \right\} = \frac{2}{T} \left\{ \int_0^T e^{-\alpha\tau} \left[ 1 - \frac{\tau}{T} \right] d\tau \right\} \\ &= \frac{2[e^{-\alpha T} + \alpha T - 1]}{\alpha^2 T^2}\end{aligned}$$

A graph of  $\sigma_{\hat{\mu}_X}^2(T)$  is shown in Fig. 19.4.2 for  $\alpha = 0.5$ , and the estimate of the mean  $\hat{\mu}_X$  is asymptotically unbiased since  $\lim_{T \rightarrow \infty} \sigma_{\hat{\mu}_X}^2(T) \rightarrow 0$ .

*Means of Estimated Autocovariances.*  $\bar{C}_X(\tau)$  and  $\hat{C}_X(\tau)$ . From Eq. (19.4.10) the mean value of  $\bar{C}_X(\tau)$  is given by

$$\begin{aligned}E[\bar{C}_X(\tau)] &= e^{-\alpha\tau} + \frac{2(e^{-50\alpha} + 50\alpha - 1)}{2500\alpha^2} \\ &\quad - \frac{1}{50 - \tau} \int_{-50}^{\tau} (e^{-\alpha\beta} + e^{-\alpha(\beta - \tau)}) \left( 1 - \frac{|\beta|}{50} \right) d\beta \\ &\quad - \frac{1}{50 - \tau} \int_{-\tau}^0 (e^{-\alpha\beta} + e^{-\alpha(\beta - \tau)}) \left( 1 - \frac{\tau}{50} \right) d\beta \\ &\quad - \frac{1}{50 - \tau} \int_0^{50 - \tau} (e^{-\alpha\beta} + e^{-\alpha(\beta - \tau)}) \left( 1 - \frac{\beta + \tau}{50} \right) d\beta, \quad 0 < \tau \leq 50\end{aligned}$$

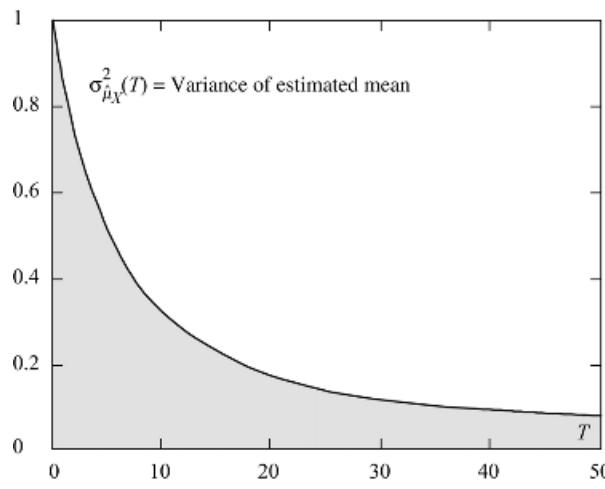


FIGURE 19.4.2

and from Eq. (19.4.13) the mean value of  $\hat{C}_X(\tau)$  is given by

$$\begin{aligned}
 E[\hat{C}_X(\tau)] &= \frac{50 - \tau}{50} \left[ e^{-\alpha\tau} + \frac{2(e^{-50\alpha} + 50\alpha - 1)}{2500\alpha^2} \right] \\
 &\quad - \frac{1}{50} \int_{-50}^{\tau} (e^{-\alpha\beta} + e^{-\alpha(\beta-\tau)}) \left( 1 - \frac{|\beta|}{50} \right) d\beta \\
 &\quad - \frac{1}{50} \int_{-\tau}^0 (e^{-\alpha\beta} + e^{-\alpha(\beta-\tau)}) \left( 1 - \frac{\tau}{50} \right) d\beta \\
 &\quad - \frac{1}{50} \int_0^{50-\tau} (e^{-\alpha\beta} + e^{-\alpha(\beta-\tau)}) \left( 1 - \frac{\beta + \tau}{50} \right) d\beta, \quad 0 < \tau \leq 50
 \end{aligned}$$

In graphing the estimated autocovariance functions, it is usual to plot the lag length  $\tau$  upto 10–20% of the length  $T$ . Expected values  $E[\bar{C}_X(\tau)]$  and  $E[\hat{C}_X(\tau)]$  are shown along with  $C_X(\tau)$  in Fig. 19.4.3 for  $\tau = 0$ –10 and tabulated in Table 19.4.1. The bias as shown in the figure is approximately equal to  $B \approx -\hat{\sigma}_{\mu_X}^2(T) = -0.0768$ .

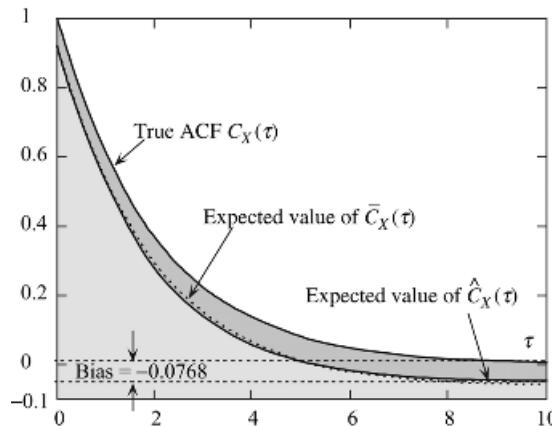


FIGURE 19.4.3

TABLE 19.4.1.

$\tau$	$C_X(\tau)$	$E[\bar{C}_X(\tau)]$	$\bar{C}_X(\tau)$	$E[\hat{C}_X(\tau)]$	$\hat{C}_X(\tau)$
0	1	0.9232	0.9232	0.9232	0.9232
1	0.606531	0.529171	0.529731	0.518588	0.519136
2	0.367879	0.291129	0.291079	0.279484	0.279436
3	0.22313	0.147612	0.14633	0.138755	0.13755
4	0.135335	0.061327	0.058535	0.056421	0.053852
5	0.082085	0.009859	0.005285	0.008873	0.004756
6	0.049787	-0.020499	-0.027013	-0.018039	-0.023771
7	0.030197	-0.038005	-0.046603	-0.032684	-0.040078
8	0.018316	-0.047664	-0.058484	-0.040038	-0.049127
9	0.011109	-0.052531	-0.065691	-0.044075	-0.053867
10	0.006738	-0.054435	-0.070062	-0.043548	-0.05605

The figure also shows that there is very little difference between  $E[\bar{C}_X(\tau)]$  and  $E[\hat{C}_X(\tau)]$ . Table 19.4.1 also shows the approximation  $E[\bar{C}_X(\tau)]$  given by Eq. (19.4.11) and the approximation  $E[\hat{C}_X(\tau)]$  given by Eq. (19.4.14).

*Covariance of ACF Estimators.* Before finding the covariance of the estimator  $\bar{C}_X(\tau)$  at two different lags  $\tau_1$  and  $\tau_2$ , we will first find the covariance  $\text{cov}[X(\tau)X(t+\tau_1), X(t)X(s+\tau_2)]$ :

$$\begin{aligned} & \text{cov}[X(\tau)X(t+\tau_1), X(s)X(s+\tau_2)] \\ &= E\{[(X(t) - \mu_X)(X(t+\tau_1) - \mu_X)][(X(s) - \mu_X)(X(s+\tau_2) - \mu_X)]\} \\ &= E[(X(t) - \mu_X)(X(t+\tau_1) - \mu_X)(X(s) - \mu_X)(X(s+\tau_2) - \mu_X)] \\ &\quad - E[(X(t) - \mu_X)(X(t+\tau_1) - \mu_X)]E[(X(s) - \mu_X)(X(s+\tau_2) - \mu_X)] \end{aligned} \quad (19.4.15)$$

Expressing  $E[(X(t) - \mu_X)(X(t+\tau_1) - \mu_X)(X(s) - \mu_X)(X(s+\tau_2) - \mu_X)]$  of Eq. (19.4.15) in terms of the fourth cumulant  $\kappa_4$  from Eq. (11.6.8), we obtain

$$\begin{aligned} & E[(X(t) - \mu_X)(X(t+\tau_1) - \mu_X)(X(s) - \mu_X)(X(s+\tau_2) - \mu_X)] \\ &= \kappa_4(s-t, \tau_1, \tau_2) + C_X(\tau_1)C_X(\tau_2) \\ &\quad + C_X(s-t)C_X(s-t+\tau_2-\tau_1) + C_X(s-t+\tau_2)C_X(s-t-\tau_1) \end{aligned}$$

and substituting in Eq. (19.4.15), we obtain

$$\begin{aligned} & \text{cov}[X(t)X(t+\tau_1), X(s)X(s+\tau_2)] \\ &= \kappa_4(s-t, \tau_1, \tau_2) + C_X(s-t)C_X(s-t+\tau_2-\tau_1) + C_X(s-t+\tau_2)C_X(s-t-\tau_1) \end{aligned} \quad (19.4.16)$$

which is an exact expression. The presence of the fourth cumulant  $\kappa_4$  in Eq. (19.4.16) renders evaluation of the covariance intractable. However, if  $X(t)$  is a Gaussian random process, then the fourth cumulant  $\kappa_4$  is zero and the covariance of  $[X(t)X(t+\tau_1), X(s)X(s+\tau_2)]$  becomes

$$\begin{aligned} & \text{cov}[X(t)X(t+\tau_1), X(s)X(s+\tau_2)] \\ &= C_X(s-t)C_X(s-t+\tau_2-\tau_1) + C_X(s-t+\tau_2)C_X(s-t-\tau_1) \end{aligned} \quad (19.4.17)$$

Assuming without loss of generality that  $0 \leq \tau_1 \leq \tau_2$ , the covariance of the estimator  $\bar{C}_X(\tau)$  at two different lags  $\tau_1$  and  $\tau_2$  can be given by

$$\begin{aligned} & \text{cov}[\bar{C}_X(\tau_1), \bar{C}_X(\tau_2)] \\ &= \text{cov}\left\{\frac{1}{(T-\tau_1)} \int_0^{T-\tau_1} X(t)X(t+\tau_1) dt, \frac{1}{(T-\tau_2)} \int_0^{T-\tau_2} X(s)X(s+\tau_2) ds\right\} \\ &= \frac{1}{(T-\tau_1)(T-\tau_2)} \int_0^{T-\tau_2} \int_0^{T-\tau_1} \text{cov}[X(t)X(t+\tau_1)X(s)X(s+\tau_2)] dt ds \end{aligned} \quad (19.4.18)$$

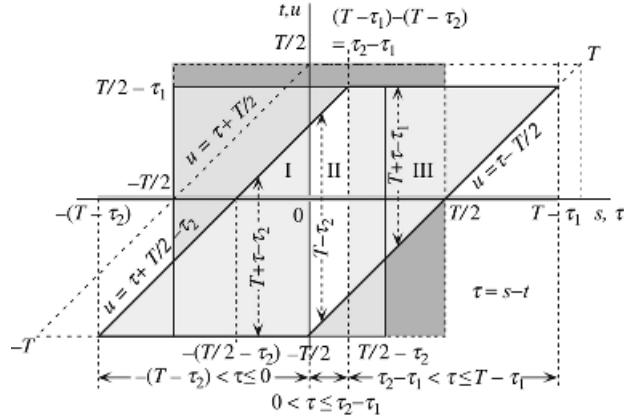


FIGURE 19.4.4

Substituting for the covariance in Eq. (19.4.18) from Eq. (19.4.16), we obtain

$$\begin{aligned} \text{cov}[\bar{C}_X(\tau_1), \bar{C}_X(\tau_2)] &= \frac{1}{(T - \tau_1)(T - \tau_2)} \int_0^{T - \tau_2} \int_0^{T - \tau_1} C_X(s - t) C_X(s - t + \tau_2 - \tau_1) \\ &\quad + C_X(s - t + \tau_2) C_X(s - t - \tau_1) dt ds \end{aligned} \quad (19.4.19)$$

We now make the transformation  $\tau = (s - t)$  and  $u = t$ , and as the Jacobian of the transformation is 1, we can use techniques similar to those employed in Eq. (19.4.8) to determine the covariance. This transformation amounts to finding the integral in the parallelogram shown in Fig. 19.4.4. The integration along the  $\tau$  axis is from  $-(T - \tau_2)$  to  $(T - \tau_1)$ . Thus, for a Gaussian process  $X(t)$ , we obtain

$$\begin{aligned} \text{cov}[\bar{C}_X(\tau_1), \bar{C}_X(\tau_2)] &= \frac{1}{(T - \tau_1)(T - \tau_2)} \int du \left\{ \int_{-(T - \tau_2)}^{T - \tau_1} C_X(\tau) C_X(\tau + \tau_2 - \tau_1) \right. \\ &\quad \left. + C_X(\tau + \tau_2) C_X(\tau - \tau_1) d\tau \right\} \end{aligned} \quad (19.4.20)$$

The integral  $\int du$  in Eq. (19.4.20) is calculated in three regions as shown in Fig. 19.4.4 in a manner similar to that used in Eqs. (19.4.9):

*Region I:*  $-(T - \tau_2) < \tau \leq 0$

$$\int_{-T/2}^{\tau + T/2 - \tau_2} du = T + \tau - \tau_2 = T \left( 1 - \frac{|\tau| - \tau_2}{T} \right) \quad (19.4.21a)$$

*Region II:*  $0 \leq (\tau_2 - \tau_1)$

$$\int_{\tau - T/2}^{\tau + T/2 - \tau_2} du = T - \tau_2 = T \left( 1 - \frac{\tau_2}{T} \right) \quad (19.4.21b)$$

Region III:  $(\tau_2 - \tau_1) < \tau \leq (T - \tau_1)$

$$\int_{\tau-T/2}^{T/2-\tau_1} du = T - \tau - \tau_1 = T \left(1 - \frac{\tau + \tau_1}{T}\right) \quad (19.4.21c)$$

Substituting Eqs. (19.4.21) into Eq. (19.4.20), the covariance between adjacent values  $\tau_1$  and  $\tau_2$  of the sample autocovariance for a Gaussian process  $X(t)$  is given by

$$\begin{aligned} \text{cov}[\bar{C}_X(\tau_1), \bar{C}_X(\tau_2)] &= \frac{1}{(T - \tau_1)(T - \tau_2)} \\ &\left\{ \begin{array}{ll} \int_{T-\tau_2}^0 (T - |\tau| - \tau_2)[C_X(\tau)C_X(\tau + \tau_2 - \tau_1) \\ \quad + C_X(\tau + \tau_2)C_X(\tau - \tau_1)]d\tau \\ + \int_0^{\tau_2-\tau_1} (T - \tau_2)[C_X(\tau)C_X(\tau + \tau_2 - \tau_1) \\ \quad + C_X(\tau + \tau_2)C_X(\tau - \tau_1)]d\tau \\ + \int_{\tau_2-\tau_1}^{T-\tau_1} (T - \tau - \tau_1)[C_X(\tau)C_X(\tau + \tau_2 - \tau_1) \\ \quad + C_X(\tau + \tau_2)C_X(\tau - \tau_1)]d\tau \end{array} \right\} \end{aligned} \quad (19.4.22)$$

Although Eq. (19.4.22) was derived with the condition  $0 \leq \tau_1 \leq \tau_2$ , it can be shown that it holds good for all  $\tau_1$  and  $\tau_2$ . The covariance is generally difficult to evaluate. For large  $T$ ,  $(T - \tau_1) \approx T \approx (T - \tau_2)$ , and the integral  $\int du \approx T$  in Eq. (19.4.21). With these approximations the distinction between  $\bar{C}_X(\tau)$  and  $\hat{C}_X(\tau)$  becomes moot and Eq. (19.4.22) becomes

$$\begin{aligned} \text{cov}[\bar{C}_X(\tau_1), \bar{C}_X(\tau_2)] &\approx \text{cov}[\hat{C}_X(\tau_1), \hat{C}_X(\tau_2)] \\ &\approx \frac{1}{T} \int_{-\infty}^{\infty} [C_X(\tau)C_X(\tau + \tau_2 - \tau_1) + C_X(\tau + \tau_2)C_X(\tau - \tau_1)]d\tau \end{aligned} \quad (19.4.23)$$

The variance of the sample autocovariance function is obtained by substituting  $\tau_1 = \tau_2 = s$  in Eqs. (19.4.20) and (19.4.22). With this substitution Eqs. (19.4.21) become  $\int du = T - s - |\tau|$ , and the general form for the variance is

$$\begin{aligned} \text{var}[\bar{C}_X(s)] &= \frac{1}{(T - |s|)^2} \\ &\times \int_{-(T-s)}^{T-s} \{(T - s - |\tau|)[C_X^2(\tau) + C_X(\tau + s)C_X(\tau - s)]\}d\tau, \quad s < T \end{aligned} \quad (19.4.24)$$

and the corresponding variance for the estimator  $\hat{C}_X(\tau)$  is obtained by substituting  $T^2$  for  $(T - |s|)^2$  in the denominator of Eq. (19.4.24). The estimator  $\hat{C}_X(\tau)$  is more often used since it yields a minimum mean-square estimator as shown in Example 19.4.2.

The approximations for the variances  $\bar{C}_X(s)$  and  $\hat{C}_X(\tau)$  corresponding to Eq. (19.4.16) for large  $T$  are

$$\text{var}[\bar{C}_X(s)] \approx \text{var}[\hat{C}_X(s)] \approx \frac{1}{T} \int_{-\infty}^{\infty} [C_X^2(\tau) + C_X(\tau + s)C_X(\tau - s)]d\tau \quad (19.4.25)$$

**Example 19.4.2** We will continue Example 19.4.1 assuming that  $X(t)$  is a zero mean Gaussian random process with autocovariance function  $C_X(\tau) = e^{-\alpha|\tau|}$ . Since we are

given that  $X(t)$  is zero mean,  $\bar{C}_X(\tau)$  will be an unbiased estimator. We have to find the variances of the two estimators  $\bar{C}_X(\tau)$  and  $\hat{C}_X(\tau)$  and calculate the minimum mean-square error for both. Substituting  $C_X(\tau) = e^{-\alpha|\tau|}$  in Eq. (19.4.24), the variance of  $\bar{C}_X(\tau)$  is given by

$$\text{var}[\bar{C}_X(s)] = \frac{1}{(T-|s|)^2} \int_{-(T-s)}^{T-s} \{(T-s-|\tau|)[e^{-2\alpha|\tau|} + e^{-\alpha|\tau+s|}e^{-\alpha|\tau-s|}]\} d\tau \quad (19.4.26)$$

Since the integrand is even, Eq. (19.4.26) can be rewritten as

$$\text{var}[\bar{C}_X(s)] = \frac{2}{(T-s)^2} \int_0^{T-s} \{(T-s-\tau)[e^{-2\alpha\tau} + e^{-\alpha|\tau+s|}e^{-\alpha|\tau-s|}]\} d\tau \quad (19.4.27)$$

where we have removed the absolute sign from the quantities where it is clear that  $\tau$  is always positive. Equation (19.4.27) can be integrated numerically to obtain the variance of the unbiased estimator. However, it is instructive to find a closed-form solution. The absolute value signs in the product term  $e^{-\alpha|\tau+s|} e^{-\alpha|\tau-s|}$  has to be carefully removed. In Fig. 19.4.5a,  $s < T/2$  and the range of integration for  $\tau$  contains two segments  $(0, s)$  and  $(s, T-s)$ .

In Fig. 19.4.5b,  $s > T/2$  and the range of integration for  $\tau$  contains only one segment  $(0, T-s)$ .

Hence we can write

$$\begin{aligned} 0 < s \leq \frac{T}{2} \\ e^{-\alpha|\tau+s|} e^{-\alpha|\tau-s|} &= \begin{cases} e^{-\alpha(\tau+s)} e^{\alpha(\tau-s)} = e^{-2\alpha s}, & 0 < \tau \leq s \\ e^{-\alpha(\tau+s)} e^{-\alpha(\tau-s)} = e^{-2\alpha\tau}, & s < \tau \leq T-s \end{cases} \\ \frac{T}{2} < s \leq T \\ e^{-\alpha|\tau+s|} e^{-\alpha|\tau-s|} &= e^{-\alpha(\tau+s)} e^{\alpha(\tau-s)} = e^{-2\alpha s}, \quad 0 < \tau \leq T-s \end{aligned} \quad (19.4.28)$$

Substituting Eq. (19.4.28) in Eq. (19.4.27), the variance of the unbiased estimator  $\hat{C}_X(s)$  can be written as follows:

$$\text{var}[\bar{C}_X(s)] = \frac{2}{(T-s)^2} \left\{ \begin{array}{l} \int_0^s \{(T-s-\tau)(e^{-2\alpha\tau} + e^{-2\alpha s})\} d\tau \\ + \int_s^{T-s} \{(T-s-\tau)(e^{-2\alpha\tau} + e^{-2\alpha s})\} d\tau, \quad 0 < s \leq \frac{T}{2} \\ \int_0^{T-s} \{(T-s-\tau)[e^{-2\alpha\tau} + e^{-2\alpha s}]\} d\tau, \quad \frac{T}{2} < s \leq T \end{array} \right\} \quad (19.4.29)$$

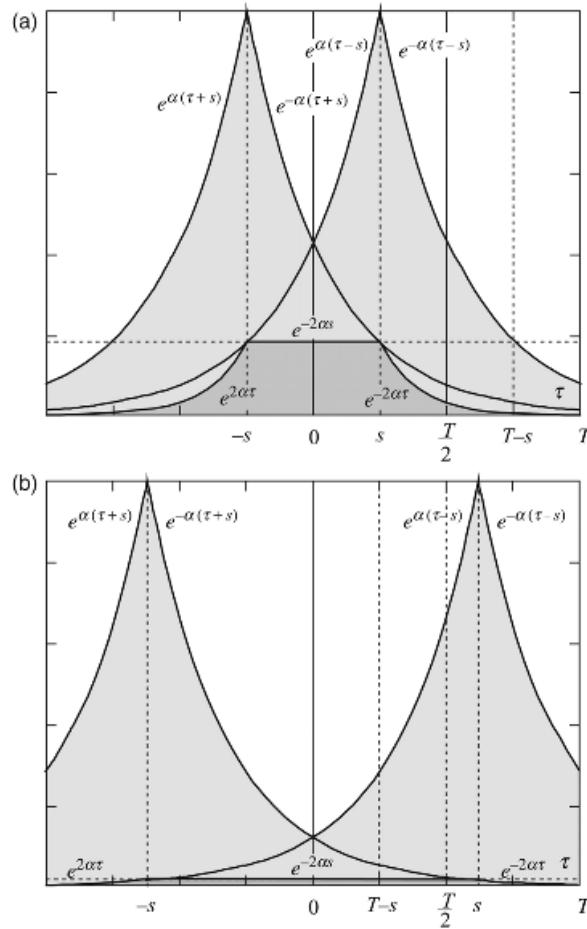


FIGURE 19.4.5

Integrating Eq. (19.4.29), we obtain the variance of the estimator  $\bar{C}_X(s)$  as

$$\begin{aligned} \text{var}[\bar{C}_X(s)] &= \frac{1}{2\alpha^2(T-s)^2} \left\{ \begin{array}{ll} \left[ e^{-2\alpha(T-s)} + 2\alpha(T-s) - 1 \right] + e^{-2\alpha(T-s)} \\ + e^{-2\alpha s} [2\alpha(T-2s) - 1 + 2\alpha^2 s(2T-3s)], & 0 < s \leq \frac{T}{2} \\ \left[ e^{-2\alpha\tau(T-s)} + 2\alpha(T-s) - 1 \right] + e^{-2\alpha s} \left[ \frac{T-s}{T} \right]^2, & \frac{T}{2} < s \leq T \end{array} \right\} \end{aligned} \quad (19.4.30)$$

The variance of the biased estimator  $\hat{C}_X(s)$  obtained by replacing  $(T-s)^2$  by  $T^2$  in the denominator of Eq. (19.4.30). The minimum mean-square errors for both the estimators

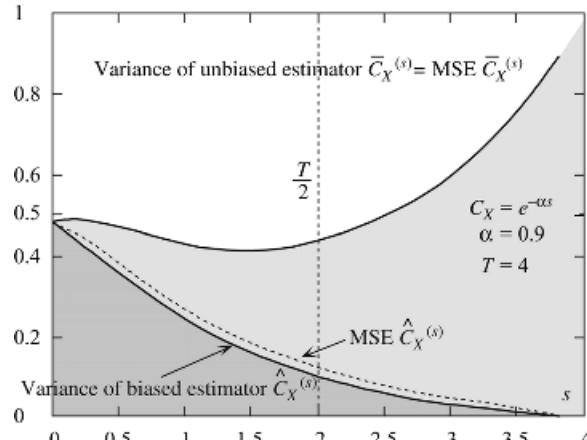


FIGURE 19.4.6

are given by

$$\begin{aligned}\bar{\varepsilon}_U^2(s) &= E[\bar{C}_X(s) - C_X(s)]^2 = \text{var}[\bar{C}_X(s)] \\ \bar{\varepsilon}_B^2(s) &= E[\hat{C}_X(s) - C_X(s)]^2 = \text{var}[\hat{C}_X(s)] + B^2\end{aligned}\quad (19.4.31)$$

The mean-square error for the unbiased estimator  $\hat{C}_X(s)$  is the same as its variance. The bias  $B$  for  $\hat{C}_X(s)$  as shown earlier for a zero mean process is  $B = -C_X(\tau)[|\tau|/T]$ , and substituting for  $B$  in Eq. (19.4.31), the mean-square error for the biased estimator  $\hat{C}_X(s)$  is

$$\bar{\varepsilon}_B^2(s) = \text{var}[\hat{C}_X(s)] + \left(\frac{\tau}{T}\right)^2 e^{-2\alpha s} \quad (19.4.32)$$

The mean-square errors and the variances are shown in Fig. 19.4.6 for  $T = 4$ ,  $\alpha = 0.9$ . It is seen from the figure that the mean-square error for a biased estimator  $\hat{C}_X(s)$  is less than that of the unbiased one  $\hat{C}_X(s)$ .

*Autocorrelation.* If  $\hat{\mu}_X$  is zero in Eqs. (19.4.5) and (19.4.6), we obtain two corresponding estimators for the autocorrelation function  $R_X(\tau)$ :

$$\text{Unbiased: } \bar{R}_X(\tau) = \frac{1}{T - |\tau|} \int_0^{T-|\tau|} X(t)X(t + \tau)dt, \quad 0 \leq \tau \leq T \quad (19.4.33a)$$

$$\text{Biased: } \hat{R}_X(\tau) = \frac{1}{T} \int_0^{T-|\tau|} X(t)X(t + \tau)dt \quad (19.4.34a)$$

$\bar{R}_X(\tau)$  is an unbiased estimator since

$$\begin{aligned}E[\bar{R}_X(\tau)] &= E\left[\frac{1}{T - |\tau|} \int_0^{T-|\tau|} X(t)X(t + \tau)dt\right] \\ &= \frac{1}{T - |\tau|} \int_0^{T-|\tau|} E[X(t)X(t + \tau)]dt = R_X(\tau)\end{aligned}\quad (19.4.33b)$$

and  $E[\hat{R}_X(\tau)]$  is a biased estimator as shown below:

$$E[\hat{R}_X(\tau)] = E\left[\frac{T - |\tau|}{T} \int_0^{T-|\tau|} X(t)X(t + \tau)dt\right] = \left[1 - \frac{|\tau|}{T}\right]R_X(\tau) \quad (19.4.34b)$$

However,  $\hat{R}_X(\tau)$ , although biased, gives a smaller mean-square error as shown in Example 19.4.2 and hence is the preferred estimator.

*Normalized Autocovariance:*

$$\hat{\rho}_X(\tau) = \frac{\hat{C}_X(\tau)}{\hat{C}_X(0)} \quad (19.4.35)$$

$\hat{\rho}_X(\tau)$  will not be an unbiased estimator.

### 19.4.2 Discrete-Time Processes

In actual practice observations are made on the sample function of a stationary ergodic random process  $X(t)$  at equally spaced time intervals  $t_1, t_2, \dots, t_n$  with corresponding random variables  $X_1, X_2, \dots, X_n$ . These observation random variables will *not be independent* and will be used in the estimation of mean and autocovariance of the random process. Since  $\{X_i\}$  are samples of a stationary random process, the means and variances of these samples are the same as those of the original process:

$$E[X_i] = \mu_X : \text{var}[X_i] = \sigma_X^2, \quad i = 1, \dots, n \quad (19.4.36)$$

#### *Estimation of the Mean*

In the previous chapter we estimated the mean  $\hat{\mu}_X$  of a random variable by minimizing its variance from a set of independent observation random variables  $\{X_1, \dots, X_n\}$ . We can use the same techniques to estimate the mean  $\hat{\mu}_X$  of the random process  $X(t)$  except that the set  $\{X_1, \dots, X_n\}$  is not independent. We define the estimated mean  $\hat{\mu}_X$  as the weighted sum of the dependent observations  $\{X_1, \dots, X_n\}$ :

$$\hat{\mu}_X = \sum_{i=1}^n a_i X_i = \mathbf{a}^T \mathbf{X} = \mathbf{X}^T \mathbf{a} \quad (19.4.37)$$

We have to determine the weight vector  $\mathbf{a}$  if possible under the criteria of unbiasedness and minimum variance. The unbiasedness criterion yields the condition

$$\sum_{i=1}^n a_i = \mathbf{a}^T \mathbf{1} = 1 \quad (19.4.38)$$

where  $\mathbf{1}$  is a unity vector defined by  $\mathbf{1} = [1, \dots, 1]^T$ . The variance of the estimator  $\hat{\mu}_X$  is given by

$$\text{var}(\hat{\mu}_X) = \sigma_{\hat{\mu}_X}^2 = E[\mathbf{a}^T (\mathbf{X} - \mathbf{\mu}_X)(\mathbf{X} - \mathbf{\mu}_X)^T \mathbf{a}] = \mathbf{a}^T \mathbf{C}_X \mathbf{a} \quad (19.4.39)$$

where  $\mathbf{C}_X$  is the autocovariance matrix of the vector random variable  $\mathbf{X}$ , defined by

$$\mathbf{C}_X = \mathbf{E}[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{X} - \boldsymbol{\mu}_X)^T]$$

$$= \begin{bmatrix} C_X(0) & C_X(1) & C_X(2) & \cdots & C_X(n-1) \\ C_X(-1) & C_X(0) & C_X(1) & \cdots & C_X(n-2) \\ C_X(-2) & C_X(-1) & C_X(0) & \cdots & C_X(n-3) \\ \vdots & \vdots & \vdots & & \vdots \\ C_X[-(n-1)] & C_X[-(n-2)] & C_X[-(n-3)] & \cdots & C_X(0) \end{bmatrix} \quad (19.4.40)$$

with  $C_X(h)$  defined as the discrete autocovariance given by

$$C_X(h) = E[(X_i - \mu_X)(X_{i+h} - \mu_X)] \quad (19.4.41)$$

and  $C_X(h) = C_X(-h)$  since  $\mathbf{C}_X$  is a real symmetric matrix. Equation (19.4.40) has to be minimized subject to the constraint of Eq. (19.4.38). The unconstrained function to be minimized after adjoining with the Lagrange multiplier  $\lambda$  is given by

$$J = \mathbf{a}^T \mathbf{C}_X \mathbf{a} + \lambda(\mathbf{a}^T \mathbf{1} - 1) \quad (19.4.42)$$

where  $\mathbf{1}$  is an  $n$  vector of ones given by  $\mathbf{1} = [1, 1, \dots, 1]^T$ . Differentiating with respect to the vector  $\mathbf{a}$ , we obtain

$$\frac{\partial J}{\partial \mathbf{a}} = \frac{\partial}{\partial \mathbf{a}} [\mathbf{a}^T \mathbf{C}_X \mathbf{a} + \lambda(\mathbf{a}^T \mathbf{1} - 1)] = 2\mathbf{C}_X \mathbf{a} + \lambda \mathbf{1} = 0$$

Hence

$$\mathbf{a} = -\frac{\lambda}{2} \mathbf{C}_X^{-1} \mathbf{1} \quad (19.4.43)$$

Solving for  $\lambda$  from the constraint equation [Eq. (19.4.38)] and substituting this result in Eq. (19.4.43), we obtain

$$\mathbf{a} = \frac{\mathbf{C}_X^{-1}}{\mathbf{1}^T \mathbf{C}_X^{-1} \mathbf{1}} \mathbf{1} \quad (19.4.44)$$

The solution for the coefficients  $\{a_i, i = 1, \dots, n\}$  will not be the same for every  $n$ . If the  $X_i$  values are independent, then  $\mathbf{C}_X = \mathbf{I}$  and Eq. (19.4.44) reduces to

$$\mathbf{a} = \frac{\frac{1}{\sigma_X^2} \mathbf{I}}{\mathbf{1}^T \frac{1}{\sigma_X^2} \mathbf{I} \mathbf{1}} \mathbf{1} = \frac{1}{\mathbf{1}^T \mathbf{1}} \mathbf{1} = \frac{1}{n} \mathbf{1} \quad (19.4.45)$$

and all coefficients have the same weight equaling  $1/n$ , a result already obtained in Eq. (18.3.9). As it is, each  $a_i$  in Eq. (19.4.44) will be different and the solution becomes cumbersome. To make the solution intuitively appealing we will assume that all the coefficients  $\{a_i\}$  have the same weight  $= 1/n$  and the estimator is given by

$$\hat{\mu}_X = \frac{1}{n} \sum_{i=1}^n X_i \quad (19.4.46)$$

The estimator given by Eq. (19.4.46) will be unbiased but will *not be minimum variance*. The variance of the estimator  $\hat{\mu}_X$  can be obtained from Eq. (19.4.39)

$$\text{var}(\hat{\mu}_X) = \sigma_{\mu_X}^2 = \frac{1}{n^2} \mathbf{1}^T \mathbf{C}_X \mathbf{1} \quad (19.4.47)$$

or, in expanded form

$$\text{var}(\hat{\mu}_X) = \sigma_{\hat{\mu}_X}^2 = \frac{1}{n} \left[ C_X(0) + 2 \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) C_X(h) \right] \quad (19.4.48)$$

The variance given by Eq. (19.4.48) is not minimum. For independent  $X_i$ s,  $C_X(h) = 0$ ,  $h \neq 0$ , and the variance  $\sigma_{\hat{\mu}_X}^2$ , given by

$$\sigma_{\hat{\mu}_X}^2 = \frac{1}{n^2} \mathbf{1}^T \mathbf{C}_X = \frac{n C_X(0)}{n^2} = \frac{\sigma_X^2}{n}$$

is a minimum, a result that has been derived previously [in Eq. (18.3.11)].

**Example 19.4.3** The autocovariance function of a random process  $X(t)$  is  $C_X(h) = e^{-\alpha|h|}$ , with  $\alpha = 0.5$  as in Example 19.4.1. The number of data points  $n = 50$ . We will find the variance of the estimated value  $\hat{\mu}_x$ . From Eq. (19.4.48) we can write the following equation for the estimated variance:

$$\text{var}(\hat{\mu}_X) = \sigma_{\hat{\mu}_X}^2(n) = \frac{1}{n} \left[ 1 + 2 \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) e^{-\alpha h} \right]$$

A closed-form solution for the summation given above can be obtained as

$$\sigma_{\hat{\mu}_X}^2(n) = \frac{1}{n^2} \cdot \frac{(n+2)e^{2\alpha} + (n-2)e^{\alpha} - ne^{3\alpha} - 2e^{-\alpha(n-2)} + 2e^{-\alpha(n-1)} - n}{(1-e^\alpha)(1-2e^\alpha+e^{2\alpha})}$$

and substituting  $\alpha = 0.5$  in this equation, we obtain

$$\sigma_{\hat{\mu}_X}^2(n) = \frac{1}{n^2} [7.3258e^{-(n/2)}(e - \sqrt{e}) + 4.083n - 7.83541]$$

and this is graphed in Fig. 19.4.7 along with the value  $C_X(0)/n$  as the variance of the estimated mean for independent observations. The variance for  $n = 50$  is  $\sigma_{\hat{\mu}_X}^2(50) = 0.0785$ . We can see from the figure that the estimated mean  $\hat{\mu}_X$  is asymptotically unbiased since the variance  $\sigma_{\hat{\mu}_X}^2(n) \rightarrow 0$  as  $n \rightarrow \infty$ .

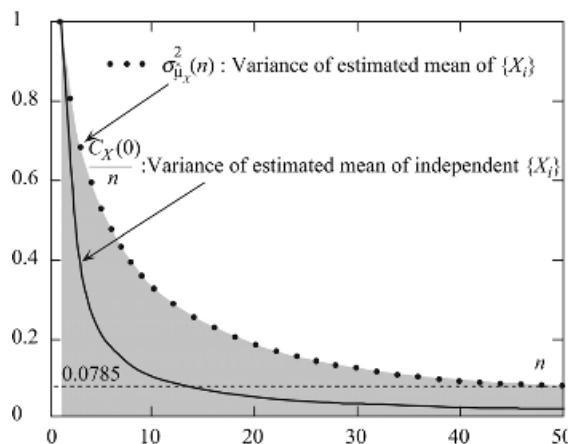


FIGURE 19.4.7

### Estimation of the Autocovariance

The autocovariance of a stationary ergodic random process  $X(t)$  has already been defined [Eq. (19.2.15)] as

$$C_X(h) = E\{[X(t) - \mu_X][X(t + h) - \mu_X]\} \quad (19.4.49)$$

where  $\tau$  has been replaced by  $h$  for convenience. The estimator  $C_X(h)$  for the ACF from the observed random variables set  $\{X_1, \dots, X_n\}$  will be of the form

$$C_X(h) = \sum_{i=1}^{n-h} a_i a_{i+h} (X_i - \hat{\mu}_X)(X_{i+h} - \hat{\mu}_X) \quad (19.4.50)$$

where  $\hat{\mu}_X$  is the unbiased estimator of the mean from Eq. (19.4.46). We have to choose, if possible, the coefficients  $\{a_i, i = 1, \dots, n\}$  such that  $E[C_X(h)] = C_X(h)$ . Equation (19.4.50) can be rewritten as follows:

$$C_X(h) = \sum_{i=1}^{n-h} a_i a_{i+h} [(X_i - \mu_X) - (\hat{\mu}_X - \mu_X)][(X_{i+h} - \mu_X) - (\hat{\mu}_X - \mu_X)]$$

Expanding this equation and taking expectations, we obtain

$$\begin{aligned} E[C_X(h)] &= \sum_{i=1}^{n-h} a_i a_{i+h} E \left[ \begin{aligned} &(X_i - \mu_X)(X_{i+h} - \mu_X) + (\hat{\mu}_X - \mu_X)^2 \\ &-(X_i - \mu_X)(\hat{\mu}_X - \mu_X) - (X_{i+h} - \mu_X)(\hat{\mu}_X - \mu_X) \end{aligned} \right] \\ &= \sum_{i=1}^{n-h} a_i a_{i+h} \{ C_X(h) + \sigma_{\hat{\mu}_X}^2 \} - \sum_{i=1}^{n-h} a_i a_{i+h} E[(X_i - \mu_X)(\hat{\mu}_X - \mu_X)] \\ &\quad - \sum_{i=1}^{n-h} a_i a_{i+h} E[(X_{i+h} - \mu_X)(\hat{\mu}_X - \mu_X)] \\ &= \left[ \sum_{i=1}^{n-h} a_i a_{i+h} (C_X(h) + \sigma_{\hat{\mu}_X}^2) \right] - (S_1 + S_2) \end{aligned} \quad (19.4.51)$$

where the summations  $S_1$  and  $S_2$  have been defined as follows:

$$\begin{aligned} S_1 &= \sum_{i=1}^{n-h} a_i a_{i+h} E[(X_i - \mu_X)(\hat{\mu}_X - \mu_X)] \\ S_2 &= \sum_{i=1}^{n-h} a_i a_{i+h} E[(X_{i+h} - \mu_X)(\hat{\mu}_X - \mu_X)] \end{aligned} \quad (19.4.52)$$

Substituting for  $\hat{\mu}_X$  from Eq. (19.4.46) in the summation  $S_1$  in Eq. (19.4.52), we can write

$$S_1 = \frac{1}{n} \sum_{i=1}^{n-h} \sum_{j=1}^n a_i a_{i+h} E[(X_i - \mu_X)(X_j - \mu_X)] = \frac{1}{n} \sum_{i=1}^{n-h} \sum_{j=1}^n a_i a_{i+h} C_X(j - i) \quad (19.4.53)$$

This equation represents the sum of every term of the  $(n - h) \times n$  matrix shown in Fig. 19.4.8.

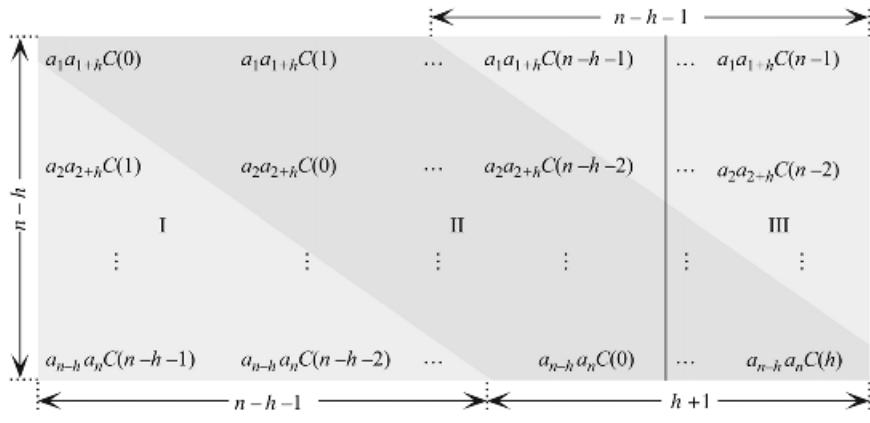


FIGURE 19.4.8

The summation is carried out in three stages. The shaded parallelogram in region II consists of the sum of  $(n - h) \times (h + 1)$  terms, and the shaded triangles in regions I and III consists of the remaining  $(n - h) \times (n - h - 1)$  terms. Equation (19.4.53) can be rewritten in terms of these summations as follows:

$$S_1 = \frac{1}{n} \left\{ \sum_{j=0}^h \sum_{i=1}^{n-h} a_i a_{i+h} C_X(j) + \sum_{j=1}^{n-h-1} \sum_{i=j+1}^{n-h} a_i a_{i+h} C_X(j) + \sum_{j=1}^{n-h-1} \sum_{i=1}^{n-h-j} a_i a_{i+h} C_X(j+h) \right\} \quad (19.4.54)$$

An equation similar to Eq. (19.4.53) for  $S_2$  is

$$\begin{aligned} S_2 &= \frac{1}{n} \sum_{i=1}^{n-h} \sum_{j=1}^n a_i a_{i+h} E[(X_{i+h} - \mu_X)(X_j - \mu_X)] \\ &= \frac{1}{n} \sum_{i=1}^{n-h} \sum_{j=1}^n a_i a_{i+h} C_X(j - i - h) \end{aligned} \quad (19.4.55)$$

A summation matrix similar to that in Fig. 19.4.8 is shown in Fig. 19.4.9.

From Fig. 19.4.9 we can write the summation in much the same way as we did in Eq. (19.4.54):

$$S_2 = \frac{1}{n} \left\{ \sum_{j=0}^h \sum_{i=1}^{n-h} a_i a_{i+h} C_X(j) + \sum_{j=1}^{n-h-1} \sum_{i=1}^{n-h-j} a_i a_{i+h} C_X(j) + \sum_{j=1}^{n-h-1} \sum_{i=j+1}^{n-h} a_i a_{i+h} C_X(j+h) \right\} \quad (19.4.56)$$

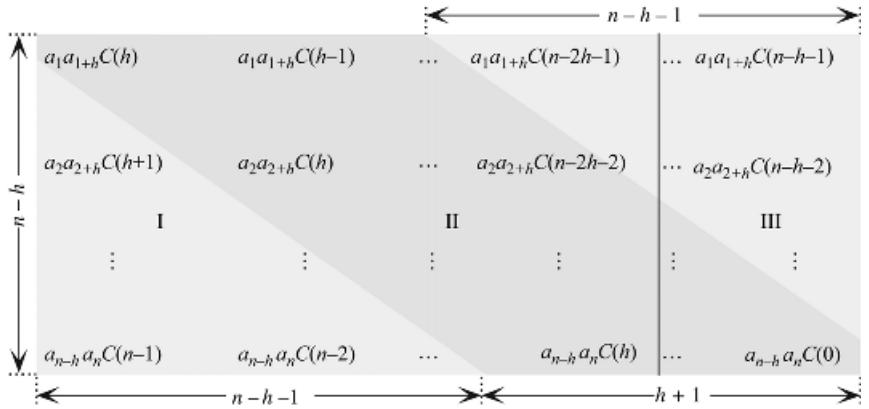


FIGURE 19.4.9

Substituting Eqs. (19.4.55) and (19.4.56) in Eq. (19.4.51), we can write  $E[C_X(h)]$  as

$$E[C_X(h)] = \left[ \sum_{i=1}^{n-h} a_i a_{i+h} \left( C_X(h) + \sigma_{\hat{\mu}_X}^2 - \frac{2}{n} \sum_{j=0}^h C_X(j) \right) \right] - \frac{1}{n} \left[ \sum_{i=1}^{n-h-j} (a_i a_{i+h} + a_{i+j} a_{i+h+j}) \right] \left[ \sum_{j=1}^{n-h-1} C_X(j) + C_X(h+j) \right] \quad (19.4.57)$$

where the variance of the sample mean  $\hat{\mu}_X$  is as follows:

$$\sigma_{\hat{\mu}_X}^2 = \frac{1}{n} \left[ C_X(0) + 2 \sum_{h=1}^{n-1} \left( 1 - \frac{h}{n} \right) C_X(h) \right] \quad (19.4.48)$$

The analysis becomes intractable finding the coefficients  $\{a_i\}$  such that  $E[C_X(h)] = C_X(h)$ . In the spirit of Eqs. (19.4.5) and (19.4.6), two estimators for the ACF  $C_X(h)$  can be defined by substituting  $a_i a_{i+h} = [1/(n-h)]$  or  $a_i a_{i+h} = (1/n)$  in Eq. (19.4.50):

$$\bar{C}_X(h) = \frac{1}{n-h} \sum_{i=1}^{n-h} (X_i - \hat{\mu}_X)(X_{i+h} - \hat{\mu}_X) \quad (19.4.58)$$

$$\hat{C}_X(h) = \frac{1}{n} \sum_{i=1}^{n-h} (X_i - \hat{\mu}_X)(X_{i+h} - \hat{\mu}_X) \quad (19.4.59)$$

The expected value  $E[\bar{C}_X(h)]$  is obtained by substituting  $a_i a_{i+h} = [1/(n-h)]$  in Eq. (19.4.57)

$$E[\bar{C}_X(h)] = \left[ C_X(h) + \sigma_{\hat{\mu}_X}^2 - \frac{2}{n} \sum_{j=0}^h C_X(j) \right] - \frac{2}{n(n-h)} \left[ \sum_{j=1}^{n-h-1} (n-h-j)[C_X(j) + C_X(h+j)] \right] \quad (19.4.60)$$

and the expected value  $E[\hat{C}_X(h)]$  is obtained by substituting  $a_i a_{i+h} = (1/n)$  in Eq. (19.4.57)

$$E[\hat{C}_X(h)] = \left[ \left(1 - \frac{h}{n}\right) \left( C_X(h) + \sigma_{\hat{\mu}_X}^2 - \frac{2}{n} \sum_{j=0}^h C_X(j) \right) \right] - \frac{2}{n^2} \left[ \sum_{j=1}^{n-h-1} (n-h-j)[C_X(j) + C_X(h+j)] \right] \quad (19.4.61)$$

where  $\sigma_{\hat{\mu}_X}^2$  in these equations is as given by Eq. (19.4.48):

$$\sigma_{\hat{\mu}_X}^2 = \frac{1}{n} \left[ C_X(0) + 2 \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) C_X(h) \right] \quad (19.4.48)$$

**Example 19.4.4** We will follow Example 19.4.3 and find  $E[\bar{C}_X(h)]$  and  $E[\hat{C}_X(h)]$  for the random process whose ACF  $C_X(h) = e^{-\alpha|h|}$  with  $\alpha = 0.5$  and the number of data points  $n = 50$ . In Example 19.4.3, we have calculated  $\sigma_{\hat{\mu}_X}^2$  for 50 data points as  $\sigma_{\hat{\mu}_X}^2(50) = 0.0785$ .  $E[\bar{C}_X(h)]$  for  $n = 50$  is obtained from Eq. (19.4.60) as

$$E[\bar{C}_X(h)] = \left[ e^{-\alpha h} + \sigma_{\hat{\mu}_X}^2(50) - \frac{2}{50} \sum_{j=0}^h e^{-\alpha j} \right] - \frac{2}{50(50-h)} \left[ \sum_{j=1}^{49-h} (50-h-j)[e^{-\alpha j} + e^{\alpha(h+j)}] \right] \quad \alpha = 0.5$$

Similarly,  $E[\hat{C}_X(h)]$  is obtained from Eq. (19.4.61):

$$E[\hat{C}_X(h)] = \left[ 1 - \frac{h}{50} \right] \left[ e^{-\alpha h} + \sigma_{\hat{\mu}_X}^2(50) - \frac{2}{50} \sum_{j=0}^h e^{-\alpha j} \right] - \frac{2}{50^2} \left[ \sum_{j=1}^{49-h} (50-h-j)[e^{-\alpha j} + e^{-\alpha(h+j)}] \right] \quad \alpha = 0.5$$

$E[\bar{C}_X(h)]$ ,  $E[\hat{C}_X(h)]$  and the true value  $C_X(h)$  are shown in Fig. 19.4.10 for lags  $h = 10$  (10% of data points). It is seen that both estimators are biased but there is very little difference between them for the 10 lags shown here.

### Variance of Autocovariance Estimators

Finding the variances of  $\bar{C}_X(h)$  and  $\hat{C}_X(h)$  is extremely tedious and is not worth the effort, particularly when the number of data points  $n$  is large. Hence the sample mean  $\hat{\mu}_X$  can be replaced by the true mean  $\mu_X$  in Eqs. (19.4.58) and (19.4.59). The unbiased estimator corresponding to Eq. (19.4.58) is given by

$$\bar{C}_X(h) = \frac{1}{n-h} \sum_{i=1}^{n-h} (X_i - \mu_X)(X_{i+h} - \mu_X) \quad (19.4.62)$$

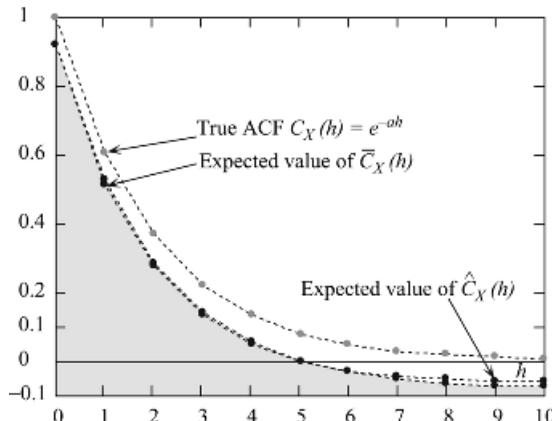


FIGURE 19.4.10

and  $E[\bar{C}_X(h)] = C_X(h)$ . The biased estimator corresponding to Eq. (19.4.59) is

$$\bar{C}_X(h) = \frac{1}{n} \sum_{i=1}^{n-h} (X_i - \mu_X)(X_{i+h} - \mu_X) \quad (19.4.63)$$

As a first step toward finding the variance  $\bar{C}_X(h)$ , we will define a random variable  $Z_i$  as  $Z_i = (X_i - \mu_X)(X_{i+h} - \mu_X)$ , and Eq. (19.4.62) becomes

$$\bar{C}_X(h) = \frac{1}{n-h} \sum_{i=1}^{n-h} Z_i \quad (19.4.64)$$

From Eq. (19.4.48) the variance of  $\bar{C}_X(h)$  can be written as follows:

$$\text{var}[\bar{C}_X(h)] = \frac{1}{n-h} \left[ C_Z(0) + 2 \sum_{g=1}^{n-h-1} \left( 1 - \frac{g}{n-h} \right) C_Z(g) \right] \quad (19.4.65)$$

The mean value of  $Z_i$  is  $C_X(h)$  for all  $i$  and the autocovariance of  $Z_i$  can be defined by

$$C_Z(g) = E[(Z_i - C_X(h))(Z_{i+g} - C_X(h))] = E[Z_i Z_{i+g}] - C_X^2(h) \quad (19.4.66)$$

where

$$E[Z_i Z_{i+g}] = E[(X_i - \mu_X)(X_{i+h} - \mu_X)(X_{i+g} - \mu_X)(X_{i+g+h} - \mu_X)] \quad (19.4.67)$$

Since  $E(X_k - \mu_X) = 0$  for all  $k$ , we can write  $E[Z_i Z_{i+g}]$  in terms of the fourth cumulant  $\kappa_4$  from Eq. (11.6.8):

$$E[Z_i Z_{i+g}] = \kappa_4(h, g) + C_X^2(h) + C_X^2(g) + C_X(g+h)C_X(g-h) \quad (19.4.68)$$

and substituting this equation in Eq. (19.4.66), we obtain

$$C_Z(g) = \kappa_4(h, g) + C_X^2(g) + C_X(g+h)C_X(g-h) \quad (19.4.69)$$

$$C_Z(0) = \kappa_4(h, 0) + C_X^2(0) + C_X^2(h) \quad (19.4.70)$$

Substituting Eqs. (19.4.69) and (19.4.70) in Eq. (19.4.65), we obtain

$$\text{var}[\bar{C}_X(h)]$$

$$= \frac{1}{n-h} \left\{ \begin{array}{l} \kappa_4(h,0) + C_X^2(0) + C_X^2(h) \\ + 2 \sum_{g=1}^{n-h-1} \left(1 - \frac{g}{n-h}\right) [\kappa_4(h,g) + C_X^2(g) + C_X(g+h)C_X(g-h)] \end{array} \right\} \quad (19.4.71)$$

For a Gaussian process the fourth cumulant  $\kappa_4$  is zero, and hence Eq. (19.4.71) becomes

$$\text{var}[\bar{C}_X(h)] = \frac{1}{n-h} \left\{ \begin{array}{l} C_X^2(0) + C_X^2(h) \\ + 2 \sum_{g=1}^{n-h-1} \left(1 - \frac{g}{n-h}\right) [C_X^2(g) + C_X(g+h)C_X(g-h)] \end{array} \right\} \quad (19.4.72)$$

Or, more compactly

$$\text{var}[\bar{C}_X(h)] = \frac{1}{(n-h)^2} \left\{ \sum_{g=-(n-h-1)}^{n-h-1} (n-h-|g|) [C_X^2(g) + C_X(g+h)C_X(g-h)] \right\}, \quad h < n \quad (19.4.73)$$

Equation (19.4.73) is the discrete counterpart of the continuous process of Eq. (19.4.24) and represents the variance of an *unbiased* estimator  $\bar{C}_X(h)$  of a Gaussian process with known mean value  $\mu_X$ .

In a similar manner the variance of the biased estimator  $\hat{C}_X(h)$  of a Gaussian process with known mean value  $\mu_X$  given by Eq. (19.4.63) can be obtained as follows:

$$\text{var}[\hat{C}_X(h)] = \frac{n-h}{n^2} \left\{ \begin{array}{l} C_X^2(0) + C_X^2(h) \\ + 2 \sum_{g=1}^{n-h-1} \left(1 - \frac{g}{n-h}\right) [C_X^2(g) + C_X(g+h)C_X(g-h)] \end{array} \right\} \quad (19.4.74)$$

**Example 19.4.5** We will continue the same way as in Example 19.4.4 and find the variances of the unbiased estimator  $\bar{C}_X(h)$  and the biased estimator  $\hat{C}_X(h)$ . We are given  $C_X(h) = e^{-\alpha|h|}$  with  $\alpha = 0.5$  and  $n = 50$ . From Eq. (19.4.73),  $\text{var}[\bar{C}_X(h)]$  is given by

$$\text{var}[\bar{C}_X(h)] = \frac{1}{n-h} \left\{ \begin{array}{l} 1 + e^{-2\alpha h} \\ + 2 \sum_{g=1}^{n-h-1} \left(1 - \frac{g}{n-h}\right) [e^{-2\alpha g} + e^{-\alpha(|g+h|)}e^{-\alpha(|g-h|)}] \end{array} \right\}$$

As in Example 19.4.2, the absolute value signs in the product term  $e^{-\alpha|g+h|}e^{-\alpha|g-h|}$  have to be carefully removed. Following the procedure in that example, we can write

$$0 \leq h \leq \frac{n-1}{2}:$$

$$e^{-\alpha|g+h|}e^{-\alpha|g-h|} = \begin{cases} e^{-\alpha(g+h)}e^{\alpha(g-h)} = e^{-2\alpha h}, & 0 \leq g \leq h \\ e^{-\alpha(g+h)}e^{-\alpha(g-h)} = e^{-2\alpha g}, & h < g \leq n-h-1 \end{cases}$$

$$\frac{n-1}{2} < h \leq n:$$

$$e^{-\alpha|g+h|}e^{-\alpha|g-h|} = e^{-\alpha(g+h)}e^{\alpha(g-h)} = e^{-2\alpha h}, \quad 0 \leq g \leq n-h-1$$

The variance equation  $\bar{C}_X(h)$ , after substituting the equations above, is

$$\text{var}[\bar{C}_X(h)]$$

$$\begin{aligned} & 1 + e^{-2\alpha h} + 2 \sum_{g=1}^h \left(1 - \frac{g}{n-h}\right) [e^{-2\alpha g} + e^{-2\alpha h}] \\ & = \frac{1}{n-h} \left\{ \begin{aligned} & +2 \sum_{g=h+1}^{n-h-1} \left(1 - \frac{g}{n-h}\right) [e^{-2\alpha g} + e^{-2\alpha g}], \quad 0 \leq h \leq \frac{n-1}{2} \\ & 2 \sum_{g=1}^{n-h-1} \left(1 - \frac{g}{n-h}\right) [e^{-2\alpha g} + e^{-2\alpha h}], \quad \frac{n-1}{2} < h \leq n \end{aligned} \right\} \end{aligned}$$

The variance of the biased estimator  $\hat{C}_X(h)$  is calculated by substituting  $(n-h)/n^2$  for  $1/(n-h)$  in the preceding equation. Both variances are shown in Fig. 19.4.11.

The result is very similar to the one for the continuous case given in Fig. 19.4.6, and the biased estimator gives a better minimum mean-square error.

#### Estimation of Autocorrelation Functions

The two estimators for the autocorrelation function are obtained by substituting  $\hat{\mu} = 0$  in the two estimators for the autocovariance function given by Eqs. (19.4.58) and (19.4.59):

$$\bar{R}_X(h) = \frac{1}{n-h} \sum_{i=1}^{n-h} X_i X_{i+h} \quad (19.4.75a)$$

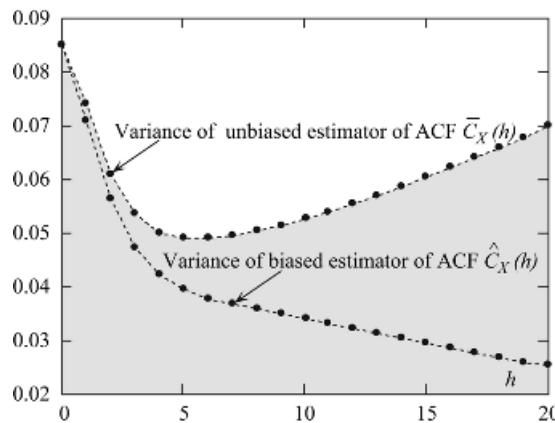


FIGURE 19.4.11

$\bar{R}_X(h)$  is an unbiased estimator since

$$E[\bar{R}_X(h)] = \frac{1}{n-h} \sum_{i=1}^{n-h} E[X_i X_{i+h}] = \frac{1}{n-h} \sum_{i=1}^{n-h} R_X(h) = R_X(h) \quad (19.4.75b)$$

$$\hat{R}_X(h) = \frac{1}{n} \sum_{i=1}^{n-h} X_i X_{i+h} \quad (19.4.76a)$$

$\hat{R}_X(h)$  is biased estimator since

$$E[\hat{R}_X(h)] = \frac{n-h}{n} \frac{1}{n-h} \sum_{i=1}^{n-h} E[X_i X_{i+h}] = \left[1 - \frac{1}{n}\right] R_X(h) \quad (19.4.76b)$$

with the bias  $B = \{[R_X(h)]/n\}$ .

Further analyses of these estimators are the same as those of the ACF estimators  $\bar{C}_X(h)$  and  $\hat{C}_X(h)$  with the estimated mean  $\hat{\mu}_X$  set equal to 0.

### **Estimation of Normalized Autocovariance Functions (NACFs)**

**Estimation of Mean** As mentioned earlier, the normalized autocovariance  $\rho_X(h)$  (NACF) plays a key role in modeling time series. The NACF for a stationary random process for discrete time can be defined by an equation analogous to Eq. (19.2.16) as follows:

$$\rho_X(h) = \frac{C_X(h)}{C_X(0)} \quad \text{and} \quad C_X(h) = \rho_X(h)C_X(0) \quad (19.4.77)$$

The unbiased estimator  $\bar{C}_X(h)$  for large  $n$  is defined in Eqs. (19.4.62). The estimated NACF  $\bar{\rho}_X(h)$  for large sample  $n$  can be defined as

$$\bar{\rho}_X(h) = \frac{\bar{C}_X(h)}{\bar{C}_X(0)} \quad (19.4.78)$$

resulting in a ratio of random variables, and the analysis becomes rather involved. We will now determine the mean of  $\bar{\rho}_X(h)$ . Since  $\bar{C}_X(h)$  is an unbiased estimator, we can define its variations about the mean value  $C_X(h)$  as  $\bar{C}_X^0(h)$  with  $E[\bar{C}_X^0(h)] = 0$  and write  $\bar{C}_X(h) = C_X(h) + \bar{C}_X^0(h)$ :

$$\begin{aligned} E[\bar{\rho}_X(h)] &= E\left[\frac{\bar{C}_X(h)}{\bar{C}_X(0)}\right] = E\left[\frac{C_X(h) + \bar{C}_X^0(h)}{C_X(0) + \bar{C}_X^0(0)}\right] \\ &= E\left\{\frac{[C_X(h) + \bar{C}_X^0(h)]\left[1 + \frac{\bar{C}_X^0(0)}{C_X(0)}\right]^{-1}}{C_X(0)}\right\} \end{aligned} \quad (19.4.79)$$

Assuming that the variations  $\bar{C}_X^0(h)$  are small, we can expand Eq. (19.4.79) in Taylor series, and neglecting second-order terms, we have

$$E[\bar{\rho}_X(h)] \approx \frac{1}{C_X(0)} E \left[ C_X(h) + \bar{C}_X^0(h) - \frac{C_X(h)\bar{C}_X^0(0)}{C_X(0)} \right] = E \left[ \frac{C_X(h)}{C_X(0)} \right] = \rho_X(h) \quad (19.4.80)$$

Thus, we have an unbiased estimator under the assumption of small variations in  $\bar{C}_X^0(h)$ .

**Estimation of Covariance of NACF** Finding the covariance between adjacent values of the estimated NACF  $\bar{\rho}_X(h)$  is very involved and is detailed elsewhere in the literature [8,24,31]. The final result for a Gaussian process  $X(t)$  is as follows:

$$\begin{aligned} & \text{cov}[\bar{\rho}_X(h), \bar{\rho}_X(h+g)] \\ &= \sum_{k=-\infty}^{\infty} \left\{ \begin{aligned} & \frac{1}{n-h} [\rho_X(k)\rho_X(k+g) + \rho_X(k+h+g)\rho_X(k-h)] \\ & - \frac{2}{n-h-g} \rho_X(h)\rho_X(k)\rho_X(k-h-g) \\ & - \frac{2}{n-h} \rho_X(h+g)\rho_X(k)\rho_X(k-h) \\ & + \frac{2}{n} \rho_X^2(k)\rho_X(h)\rho_X(h+g) \end{aligned} \right\} \end{aligned} \quad (19.4.81)$$

The variance of  $\bar{\rho}_X(h)$  is obtained by substituting  $g = 0$  in Eq. (19.4.81). The result is

$$\text{var}[\bar{\rho}_X(h)] = \sum_{k=-\infty}^{\infty} \left\{ \begin{aligned} & \frac{1}{n-h} [\rho_X^2(k) + \rho_X(k+h)\rho_X(k-h)] \\ & - \frac{4}{n-h} \rho_X(h)\rho_X(k)\rho_X(k-h) + \frac{2}{n} \rho_X^2(k)\rho_X^2(h) \end{aligned} \right\} \quad (19.4.82)$$

### Large-Lag Approximations

For most stationary random processes the NACF becomes negligibly small after some lag  $q$ , that is,  $\rho_X(h) \approx 0$ ,  $|h| > q$ . Hence, under the assumption  $\rho_X(h) = 0$  for  $|h| > q$  in Eq. (19.4.81), we can write the large-lag covariance as

$$\text{cov}[\bar{\rho}_X(h), \bar{\rho}_X(h+g)] \approx \frac{1}{n-h} \sum_{k=-q}^q \rho_X(k)\rho_X(k+g), \quad |h| > q \quad (19.4.83)$$

Similarly, the large-lag variance after substituting  $\rho_X(h)=0$  for  $|h| > q$  in Eq. (19.4.82) is

$$\text{var}[\bar{\rho}_X(h)] \approx \frac{1}{n-h} \sum_{k=-q}^q \rho_X^2(k), \quad |h| > q \quad (19.4.84)$$

If the estimator for ACF is  $\hat{C}_X(h)$ , then the divisor  $(n - h)$  in Eqs. (19.4.83) and (19.4.84) is replaced by the divisor  $n$ . Thus the variance equation [Eq. (19.4.84)] becomes

$$\text{var}[\hat{\rho}_X(h)] \approx \frac{1}{n} \sum_{k=-q}^q \rho_X^2(k), \quad |h| > q \quad (19.4.85)$$

It has been shown [3,25] that if the true NACF  $\rho_X(h)$  is zero for  $|h| > q$ , then the estimated NACF  $\bar{\rho}_X(h)$  is normally distributed with zero mean and variance given by Eq. (19.4.84). This property is called *Anderson's theorem*.

### Discrete White-Noise Process

A zero mean stationary discrete white-noise process  $\{X_i\}$  has the following property:  $E[X_i X_j] = \delta_{j-i}$  where the Krönecker delta  $\delta_{j-i} = 1$ ,  $j = i$  and  $\delta_{j-i} = 0$ ,  $j \neq i$ , an equation analogous to Eq. (19.2.17) for continuous white noise. As a consequence, the NACF  $\rho_X(h)$  has the property

$$\rho_X(h) = \begin{cases} 1, & h = 0 \\ 0, & h \neq 0 \end{cases} \quad (19.4.86)$$

Hence the variance of the estimated  $\bar{\rho}_X(h)$  for white noise is obtained by substituting  $\rho_X(k) = 0$  for  $k \neq 0$  in Eq. (19.4.84) to yield

$$\text{var}[\bar{\rho}_X(h)] \approx \frac{1}{n-h} \quad (19.4.87)$$

According to Anderson's theorem, the estimated NACF  $\bar{\rho}_X(h)$  of a stationary discrete white-noise process is Gaussian-distributed with zero mean and variance  $1/(n-h)$ . As a result, we can construct a test for white noise from knowledge of  $\bar{\rho}_X(h)$ . Since the variance is  $1/(n-h)$ , the 95% confidence interval ( $\pm 2\sigma$ ) is  $\pm 2/\sqrt{n-h}$ .

If the estimated NACF is given by

$$\hat{\rho}_X(h) = \frac{\hat{C}_X(h)}{\hat{C}_X(0)} \quad (19.4.88)$$

then, from Eq. (19.4.85), the variance of the estimator  $\hat{\rho}_X(h)$  is

$$\text{var}[\hat{\rho}_X(h)] \approx \frac{1}{n} \quad (19.4.89)$$

and the corresponding confidence interval is  $\pm \frac{2}{\sqrt{n}}$ .

**Example 19.4.6** We have generated 100 points of a zero mean unit variance Gaussian white noise  $\{X_i\}$  by the computer. The plot is shown in Fig. 19.4.12.

The estimated NACFs  $\bar{\rho}_X(h)$  and  $\hat{\rho}_X(h)$  of this sequence were determined from Eqs. (19.4.78) and (19.4.88) and plotted in Fig. 19.4.13 along with their confidence intervals for 50 lags.

In the plot of the estimator  $\bar{\rho}_X(h)$ , only two points out of 50 (4%) exceed the confidence limits, whereas in the estimator  $\hat{\rho}_X(h)$ , no points exceed the confidence limits. Both estimators pass the white-noise test at the 5% significance level comfortably, and hence we conclude that the process  $\{X_i\}$  is indeed a white-noise sequence.

**Example 19.4.7** We will generate 100 points of the discrete-time process for  $\phi = 0.8$  and  $-0.8$  in Example 19.2.11 with  $\sigma_X^2 = 1$

$$X_i = \phi X_{i-1} + \nu_i, \quad i = 1, \dots, 100$$

and estimate the NACF and compare it to the true NACF.

The computer-generated random sequence is shown in Fig. 19.4.14 for  $\phi = 0.8$  and  $-0.8$ . The figure shows that for  $\phi = -0.8$ , the process is more oscillatory. The

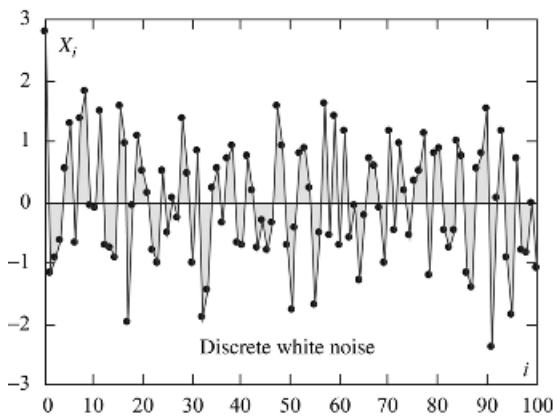


FIGURE 19.4.12

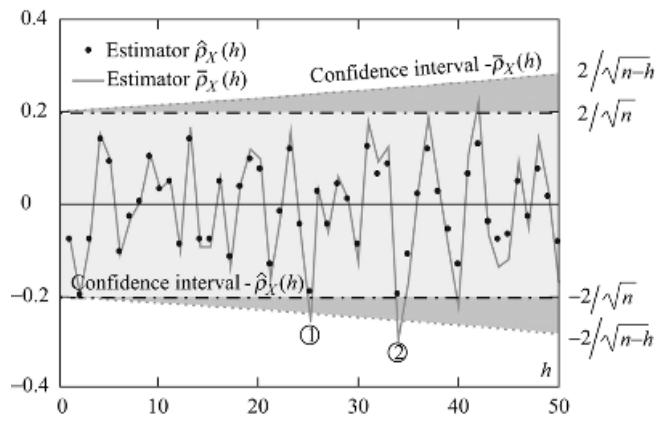


FIGURE 19.4.13

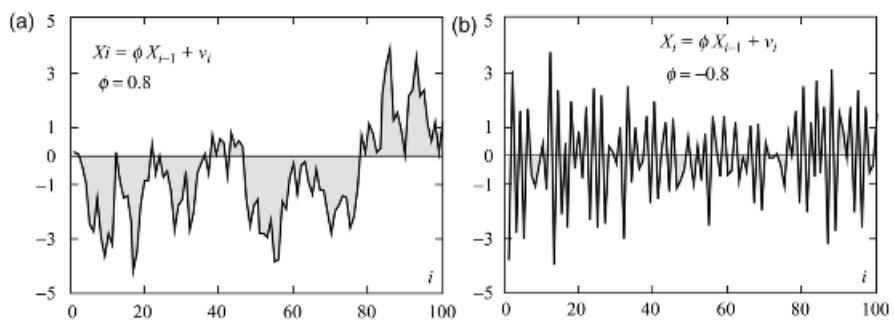


FIGURE 19.4.14

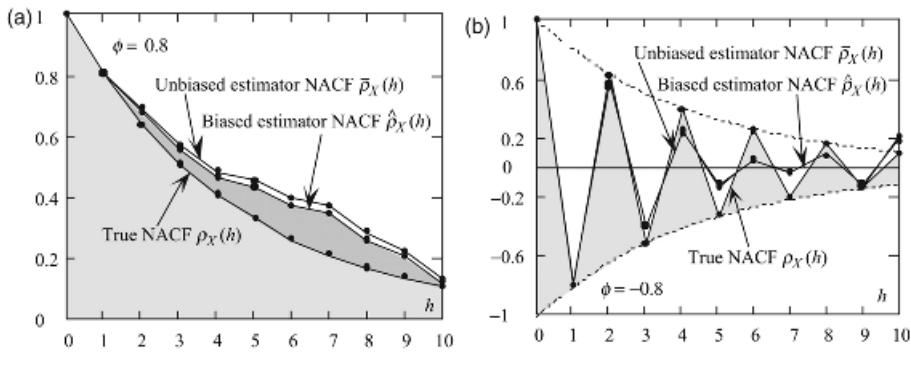


FIGURE 19.4.15

estimated NACFs  $\bar{\rho}_X(h)$  and  $\hat{\rho}_X(h)$  are determined from Eqs. (19.4.78) and (19.4.88) respectively. They are shown in Fig. 19.4.15 for  $\phi = 0.8$  and  $-0.8$  for 10 lags.

Here also the NACFs for  $\phi = -0.8$  show oscillatory behavior. For the 10 lags shown, there is very little difference between the two estimators  $\bar{\rho}_X(h)$  and  $\hat{\rho}_X(h)$ . However, the discrepancy between the true NACF and the estimators for higher lags is to be expected.

## 19.5 POWER SPECTRAL DENSITY

### 19.5.1 Continuous Time

In signal analysis power spectra are associated with Fourier transforms that transform signals from the time domain to the frequency domain. The same concept is also applicable to stationary random processes. The correlation functions represent stationary processes in the time domain. We can transform them to the frequency domain by taking their Fourier transforms. The *power spectral density* (psd) function  $S_X(\omega)$  of a real stationary random process  $X(t)$  is defined as the Fourier transform of the autocorrelation function:

$$S_X(\omega) = \text{FT}[R_X(\tau)] = \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega\tau} d\tau \quad (19.5.1)$$

From the Fourier inversion theorem we can obtain the autocorrelation function from the power spectral density:

$$R_X(\tau) = \text{IFT}[S_X(\omega)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) e^{j\omega\tau} d\omega \quad (19.5.2)$$

Equations (19.5.1) and (19.5.2) are called the *Wiener–Khinchine theorem*.

Since  $R_X(0) = E[X^2(t)]$ , the average power in the random process, we obtain the following from Eq. (19.5.2):

$$R_X(0) = E[X^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) d\omega = \int_{-\infty}^{\infty} S_X(f) df \quad (19.5.3)$$

Thus,  $S_X(f)$  represents the average power per hertz, and hence the term *power spectral density*. Since  $R_X(\tau)$  is an even function, we can rewrite Eq. (19.5.1) as

$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau)[\cos(\omega\tau) + j\sin(\omega\tau)]d\tau = \int_{-\infty}^{\infty} R_X(\tau)\cos(\omega\tau)d\tau \quad (19.5.4)$$

and  $S_X(\omega)$  is also an even function. Hence Eq. (19.5.2) can be rewritten as

$$R_X(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega)\cos(\omega\tau)d\omega \quad (19.5.5)$$

The *cross-spectral density* (csd)  $S_{XY}(\omega)$  of two real stationary random processes  $X(t)$  and  $Y(t)$  is defined as the Fourier transform of the cross-correlation function,  $R_{XY}(\tau)$

$$S_{XY}(\omega) = \text{FT}[R_{XY}(\tau)] = \int_{-\infty}^{\infty} R_{XY}(\tau)e^{-j\omega\tau}d\tau \quad (19.5.6)$$

and the inverse Fourier transform of  $S_{XY}(\omega)$  gives the cross-correlation function:

$$R_{XY}(\tau) = \text{IFT}[S_{XY}(\omega)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{XY}(\omega)e^{j\omega\tau}d\omega \quad (19.5.7)$$

The cross-spectral density  $S_{XY}(\omega)$  will be complex, in general, even when the random processes  $X(t)$  and  $Y(t)$  are real.

The Fourier transforms can also be obtained from the tables in Appendix A.

**Example 19.5.1** The power spectral density (psd) of the random binary wave of Example 19.2.6 is to be determined. The autocorrelation of the random process  $X(t)$  is given by

$$R_X(\tau) = \begin{cases} A^2 \left(1 - \frac{|\tau|}{T}\right), & |\tau| < T \\ 0, & \text{otherwise} \end{cases}$$

The psd  $S_X(\omega)$  given by

$$\begin{aligned} S_X(\omega) &= \int_{-T}^0 A^2 \left(1 + \frac{\tau}{T}\right) e^{-j\omega\tau} d\tau + \int_0^T A^2 \left(1 - \frac{\tau}{T}\right) e^{-j\omega\tau} d\tau \\ &= \frac{A^2}{\omega^2 T} [(1 - e^{j\omega T} + j\omega T) + (1 - e^{-j\omega T} - j\omega T)] \\ &= A^2 T \left[ \frac{\sin(\omega T/2)}{\omega T/2} \right]^2 \end{aligned}$$

is shown in Fig. 19.5.1. As we can see from figure, the psd is an even function.

**Example 19.5.2** We will find the psd from the autocorrelation  $R_X(\tau) = \frac{1}{4}[1 + e^{-2\lambda|\tau|}]$  of the random telegraph wave given in Example 19.2.7. Taking the Fourier transform, the psd

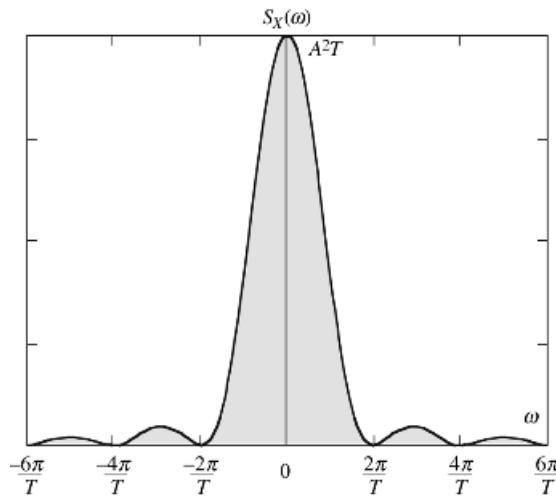


FIGURE 19.5.1

is given by

$$\begin{aligned}
 S_X(\omega) &= \frac{1}{4} \int_{-\infty}^{\infty} e^{-j\omega\tau} d\tau + \frac{1}{4} \int_{-\infty}^0 e^{2\lambda\tau} e^{-j\omega\tau} d\tau + \frac{1}{4} \int_0^{\infty} e^{-2\lambda\tau} e^{-j\omega\tau} d\tau \\
 &= \frac{1}{4} \left[ 2\pi\delta(\omega) + \frac{4\lambda}{4\lambda^2 + \omega^2} \right] = \frac{\pi}{2} \delta(\omega) + \frac{\lambda}{4\lambda^2 + \omega^2}
 \end{aligned}$$

The impulse function  $(\pi/2)\delta(\omega)$  represents the direct-current (dc) value of  $R_X(\tau) = \frac{1}{4}$ . The psd is shown in Fig. 19.5.2 for  $\lambda = 1$ .

**Example 19.5.3** We will find the psd from the autocorrelation  $R_X(\tau) = e^{-2\lambda|\tau|} \cos(\omega_0\tau)$  of Example 19.2.8. We can take the Fourier transform of  $R_X(\tau)$  directly, but it is easier to

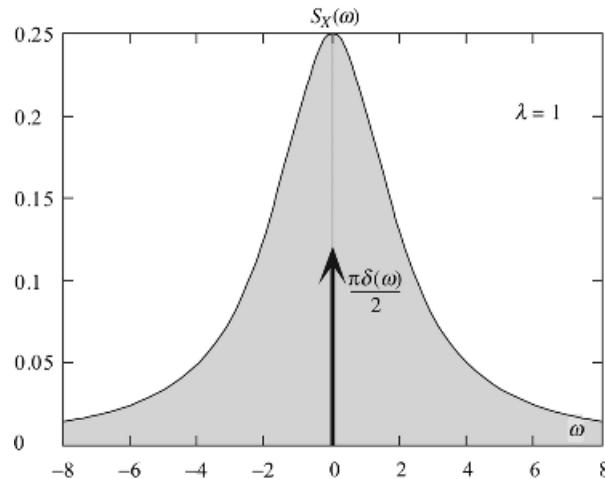


FIGURE 19.5.2

evaluate the FT by using the frequency convolution property of the FT as follows:

$$\text{FT}[e^{-2\lambda|\tau|}] = \frac{4\lambda^2}{4\lambda^2 + \omega^2} \quad \text{and} \quad \text{FT}[\cos(\omega_0\tau)] = \pi[\delta(\omega + \omega_0) + \delta(\omega - \omega_0)]$$

Using the frequency convolution property  $x(t)y(t) \Leftrightarrow (1/2\pi)X(\omega) * Y(\omega)$ , we have

$$e^{-2\lambda|\tau|} \cos(\omega_0\tau) \Leftrightarrow \frac{1}{2\pi} \frac{4\lambda^2}{4\lambda^2 + \omega^2} * \pi[\delta(\omega + \omega_0) + \delta(\omega - \omega_0)]$$

or

$$S_X(\omega) = 2\lambda \left[ \frac{1}{4\lambda^2 + (\omega + \omega_0)^2} + \frac{1}{4\lambda^2 + (\omega - \omega_0)^2} \right]$$

and simplifying, we obtain

$$S_X(\omega) = \frac{4\lambda(\omega^2 + \omega_0^2 + 4\lambda^2)}{\omega^4 - 2(\omega_0^2 - 4\lambda^2)\omega^2 + (\omega_0^2 + 4\lambda^2)^2}$$

With  $\lambda = 1$  and  $\omega_0 = 2\pi$ , the psd becomes

$$S_X(\omega) = \frac{4(\omega^2 + 4\pi^2 + 4\lambda^2)}{\omega^4 - 8(\pi^2 - 1)\omega^2 + 16(\pi^2 + 1)^2}$$

The psd  $S_X(\omega)$  is shown in Fig. 19.5.3.

**Example 19.5.4 (Bandlimited Process)** The autocorrelation function  $R_X(\tau) = \sin(\omega_0\tau)/\pi\tau$  of a stationary random process  $X(t)$  is shown in Fig. 19.5.4 for  $\omega_0 = 2\pi$ . We have to find the psd  $S_X(\omega)$ .

From item 2 in the FT table, the Fourier transform of  $R_X(\tau)$  can be obtained as follows:

$$\frac{\sin(\omega_0\tau)}{\pi\tau} \Leftrightarrow p_{\omega_0}(\omega), \quad \text{hence} \quad S_X(\omega) = p_{\omega_0}(\omega)$$

The psd  $S_X(\omega)$  is shown in Fig. 19.5.5 for  $\omega_0 = 2\pi$ .

The random process that has a psd as in Fig. 19.5.5 is called a *bandlimited signal* since the frequency spectrum exists only between  $-2\pi$  and  $2\pi$ .

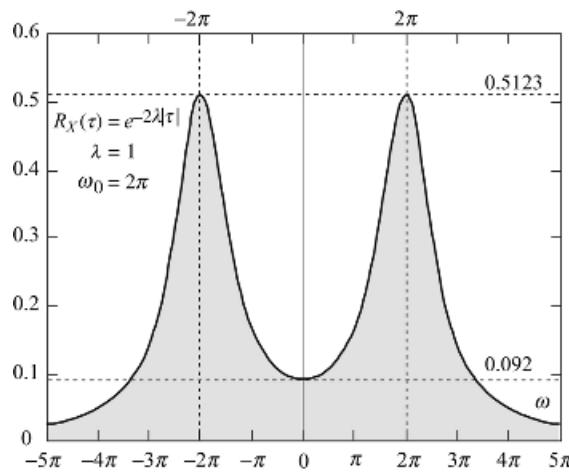


FIGURE 19.5.3

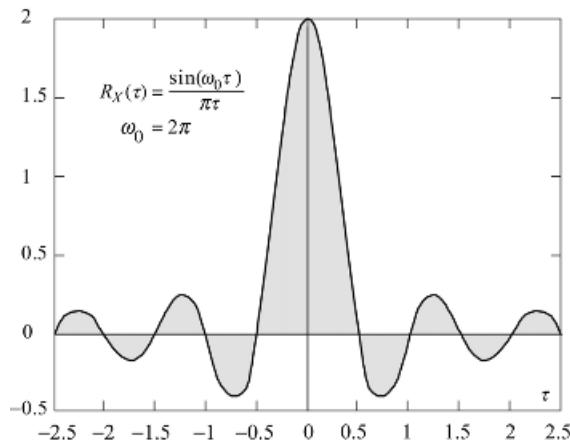


FIGURE 19.5.4

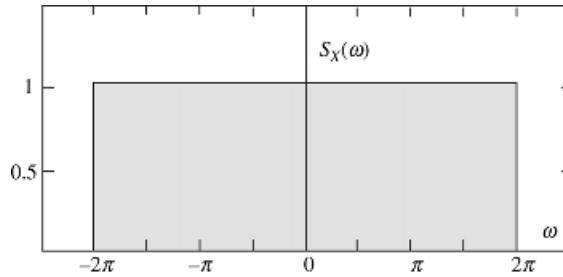


FIGURE 19.5.5

**Example 19.5.5 (Bandlimited Process)** The autocorrelation function  $R_X(\tau)$

$$R_X(\tau) = \frac{\sin[\omega_0(\tau - \tau_0)]}{\pi(\tau - \tau_0)} + \frac{\sin[\omega_0(\tau + \tau_0)]}{\pi(\tau + \tau_0)}$$

of a bandlimited random process  $X(t)$  is shown in Fig. 19.5.6 for  $\omega_0 = 2\pi$  and  $\tau_0 = 3$ .

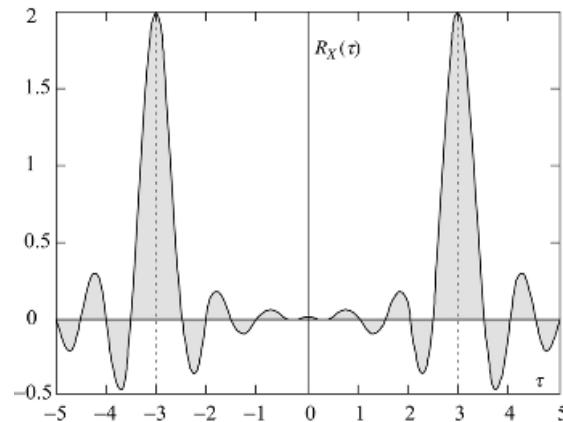


FIGURE 19.5.6

Since  $\sin(\omega_0\tau)/\pi\tau \Leftrightarrow p_{\omega_0}(\omega)$  using the time-shifting property  $x(t \pm t_0) \Leftrightarrow X(\omega)e^{\pm j\omega t_0}$  of the FT, we have

$$\frac{\sin[\omega_0(\tau - \tau_0)]}{\pi(\tau - \tau_0)} \Leftrightarrow p_{\omega_0}(\omega)e^{-j\omega\tau_0}, \quad \frac{\sin[\omega_0(\tau + \tau_0)]}{\pi(\tau + \tau_0)} \Leftrightarrow p_{\omega_0}(\omega)e^{j\omega\tau_0}$$

or

$$\frac{\sin[\omega_0(\tau - \tau_0)]}{\pi(\tau - \tau_0)} + \frac{\sin[\omega_0(\tau + \tau_0)]}{\pi(\tau + \tau_0)} \Leftrightarrow p_{\omega_0}(\omega)[e^{j\omega\tau_0} + e^{-j\omega\tau_0}]$$

and

$$S_X(\omega) = 2p_{\omega_0}(\omega)\cos(\omega\tau_0)$$

The psd is shown in Fig. 19.5.7 for  $\omega_0 = 2\pi$  and  $\tau_0 = 3$ .

**Example 19.5.6 (Bandpass Process)** The autocorrelation function  $R_X(\tau)$

$$R_X(\tau) = \frac{2 \sin(\omega_0\tau)}{\pi\tau} \cos(\omega_c\tau)$$

of a bandpass random process  $X(t)$  is shown in Fig. 19.5.8 for  $\omega_0 = 2\pi$  and  $\omega_c = 16\pi$ . We will find the psd using the frequency convolution property of the FT:

$$x(t)y(t) \Leftrightarrow \frac{1}{2\pi} X(\omega) * Y(\omega)$$

Since  $2 \sin(\omega_0\tau)/\pi\tau \Leftrightarrow 2p_{\omega_0}(\omega)$  and  $\cos(\omega_c\tau) \Leftrightarrow \pi[\delta(\omega + \omega_c) + \delta(\omega - \omega_c)]$ , we have

$$\frac{2 \sin(\omega_0\tau)}{\pi\tau} \cos(\omega_c\tau) \Leftrightarrow \frac{1}{2\pi} 2p_{\omega_0}(\omega) * \pi[\delta(\omega + \omega_c) + \delta(\omega - \omega_c)]$$

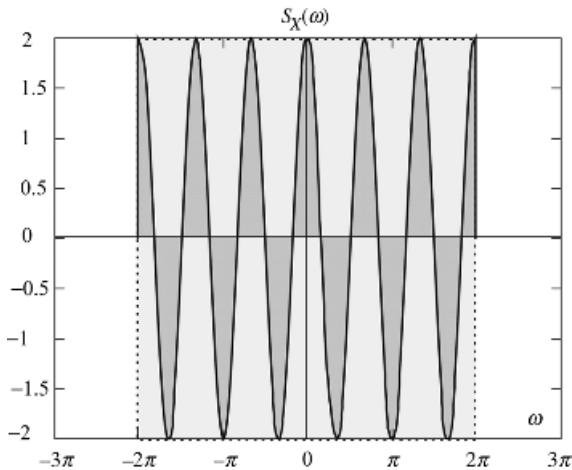


FIGURE 19.5.7

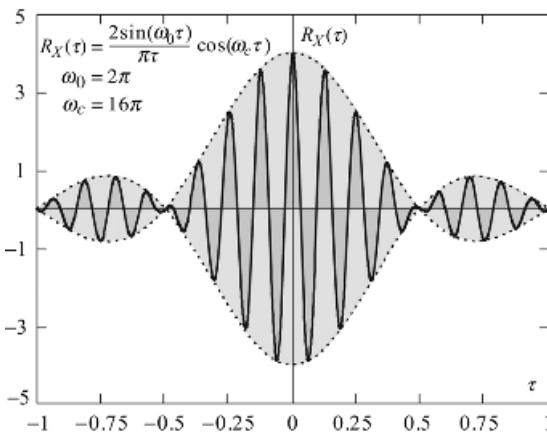


FIGURE 19.5.8

and the psd

$$S_X(\omega) = p_{\omega_0}(\omega + \omega_c) + p_{\omega_0}(\omega - \omega_c)$$

The function  $S_X(\omega)$  is shown in Fig. 19.5.9 for  $\omega_0 = 2\pi$  and  $\omega_c = 16\pi$ .

**Example 19.5.7 (White Noise)** A white-noise process has the autocorrelation function  $R_X(\tau) = \sigma_X^2 \delta(\tau)$  given in Eq. (19.2.17). The psd of white noise is given by

$$S_X(\omega) = \sigma_X^2, \quad -\infty < \omega < \infty$$

and has a flat spectrum. Since the process has all frequencies with equal power, it is called “white noise,” analogously to white light. It has infinite energy since  $R_X(0) = (1/2\pi) \int_{-\infty}^{\infty} \sigma_X^2 d\omega = \infty$ , and hence it is an idealization.

**Example 19.5.8** The cross-correlation function of two random processes  $X(t) = A \cos(\omega_0 t) + B \sin(\omega_0 t)$  and  $Y(t) = -A \sin(\omega_0 t) + B \cos(\omega_0 t)$  in Example 19.2.9 is  $R_{XY}(\tau) = -\sigma^2 \sin(\omega_0 \tau)$ ,  $E[A^2] = E[B^2] = \sigma^2$ . We will find the cross-spectral density  $S_{XY}(\omega)$ .

Taking FT of  $-\sigma^2 \sin(\omega_0 \tau)$ , from tables, we have

$$S_{XY}(\omega) = -j\pi\sigma^2 [\delta(\omega + \omega_0) - \delta(\omega - \omega_0)]$$

and this is shown in Fig. 19.5.10.

Note that  $S_{XY}(\omega)$  is neither even nor real.

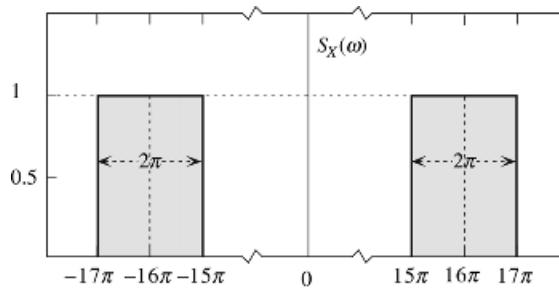


FIGURE 19.5.9

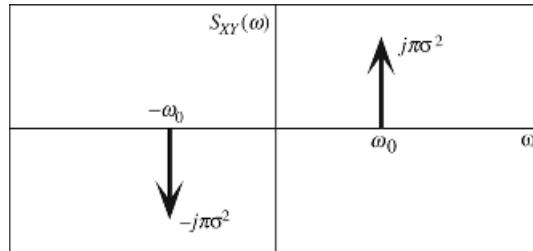


FIGURE 19.5.10

**Example 19.5.9** The cross-spectral density is to be found for the cross-correlation function  $R_{XY}(\tau)$  derived in Example 19.2.10:

$$R_{XY}(\tau) = \frac{1}{2} + \begin{cases} \frac{4}{15} e^{-(\tau/2)} - \frac{1}{6} e^{-2\tau}, & \tau > 0 \\ \frac{1}{10} e^{2\tau}, & \tau \leq 0 \end{cases}$$

Taking FT of each term in  $R_{XY}(\tau)$ , we have

$$\begin{aligned} \frac{4}{15} e^{-(\tau/2)} &\iff \frac{4}{15} \frac{1}{1/2 + j\omega}; \quad \frac{1}{6} e^{-2\tau} \iff \frac{1}{6} \frac{1}{2 + j\omega} \\ \frac{1}{2} &\iff \pi\delta(\omega); \quad \frac{1}{10} e^{2\tau} \iff \frac{1}{10} \frac{1}{2 - j\omega} \end{aligned}$$

The cross-spectral density  $S_{XY}(\omega)$  is obtained by adding the FT terms shown above:

$$S_{XY}(\omega) = \pi\delta(\omega) + \frac{4}{15} \frac{1}{\frac{1}{2} + j\omega} + \frac{1}{6} \frac{1}{2 + j\omega} + \frac{1}{10} \frac{1}{2 - j\omega} = \frac{1}{3} \frac{8 + 2\omega^2 + 3j\omega}{(1 + 2j\omega)(4 + \omega^2)} + \pi\delta(\omega)$$

The cross-spectral density  $S_{XY}(\omega)$  is complex and possesses no symmetry, unlike the power spectral density. The constant term  $\frac{1}{2}$  in  $R_{XY}(\tau)$  gives rise to the impulse function in the frequency domain. The amplitude  $|S_{XY}(\omega)|$  and the phase  $\arg\{S_{XY}(\omega)\}$  spectra are graphed in Fig. 19.5.11.

### Properties of Power Spectral Densities of Stationary Random Processes

1.  $S_X(\omega)$  is a real function. In general, the Fourier transform  $X(\omega)$  of any function  $X(t)$  will be complex. However, the ACF  $R_X(\tau)$  is an even function and satisfies the relation  $R_X(\tau) = R_X(-\tau)$ . Hence, from the definition of psd, we obtain

$$\begin{aligned} S_X(\omega) &= \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega\tau} d\tau = \int_{-\infty}^{\infty} R_X(\tau) [\cos(\omega\tau) + j \sin(\omega\tau)] d\tau \\ &= \int_{-\infty}^{\infty} R_X(\tau) \cos(\omega\tau) d\tau \end{aligned}$$

since the imaginary part  $\int_{-\infty}^{\infty} R_X(\tau) j \sin(\omega\tau) d\tau = 0$  because an even function multiplying an odd function results in an odd function and the integral of an odd function over  $(-\infty, \infty)$  is zero.

2. From property 1 the psd  $S_X(\omega)$  is an even function and, hence it is a function of  $\omega^2$ . As a consequence  $S_X(\omega) = S_X(-\omega)$ . (19.5.8)
3.  $S_X(\omega) \geq 0$ : The psd is a nonnegative function of  $\omega$ . (19.5.9)

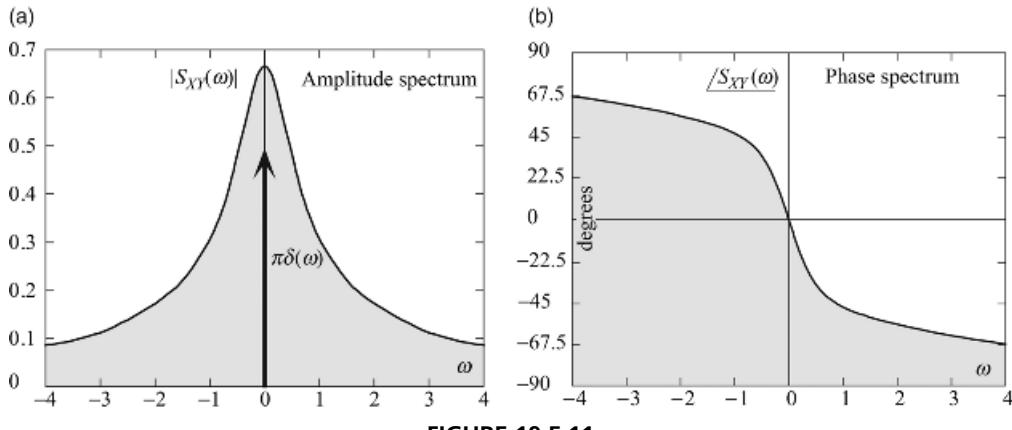


FIGURE 19.5.11

4. From the Fourier transform properties:

$$R_X(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(\omega) d\omega = \int_{-\infty}^{\infty} S_X(f) df = E[X^2(t)] \quad (19.5.10)$$

$$S_X(0) = \int_{-\infty}^{\infty} R_X(\tau) d\tau$$

5. If  $Z(t) = X(t) + Y(t)$ , then from Eq. (19.2.30), we have

$$R_Z(\tau) = R_X(\tau) + R_Y(\tau) + R_{XY}(\tau) + R_{YX}(\tau) \quad (19.5.11)$$

$$S_Z(\omega) = S_X(\omega) + S_Y(\omega) + S_{XY}(\omega) + S_{YX}(\omega)$$

If  $X(t)$  and  $Y(t)$  are orthogonal, then Eq. (19.5.11) reduces to  $S_Z(\omega) = S_X(\omega) + S_Y(\omega)$

#### Alternate Form for Power Spectral Density

The psd can also be obtained directly from the stationary random process  $X(t)$ . We will truncate  $X(t)$  in the interval  $(-T, T)$  and define the random process  $X_T(t)$  as

$$X_T(t) = \begin{cases} X(t), & -T \leq t \leq T \\ 0, & \text{otherwise} \end{cases} \quad (19.5.12)$$

The FT  $X_T(\omega)$  of  $X_T(t)$ , given by  $X_T(\omega) = \int_{-T}^T X(t) e^{-j\omega t} dt$ , is a random variable. The quantity  $S_T(\omega)$ , defined by

$$S_T(\omega) = \frac{1}{2T} \left| \int_{-T}^T X(t) e^{-j\omega t} dt \right|^2 = \frac{|X_T(\omega)|^2}{2T} = \frac{X_T(\omega) X_T^*(\omega)}{2T} \quad (19.5.13)$$

is called the *periodogram*. The periodogram represents the power of the sample function  $X(t)$  at the frequency  $\omega$ . Equation (19.5.13) can be expanded as follows:

$$S_T(\omega) = \frac{1}{2T} \int_{-T}^T \int_{-T}^T X(t) X(s) e^{-j\omega(t-s)} dt ds \quad (19.5.14)$$

We can make the transformation  $\tau = t - s$  and  $u = s$  in Eq. (19.5.14) and perform the integration in the  $(\tau, u)$  plane. The Jacobian  $\| J \|$  of the transformation from Eq. (19.3.5) is 1,

and Eq. (19.5.14) can be transformed as follows:

$$S_T(\omega) = \frac{1}{2T} \int \left[ \int X(u)X(u+|\tau|)du \right] e^{-j\omega\tau} d\tau \quad (19.5.15)$$

We can obtain the limits of integration from Fig. 19.3.1 and write Eq. (19.5.15) as

$$\begin{aligned} S_T(\omega) &= \int_0^{2T} \left[ \frac{1}{2T} \int_{-T+\tau}^T X(u)X(u+\tau)du \right] e^{-j\omega\tau} d\tau \\ &\quad + \int_{-2T}^0 \left[ \frac{1}{2T} \int_{-T}^{T-|\tau|} X(u)X(u+\tau)du \right] e^{-j\omega\tau} d\tau \\ &= \int_0^{2T} \hat{R}_X(\tau) e^{-j\omega\tau} d\tau + \int_{-2T}^0 \hat{R}_X(\tau) e^{-j\omega\tau} d\tau \\ &= \int_{-2T}^{2T} \hat{R}_X(\tau) e^{-j\omega\tau} d\tau \end{aligned} \quad (19.5.16)$$

where we have used the definition for  $\hat{R}_X(\tau)$  as in Eq. (19.4.34a). Using Eq. (19.4.34b), the expected value of  $S_T(\omega)$  can be written as

$$E[S_T(\omega)] = \int_{-2T}^{2T} E[\hat{R}_X(\tau)] e^{-j\omega\tau} d\tau = \int_{-2T}^{2T} R_X(\tau) \left( 1 - \frac{|\tau|}{2T} \right) e^{-j\omega\tau} d\tau \quad (19.5.17)$$

Taking the limit of  $E[S_T(\omega)]$  in Eq. (19.5.17) as  $T \rightarrow \infty$ , we have

$$\begin{aligned} \lim_{T \rightarrow \infty} E[S_T(\omega)] &= \lim_{T \rightarrow \infty} \left\{ \int_{-\infty}^{\infty} R_X(\tau) \left( 1 - \frac{|\tau|}{2T} \right) e^{-j\omega\tau} d\tau \right\} \\ &= \lim_{T \rightarrow \infty} \text{FT} \left[ R_X(\tau) \left( 1 - \frac{|\tau|}{2T} \right) \right] \end{aligned} \quad (19.5.18)$$

Using the frequency convolution property of the FT, we obtain

$$x(t)y(t) \iff \frac{1}{2\pi} X(\omega) * Y(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(p)Y(\omega-p)dp$$

and with  $R_X(\tau) \Leftrightarrow S_X(\omega)$  and

$$\left( 1 - \frac{|\tau|}{2T} \right) \Leftrightarrow 2T \frac{\sin^2 \omega T}{(\omega T)^2}$$

Eq. (19.5.18) can be rewritten as follows:

$$\begin{aligned} \lim_{T \rightarrow \infty} E[S_T(\omega)] &= \lim_{T \rightarrow \infty} \text{FT} \left[ R_X(\tau) \left( 1 - \frac{|\tau|}{2T} \right) \right] \\ &= \lim_{T \rightarrow \infty} \int_{-\infty}^{\infty} S_X(p)T \frac{\sin^2(\omega-p)T}{\pi[(\omega-p)T]^2} dp \end{aligned} \quad (19.5.19)$$

From Ref. 36,

$$\lim_{T \rightarrow \infty} T \frac{\sin^2(\omega T)}{\pi(\omega T)^2} \rightarrow \delta(\omega) \quad \text{and} \quad \lim_{T \rightarrow \infty} T \frac{\sin^2(\omega - p)T}{\pi[(\omega - p)T]^2} = \delta(\omega - p)$$

and substituting this result in Eq. (19.5.19), we have

$$\lim_{T \rightarrow \infty} E[S_T(\omega)] = \int_{-\infty}^{\infty} S_X(p) \delta(\omega - p) dp = S_X(\omega) \quad (19.5.20)$$

Thus, the expected value of the periodogram as  $T \rightarrow \infty$  yields the psd  $S_X(\omega)$ .

### Estimation of Power Spectral Density

We have defined an estimator  $\hat{R}_X(\tau)$  for the autocorrelation function in Eq. (19.4.34a). It is intuitive to express the estimator  $\hat{S}_X(\omega)$  for the psd as the FT of  $\hat{R}_X(\tau)$  as follows:

$$\hat{S}_{XT}(\omega) = \int_{-T}^T \hat{R}_X(\tau) e^{-j\omega\tau} d\tau : \hat{S}_X(\omega) = \lim_{T \rightarrow \infty} \hat{S}_{XT}(\omega) \quad (19.5.21)$$

This equation corresponds to Eq. (19.5.16). The periodogram is an asymptotically unbiased estimator for  $S_X(\omega)$  since  $\lim_{T \rightarrow \infty} E[\hat{S}_X(\omega)] = S_X(\omega)$ . We may be tempted to conclude that since  $\hat{R}_X(\tau)$  is a consistent estimator of  $R_X(\tau)$ , the Fourier transform of  $\hat{R}_X(\tau)$  will also be a consistent estimator of  $S_X(\omega)$ . This is not true because  $\hat{S}_X(\omega)$  fails to converge to  $S_X(\omega)$  for  $T \rightarrow \infty$ . We will show heuristically that  $\text{var}[\hat{S}_{X_T}(\omega)]$  does not approach 0 as  $T \rightarrow \infty$ . From Eq. (19.5.14) the second moment  $E[\hat{S}_{X_T}(\omega)]^2$  is given by

$$\begin{aligned} E[\hat{S}_{X_T}(\omega)]^2 &= \frac{1}{4T^2} \int_{-T}^T \int_{-T}^T \int_{-T}^T \int_{-T}^T E[X(t)X(s)X(v)X(u)] \\ &\quad \times e^{-j\omega(t-s+v-u)} dt ds dv du \end{aligned} \quad (19.5.22)$$

If the process  $X(t)$  is Gaussian, then from Eq. (11.6.8) or Example 19.3.4, we obtain

$$\begin{aligned} E[X(t)X(s)X(v)X(u)] &= R_X(t-s)R_X(v-u) + R_X(t-v)R_X(s-u) \\ &\quad + R_X(t-u)R_X(s-v) \end{aligned} \quad (19.5.23)$$

Substituting Eq. (19.5.23) into Eq. (19.5.22), we have

$$\begin{aligned} E[\hat{S}_{X_T}(\omega)]^2 &= \frac{1}{4T^2} \int_{-T}^T \int_{-T}^T \int_{-T}^T \int_{-T}^T [R_X(t-s)R_X(v-u) \\ &\quad + R_X(t-u)R_X(s-u) + R_X(t-u)R_X(s-v)] \\ &\quad \times e^{-j\omega(t-s+v-u)} dt ds dv du \end{aligned} \quad (19.5.24)$$

Rearranging Eq. (19.5.24), we obtain

$$\begin{aligned}
 E[\hat{S}_{X_T}(\omega)]^2 &= \frac{1}{2T} \int_{-T}^T \int_{-T}^T R_X(t-s)e^{-j\omega(t-s)} dt ds \\
 &\quad \times \frac{1}{2T} \int_{-T}^T \int_{-T}^T R_X(v-u)e^{-j\omega(v-u)} dv du \\
 &\quad + \frac{1}{2T} \int_{-T}^T \int_{-T}^T R_X(t-u)e^{-j\omega(t-u)} dt du \\
 &\quad \times \frac{1}{2T} \int_{-T}^T \int_{-T}^T R_X(v-s)e^{-j\omega(v-s)} dv ds \\
 &\quad + \frac{1}{2T} \int_{-T}^T \int_{-T}^T R_X(t-v)e^{-j\omega(t+v)} dt dv \\
 &\quad \times \frac{1}{2T} \int_{-T}^T \int_{-T}^T R_X(s-u)e^{j\omega(s+u)} ds du
 \end{aligned} \tag{19.5.25}$$

Using Eq. (19.5.21), we can simplify Eq. (19.5.25) as follows:

$$E[\hat{S}_{X_T}(\omega)]^2 = 2\{E[\hat{S}_{X_T}(\omega)]\}^2 + \left\{ \frac{1}{4T^2} \left| \int_{-T}^T \int_{-T}^T R_X(t-v)e^{-j\omega(t+v)} dt dv \right|^2 \right\} \tag{19.5.26}$$

As  $T \rightarrow \infty$ , the second term on the righthand side of Eq. (19.5.26) tends to zero. Hence

$$E[\hat{S}_{X_T}(\omega)]^2 \geq 2\{E[\hat{S}_{X_T}(\omega)]\}^2, \quad \omega \neq 0$$

Subtracting  $\{E[\hat{S}_{X_T}(\omega)]\}^2$  from both sides of this equation, we obtain

$$E[\hat{S}_{X_T}(\omega)]^2 - \{E[\hat{S}_{X_T}(\omega)]\}^2 \geq \{E[\hat{S}_{X_T}(\omega)]\}^2, \quad \omega \neq 0$$

or

$$\text{var}[\hat{S}_{X_T}(\omega)] \geq \{E[\hat{S}_{X_T}(\omega)]\}^2, \quad \omega \neq 0 \tag{19.5.27}$$

Substitution of Eq. (19.5.20)  $\lim_{T \rightarrow \infty} E[\hat{S}_{X_T}(\omega)] = S_X(\omega)$ , in Eq. (19.5.27) results in

$$\lim_{T \rightarrow \infty} \text{var}[\hat{S}_{X_T}(\omega)] \approx S_X^2(\omega), \quad \omega \neq 0 \tag{19.5.28}$$

Thus, as  $T \rightarrow \infty$ ,  $\text{var}[\hat{S}_{X_T}(\omega)]$  does not go to zero but is approximately equal to the psd  $S_X^2(\omega)$ , and hence  $\hat{S}_X(\omega)$  is not a consistent estimator. To make  $\hat{S}_X(\omega)$  consistent, spectral windowing [25] is employed. However, with spectral windowing the asymptotically unbiased nature of the estimator is lost.

### 19.5.2 Discrete Time

As in Eq. (19.5.1), we can also define power spectral density for a discrete-time stationary ergodic random process  $\{X_i = X(t_i), i = 0, \pm 1, \dots\}$ , where the intervals are equally spaced. The autocorrelation function  $R_X(h)$  has been defined in Eq. (19.2.34). The psd  $S_X(\omega)$  is the Fourier transform of  $R_X(h)$  and is given by

$$S_X(\omega) = \text{FT}[R_X(h)] = \sum_{h=-\infty}^{\infty} R_X(h) e^{-j\omega h} \quad (19.5.29)$$

Since the discrete-time Fourier transforms are periodic, the psd for discrete-time random process  $X_i$  as given by Eq. (19.5.29) is periodic with period  $2\pi$ .

The inverse FT of  $S_X(\omega)$  is the autocorrelation function  $R_X(h)$ , given by

$$R_X(h) = \text{IFT}[S_X(\omega)] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_X(\omega) e^{j\omega h} d\omega \quad (19.5.30)$$

**Example 19.5.10 (Discrete Analog of Example 19.5.1)** The autocorrelation  $R_X(h)$  of the discrete-time process  $\{X_i\}$  corresponding to the continuous-time process  $X(t)$  of Example 19.5.1 is given by

$$R_X(h) = \begin{cases} A^2 \left(1 - \frac{|h|}{n}\right), & |h| < n \\ 0, & \text{otherwise} \end{cases}$$

and is shown in Fig. 19.5.12 for  $A = 1$  and  $n = 10$ .

The psd is obtained from Eq. (19.5.29) for finite  $n$  as

$$\begin{aligned} S_X(\omega) &= \sum_{h=-(n-1)}^{n-1} A^2 \left(1 - \frac{|h|}{n}\right) e^{-j\omega h} \\ &= A^2 \left[ 1 + \sum_{h=-(n-1)}^{-1} \left(1 + \frac{h}{n}\right) e^{-j\omega h} + \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) e^{-j\omega h} \right] \\ &= A^2 \left[ 1 + \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) (e^{j\omega h} + e^{-j\omega h}) \right] = A^2 \left[ 1 + 2 \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) \cos(\omega h) \right] \\ &= \frac{A^2}{n} \left[ \frac{1 - \cos(n\omega)}{1 - \cos(\omega)} \right] = \frac{A^2}{n} \left[ \frac{\sin(n\omega/2)}{\sin(\omega/2)} \right]^2 \end{aligned}$$

and

$$S_X(0) = \frac{A^2 n^2}{n} = A^2 n$$

The psd is graphed in Fig. 19.5.13 for  $A = 1$  and  $n = 5$  for values of  $\omega$  between  $-2\pi$  and  $2\pi$ , and the periodic nature of the psd is evident.

Figure 19.5.13 is similar to Fig. 19.5.1 except for the periodicity.

**Example 19.5.11 (Discrete Analog of Example 19.5.2)** The discrete AC function  $R_X(h)$  corresponding to Example 19.5.2 is given by

$$R_X(h) = \sigma_X^2 [1 + e^{-2\lambda|h|}], \quad -\infty < h < \infty$$

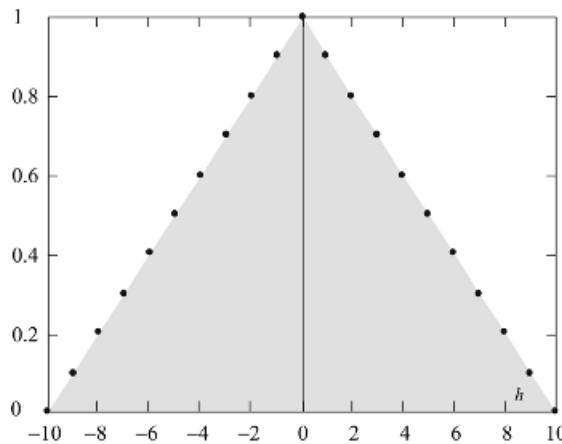


FIGURE 19.5.12

where 1 is a discrete unity function. The psd is determined as follows:

$$\begin{aligned}
 S_X(\omega) &= \sum_{h=-\infty}^{\infty} \sigma_X^2 (1 + e^{-2\lambda|h|}) e^{-j\omega h} \\
 &= \sigma_X^2 \left[ \sum_{h=-\infty}^{\infty} 1 \cdot e^{-j\omega h} + \sum_{h=-\infty}^{-1} e^{2\lambda h} e^{-j\omega h} + \sum_{h=1}^{\infty} e^{-2\lambda h} e^{-j\omega h} \right] \\
 &= \sigma_X^2 \left\{ \sum_{h=-\infty}^{\infty} 2\pi\delta(\omega - 2\pi h) + 1 + \sum_{h=1}^{\infty} [e^{-h(2\lambda+j\omega)} + e^{-h(2\lambda-j\omega)}] \right\} \\
 &= \sigma_X^2 \left\{ \sum_{k=-\infty}^{\infty} 2\pi\delta(\omega - 2\pi h) + \frac{1 - e^{-4\lambda}}{1 + e^{-4\lambda} - 2e^{-2\lambda} \cos(\omega)} \right\}
 \end{aligned}$$

where  $\delta(\omega)$  is the Dirac delta function. The delta function train corresponds the discrete constant  $\sigma_X^2 = \frac{1}{4}$  in the frequency domain. The plot of  $S_X(\omega)$  is shown in Fig. 19.5.14

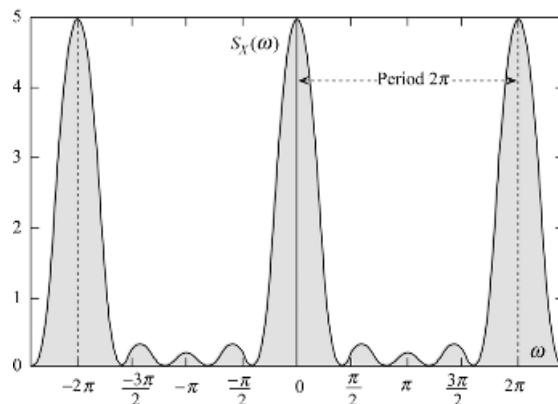


FIGURE 19.5.13

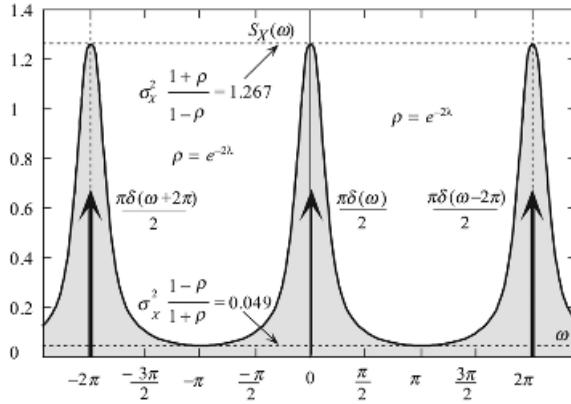


FIGURE 19.5.14

with  $\sigma_X^2 = \frac{1}{5}$  and  $\lambda = \frac{1}{5}$ . The graph is very similar to Fig. 19.5.2 except that it is periodic with period equal to  $2\pi$ .

#### Alternate Form for Power Spectral Density

The discrete psd can also be obtained from the stationary discrete-time random process  $\{X_k\}$ . We will truncate this process in the interval  $(0, N - 1)$  and write

$$X_{kN} = \begin{cases} X_k, & 0 \leq k \leq N - 1 \\ 0, & \text{otherwise} \end{cases} \quad (19.5.31)$$

The discrete-time FT of the sequence  $\{X_{kN}\}$  is given by  $X_N(\omega) = \sum_{k=0}^{N-1} X_k e^{-j\omega k}$ , which is a random variable. We now define the quantity  $S_N(\omega)$  as

$$S_N(\omega) = \frac{1}{N} \left| \sum_{k=0}^{N-1} X_k e^{-j\omega k} \right|^2 = \frac{|X_N(\omega)|^2}{N} = \frac{X_N(\omega) X_N^*(\omega)}{N} \quad (19.5.32)$$

which is the periodogram for the discrete random process. Equation (19.5.32) can be expanded as follows:

$$S_N(\omega) = \frac{1}{N} \sum_{m=0}^{N-1} \sum_{k=0}^{N-1} X_k X_m e^{-j\omega(k-m)} \quad (19.5.33)$$

Substituting  $h = (k - m)$  and  $j = m$  in Eq. (19.5.33), the summation is carried out along the diagonal with limits found by referring to Fig. 19.4.1, resulting in

$$\begin{aligned} S_N(\omega) &= \sum_{h=0}^{N-1} \left[ \frac{1}{N} \sum_{m=-[(N-1)/2]+h}^{(N-1)/2} X_m X_{m+h} \right] e^{-j\omega h} \\ &+ \sum_{h=-(N-1)}^0 \left[ \frac{1}{N} \sum_{m=-[(N-1)/2]}^{[(N-1)/2]-|h|} X_m X_{m+h} \right] e^{-j\omega h} \end{aligned} \quad (19.5.34)$$

The estimated autocorrelation  $\hat{R}_X(h)$  from Eq. (19.4.76) is

$$\hat{R}_X(h) = \frac{1}{N} \sum_{i=0}^{N-h-1} X_i X_{i+h} \quad (19.5.35)$$

and substituting Eq. (19.5.35) in Eq. (19.5.34), we obtain

$$\begin{aligned} S_N(\omega) &= \sum_{h=0}^{N-1} \hat{R}_X(h) e^{-j\omega h} + \sum_{h=-(N-1)}^0 \hat{R}_X(h) e^{-j\omega h} \\ &= \sum_{h=-(N-1)}^{N-1} \hat{R}_X(h) e^{-j\omega h} \end{aligned} \quad (19.5.36)$$

Taking expectations and substituting Eq. (19.4.76b) in Eq. (19.5.36), we obtain

$$E[S_N(\omega)] = \sum_{h=-(N-1)}^{N-1} E[\hat{R}_X(h)] e^{-j\omega h} = \sum_{h=-(N-1)}^{N-1} \left[ 1 - \frac{|h|}{N} \right] R_X(h) e^{-j\omega h} \quad (19.5.37)$$

Taking the limit as of Eq. (19.5.37) as  $N \rightarrow \infty$ , we have

$$\begin{aligned} \lim_{N \rightarrow \infty} E[S_N(\omega)] &= \lim_{N \rightarrow \infty} \sum_{h=-(N-1)}^{N-1} \left[ 1 - \frac{|h|}{N} \right] R_X(h) e^{-j\omega h} \\ &= \lim_{N \rightarrow \infty} \text{DTFT} \left[ R_X(h) \left( 1 - \frac{|h|}{N} \right) \right] \end{aligned} \quad (19.5.38)$$

We use the frequency convolution property of the DTFT

$$x_n y_n \iff \frac{1}{2\pi} X(\omega) * Y(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(p) Y(\omega - p) dp$$

with  $R_X(h) \Leftrightarrow S_X(\omega)$  and from Example 19.5.10

$$\left( 1 - \frac{|h|}{N} \right) \Leftrightarrow \frac{1}{N} \left[ \frac{\sin(\omega N/2)}{\sin(\omega/2)} \right]^2$$

Eq. (19.5.38) can be written as

$$\begin{aligned} \lim_{N \rightarrow \infty} E[S_N(\omega)] &= \lim_{N \rightarrow \infty} \text{DTFT} \left[ R_X(h) \left( 1 - \frac{|h|}{N} \right) \right] \\ &= \lim_{N \rightarrow \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} S_X(p) \frac{1}{N} \left[ \frac{\sin[(\omega - p)N/2]}{\sin[(\omega - p)/2]} \right] dp \end{aligned} \quad (19.5.39)$$

Using the result

$$\lim_{N \rightarrow \infty} \frac{1}{N} \left[ \frac{\sin(\omega N/2)}{\sin(\omega/2)} \right] \rightarrow \sum_{k=-\infty}^{\infty} \delta(\omega - 2k\pi)$$

in Eq. (19.5.39), we obtain

$$\lim_{N \rightarrow \infty} E[S_N(\omega)] = \lim_{N \rightarrow \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} S_X(p) \delta(\omega - p) dp = S_X(\omega) \quad (19.5.40)$$

Thus, the periodogram  $S_N(\omega)$  is an asymptotically unbiased estimator of the psd  $S_X(\omega)$ . The variance of  $S_N(\omega)$  can be obtained by using techniques similar to the continuous case, and we can show a result similar to Eq. (19.5.28):

$$\lim_{N \rightarrow \infty} \text{var}[S_N(\omega)] \approx S_X^2(\omega), \quad \omega \neq 0 \quad (19.5.41)$$

This equation shows that the estimator  $S_N(\omega)$  is not a consistent estimator of  $S_X(\omega)$ .

**Example 19.5.12** To check the validity of Eq. (19.5.41), a computer simulation of discrete white noise of zero mean and unit variance was performed. The power spectral density of the white noise is the variance, or  $S_X(\omega) = \sigma_X^2 = 1$ . The number  $2^m$  of data points with  $m = 7, 9, 10, 11$  were chosen so that they could fit into the discrete Fourier transform algorithm. The psd  $S(\omega)$  was estimated for  $N = 128, 512, 1024$ , and 2048 points using Eq. (19.5.32). Because of the symmetry about  $N/2$ , the estimates  $S_{128}(\omega)$ ,

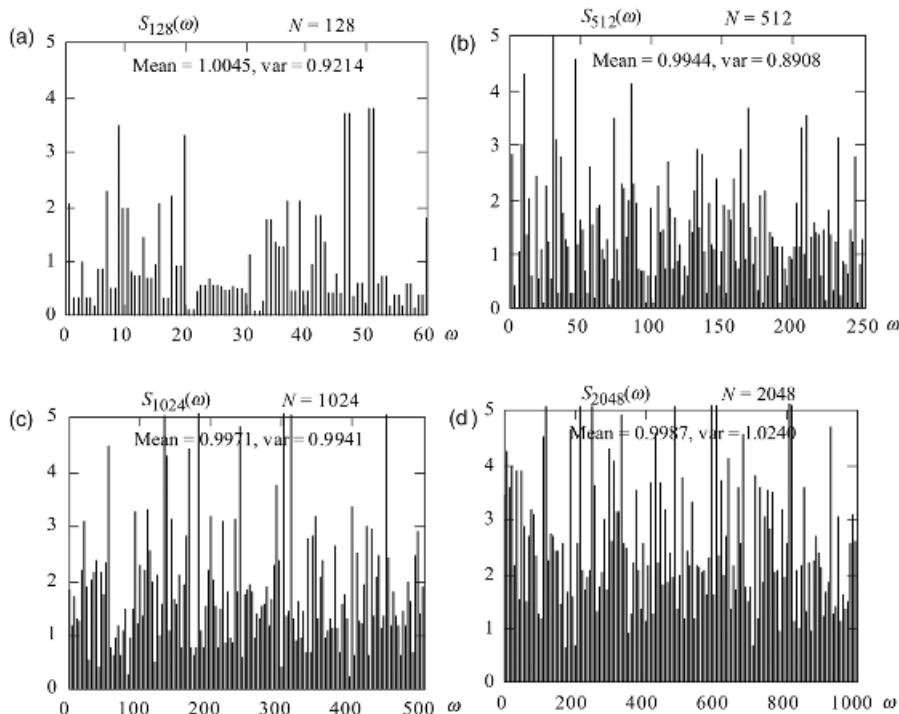


FIGURE 19.5.15

TABLE 19.5.1.

#	PSD	$N$	Mean	Variance
1	$S_X(\omega)$	—	1	0
2	$S_{128}(\omega)$	128	1.0045	0.9214
3	$S_{512}(\omega)$	512	0.9944	0.8908
4	$S_{1024}(\omega)$	1024	0.9971	0.9941
5	$S_{2048}(\omega)$	2048	0.9987	1.0240

$S_{512}(\omega)$ ,  $S_{1024}(\omega)$ , and  $S_{2048}(\omega)$  are graphed for only half the number of data points in Figs. 19.5.15a–19.5.15d.

The estimated psd's and their variances are shown in Table 19.5.1.

The table shows that the estimator  $S_N(\omega)$  is unbiased because the means for all  $N$  are nearly equal to 1 as expected. However, the variances of the psd for all  $N$  are nearly the same, equaling the derived value  $\sigma_x^4 = 1$ , which is the expected result according to Eq. (19.5.41). Thus, the simulation of discrete white noise confirms the desired theoretical result.

# Classification of Random Processes

## 20.1 SPECIFICATIONS OF RANDOM PROCESSES

In the last chapter we discussed the fact that a random process  $X(t)$  is a random variable indexed on a time parameter and the random processes were classified as only nonstationary and stationary. We can also give other specifications of random processes depending on time  $t$  and the state  $X$ . If the timepoints are a set of points like  $\{t_n, n = 0, \pm 1, \pm 2, \dots\}$ , they are *discrete*, and if they are intervals of time on the real line such as  $\{t \leq T\}$  or  $\{-T_1 < t < T_2\}$ , they are *continuous*. The state of a process is the set of values of a sample function  $X_i$  or  $X(t)$ . If this set is finite or countably infinite, the state space is discrete and if it is uncountably infinite, the state space is continuous. On the basis of these definitions, we can specify four sets of random processes as follows.

### 20.1.1 Discrete-State Discrete-Time (DSDT) Process

**Example 20.1.1** The random variable  $X_n$  denotes the position of a particle at time  $n$ . At each time  $i$  the particle undergoes a jump of  $+1$  with probability  $p$  and a jump of  $-1$  with probability  $q$  and stays in the same position with probability  $1 - p - q$ . The jumps are independent. If  $Y_i$  is a random variable representing the jump at the  $i$ th step, then the sequence  $\{Y_i, i = 1, 2, \dots\}$  are independent with  $P\{Y_i = 1\} = p, P\{Y_i = -1\} = q$  and  $P\{Y_i = 0\} = 1 - p - q$ . Initially the particle is at  $X_0 = 0$ . The DSDT process  $X_i$  is given by

$$X_i = X_{i-1} + Y_i, \quad i = 1, \dots, n \quad (20.1.1)$$

This process, called *random walk*, is represented in Fig. 20.1.1.

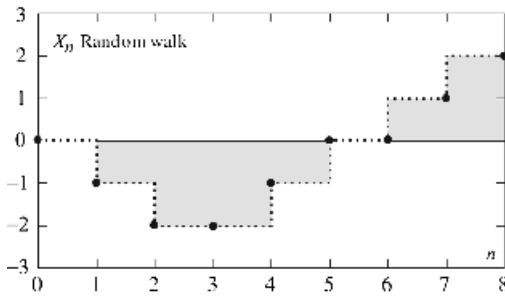


FIGURE 20.1.1

### 20.1.2 Discrete-State Continuous-Time (DSCT) Process

**Example 20.1.2** A random process  $N(t)$  has an increment of unity at each random time  $t_k$  where the random times  $\{t_k\}$  are Poisson-distributed with  $\lambda$  as the occurrence of events per unit time.

The probability of occurrence of  $k$  events in the time interval  $(t_1, t_2)$  is given by

$$P\{N(t_2) - N(t_1) = k\} = \frac{\lambda(t_2 - t_1)^k}{k!} e^{-\lambda(t_2 - t_1)}, \quad k = 0, 1, \dots, \quad (20.1.2)$$

Such a process is known as a *Poisson* or *counting process* and is shown in Fig. 20.1.2.

### 20.1.3 Continuous-State Discrete-Time (CSDT) Process

**Example 20.1.3** In Example 20.1.1 if the jumps are continuously distributed with probability density function  $f_X(x)$  as shown in Fig. 20.1.3, then we have a continuous-state discrete-time process. This process is also a random walk with continuous state space. It also characterizes a sample function discretized at equal intervals of time.

### 20.1.4 Continuous-State Continuous-Time (CSCT) Process

**Example 20.1.4** In this process both the state and the time are continuous. For example, in the extrusion of plastic bags, the thickness of the plastic sheet will be continuously

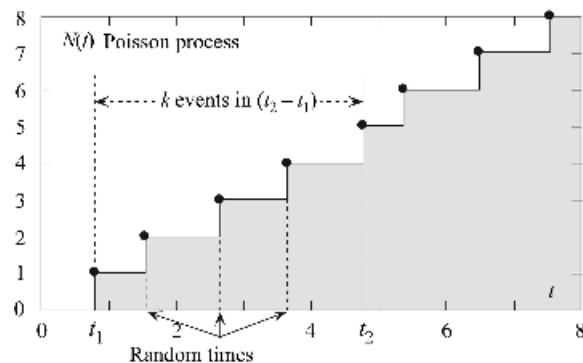


FIGURE 20.1.2

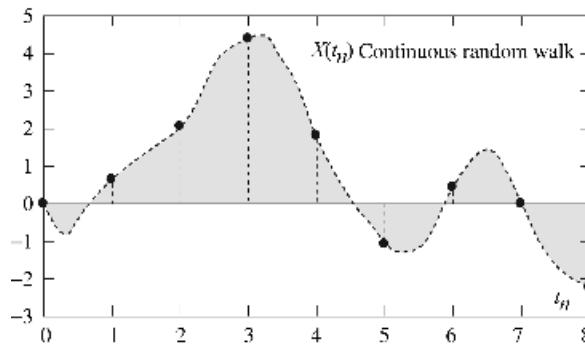


FIGURE 20.1.3

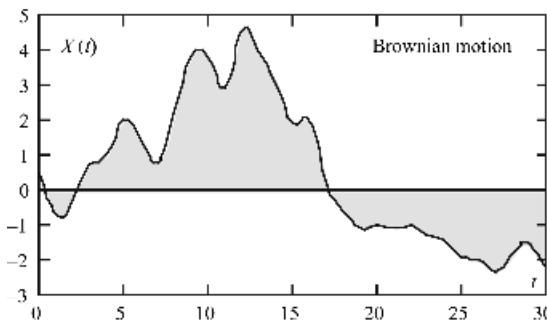


FIGURE 20.1.4

varying with respect to time with the statistics being constant over long periods of time. In the random-walk example (Example 20.1.1), if both the differential time  $\Delta t$  and the differential state  $\Delta x$  tend to zero, and if certain limiting conditions are satisfied, then we get a Brownian motion whose sample function is shown in Fig. 20.1.4.

We will now discuss different types of random processes.

## 20.2 POISSON PROCESS

A random process  $\{N(t), t \geq 0\}$  represents the total number of events (detection of photons or electrons) that have occurred prior to time  $t$ . Denoting the time of occurrence of an event as a random variable  $T$ , we make the following assumptions about the process  $\{N(t), t \geq 0\}$ :

1.  $P\{\text{single event } T \text{ in the interval } t, t + \Delta t\} = P\{N(t + \Delta t) - N(t) = 1\} = P\{N(\Delta t) = 1\} = \lambda\Delta t + o(\Delta t)$ , where any function  $g(x)$  is of  $o(\Delta t)$  if  $\lim_{\Delta t \rightarrow 0} [g(\Delta t)/\Delta t] \rightarrow 0$ .
2.  $P\{\text{no event in the interval } t, t + \Delta t\} = P\{N(t + \Delta t) - N(t) = 0\} = P\{N(\Delta t) = 0\} = 1 - \lambda\Delta t + o(\Delta t)$ .
3.  $P\{\text{more than two events in the interval } t, t + \Delta t\} = P\{N(\Delta t) \geq 2\} = o(\Delta t)$ .
4. Events occurring in nonoverlapping time intervals are statistically independent:

$$\begin{aligned} P\{[N(t_2) - N(t_1)][N(t_4) - N(t_3)]\} &= P\{[N(t_2 - t_1)][N(t_4 - t_3)]\} \\ &= P\{N(t_2 - t_1)\}P\{N(t_4 - t_3)\} \quad \text{for } t_1 < t_2 < t_3 < t_4 \end{aligned} \quad (20.2.1)$$

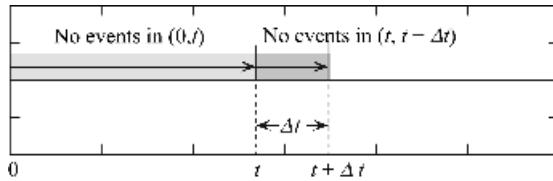


FIGURE 20.2.1

Processes satisfying Eq. (20.2.1) are called *independent increment* processes. Given these assumptions, we will find the probability of  $n$  events occurring in time interval  $t$ :

$$P\{n \text{ events in the time interval } t\} = P\{N(t) = n\} = P_N(n, t)$$

We will first find the probability of no event in the interval  $t + \Delta t = P_N\{0, t + \Delta t\}$ . The interval  $(0, t + \Delta t)$  can be partitioned as  $(0, t)$  and  $(t, t + \Delta t)$  as shown in Fig. 20.2.1.

Thus,  $P_N\{0, t + \Delta t\}$  can be given as probability of no event in  $(0, t)$  and probability of no event in  $(t, t + \Delta t)$ . Since these two intervals are nonoverlapping, we can write from assumption 4

$$P_N\{0, t + \Delta t\} = P_N\{0, t\}P_N\{0, \Delta t\} \quad (20.2.2)$$

and from assumption 2

$$P_N(0, t + \Delta t) = P_N(0, t)(1 - \lambda\Delta t) + o(\Delta t)$$

or

$$P_N(0, t + \Delta t) - P_N(0, t) = -P_N(0, t)\lambda\Delta t + o(\Delta t) \quad (20.2.3)$$

Dividing Eq. (20.2.3) by  $\Delta t$  and taking the limit as  $\Delta t \rightarrow 0$ , we have

$$\frac{dP_N(0, t)}{dt} = -\lambda P_N(0, t) \quad (20.2.4)$$

The initial condition for this first-order differential equation is obtained from  $P\{\text{no events in zero time}\}$  equals 1. Hence the solution to Eq. (20.2.4) with initial condition  $P_N(0, 0) = 1$  is given by

$$P_N(0, t) = e^{-\lambda t}u(t)$$

which is the probability that no event  $T$  occurs in the time interval  $(0, t)$ . Or

$$P\{N(t) = 0\} = P_N(0, t) = P(T > t) = e^{-\lambda t}u(t) \quad (20.2.5)$$

Hence, the probability of occurrence of an event  $T$  in the interval  $(0, t)$  is the complement of Eq. (20.2.5) and is written as

$$P(T \leq t) = F_T(t) = (1 - e^{-\lambda t})u(t)$$

and the density function is

$$f_T(t) = \lambda e^{-\lambda t}u(t) \quad (20.2.6)$$

Equation (20.2.5) will serve as the initial condition for the solution of  $P_N\{n, t\}$ .

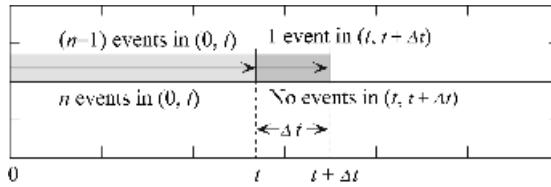


FIGURE 20.2.2

From Fig. 20.2.2 it can be observed that  $n$  events occurring in the interval  $(0, t + \Delta t)$  can be partitioned as  $n$  events in  $(0, t)$  AND no event in  $(t, t + \Delta t)$  OR  $(n - 1)$  events in  $(0, t)$  AND 1 event in  $(t, t + \Delta t)$ . From assumptions 1–4, we can write

$$P_N(n, t + \Delta t) = P_N(n, t)[1 - \lambda \Delta t + o(\Delta t)] + P_N(n - 1, t) \lambda \Delta t + o(\Delta t) \quad (20.2.7)$$

Rearranging terms in Eq. (20.2.7), dividing throughout by  $\Delta t$ , and taking the limit as  $\Delta t \rightarrow 0$ , we obtain a differential–difference equation

$$\lim_{\Delta t \rightarrow 0} \frac{P_N(n, t + \Delta t) - P_N(n, t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} [-P_N(n, t)\lambda + P_N(n - 1, t)\lambda + o(\Delta t)]$$

or

$$\frac{dP_N(n, t)}{dt} = -\lambda P_N(n, t) + \lambda P_N(n - 1, t) \quad (20.2.8)$$

Equation (20.2.8) is a differential–difference equation and the two initial conditions for solving the are

$$1. P_N(0, t) = e^{-\lambda t} u(t)$$

$$2. P_N(n, 0) = 0, n > 0$$

Equation (20.2.8) can be solved recursively as follows:

$$n = 1: \quad \frac{dP_N(1, t)}{dt} = -\lambda P_N(1, t) + \lambda e^{-\lambda t} \quad \text{or} \quad P_N(1, t) = \lambda t e^{-\lambda t}, \quad t \geq 0$$

$$n = 2: \quad \frac{dP_N(2, t)}{dt} = -\lambda P_N(2, t) + \lambda P_N(1, t) \quad \text{or} \quad P_N(2, t) = \frac{(\lambda t)^2}{2} e^{-\lambda t}, \quad t \geq 0$$

Proceeding in a similar manner, we obtain the result for  $n$  events occurring in interval  $t$  as

$$P_N(n, t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad t \geq 0 \quad (20.2.9)$$

The discrete-state continuous-time (DSCT) random process satisfying Eq. (20.2.9) is called a *Poisson random process* and has been shown in Fig. 20.1.2.

### Solution of Eq. (20.2.7) Using Generating Functions

The differential–difference equation [Eq. (20.2.8)] can also be solved using generating functions. From Eq. (11.3.1) we can define the generating function for  $P_N(n, t)$  as

$$G(z, t) = \sum_{n=0}^{\infty} P_N(n, t) z^n \quad (20.2.10)$$

Multiplying Eq. (20.2.8) by  $z^n$  and summing over all  $n$  from 0 to  $\infty$ , we have

$$\frac{d}{dt} \sum_{n=0}^{\infty} P_N(n, t) z^n = -\lambda \sum_{n=0}^{\infty} P_N(n, t) z^n + \lambda \sum_{n=0}^{\infty} P_N(n-1, t) z^n$$

or

$$\frac{d}{dt} G(z, t) = \lambda(z-1)G(z, t) \quad (20.2.11)$$

For a fixed  $z$ , Eq. (20.2.11) is an ordinary differential equation in  $t$ . The initial condition  $G(z, 0)$  is obtained as follows:

$$G(z, 0) = \sum_{n=0}^{\infty} P_N(n, 0) z^n = P_N(0, 0) = 1, \quad \text{since } P_N(n, 0) = 0 \text{ for } n \geq 1$$

With this initial condition, the solution to Eq. (20.2.11) is

$$G(z, t) = e^{\lambda t(z-1)}, \quad t \geq 0 \quad (20.2.12)$$

The power-series expansion of Eq. (20.2.12) yields

$$G(z, t) = e^{-\lambda t} \left[ 1 + z \frac{\lambda t}{1!} + z^2 \frac{(\lambda t)^2}{2!} + \cdots + z^n \frac{(\lambda t)^n}{n!} + \cdots \right]$$

Hence,  $P_N(n, t)$  is the coefficient of  $z^n$  in the preceding equation and is given by

$$P_N(n, t) = P\{N(t) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad t \geq 0 \quad (20.2.13)$$

The statistics of the Poisson process as derived in Example 19.1.7 are

*Mean:*

$$E[N(t)] = \mu_N(t) = \lambda t$$

*Variance:*

$$\text{var}[N(t)] = \sigma_N^2(t) = \lambda t$$

*Autocorrelation:*

$$E[N(t_1)N(t_2)] = R_N(t_1, t_2) = \lambda^2 t_1 t_2 + \lambda \min(t_1, t_2)$$

*Autocovariance:*

$$E\{[N(t_1) - \mu_N(t_1)][N(t_2) - \mu_N(t_2)]\} = C_N(t_1, t_2) = \lambda \min(t_1, t_2)$$

*Normalized Autocovariance:*

$$\rho_N(t_1, t_2) = \frac{C_N(t_1, t_2)}{\sigma_N(t_1)\sigma_N(t_2)} = \frac{1}{\lambda} \min\left(\sqrt{\frac{t_1}{t_2}}, \sqrt{\frac{t_2}{t_1}}\right)$$

Since the mean is a function of time  $t$  and the autocorrelation is a function of  $t_1$  and  $t_2$ , the Poisson process is a nonstationary process.

### **Distribution of Interarrival Times**

We will now find the distribution of interarrival times for the Poisson process  $N(t)$ . Let  $T_1$  be the random time of occurrence of the first event after time  $t = 0$ ,  $T_2$  the random time

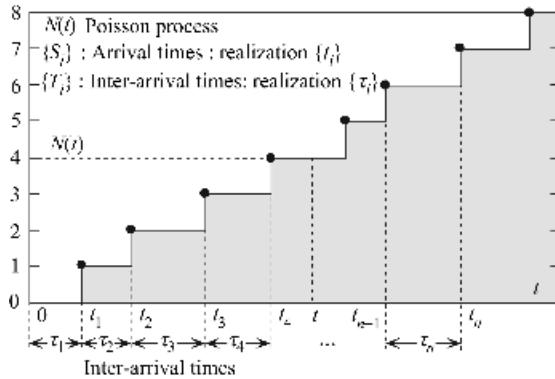


FIGURE 20.2.3

between the occurrence of the first and second events, and  $T_n$  the random time between the occurrence of the  $(n-1)$ st and  $n$ th events as shown in Fig. 20.2.3. The sequence of times  $\{T_i, i = 1, \dots, n\}$  is called *interarrival* times. Let  $\{\tau_i, i = 1, \dots, n\}$  be the realizations of these random interarrival times. Let  $S_n$  be the random variable representing the arrival time of the  $n$ th event or the waiting time until the  $n$ th event:

$$S_1 = T_1, S_2 = T_1 + T_2, \dots, S_n = T_1 + T_2 + \dots + T_n = \sum_{i=1}^n T_i \quad (20.2.14)$$

If  $\{t_i, i = 1, \dots, n\}$  are the realizations of the random arrival times  $\{S_i, i = 1, \dots, n\}$ , then

$$t_1 = \tau_1, t_2 = \tau_1 + \tau_2, \dots, t_n = \tau_1 + \tau_2 + \dots + \tau_n = \sum_{i=1}^n \tau_i \quad (20.2.15)$$

We will find the distributions of these interarrival times  $\{T_i\}$  and the waiting time for the  $n$ th event  $S_n$ .

The Poisson process can also be specified in terms of the arrival times  $\{S_i\}$  as

$$N(t) = \sum_{i=1}^{\infty} u(t - S_i) \quad (20.2.16a)$$

where  $u(t)$  is a unit step function. Referring to Fig. 20.2.3,  $N(t)$  is the total number of events that have occurred in the time interval  $t$ , and each event causes an increment by 1. Hence Eq. (20.2.16a) can be simplified as

$$N(t) = \sum_{i=1}^{\infty} u(t - S_i) = \sum_{i=1}^{N(t)} u(t - S_i) = \sum_{i=1}^{N(t)} 1, \quad t \geq 0 \quad (20.2.16b)$$

By definition, the first event  $\{T_1 > t\}$  occurs after time  $t$ , and hence there are no events occurring in the interval  $(0, t)$ . Hence from Eq. (20.2.5), we obtain

$$P\{T_1 > t\} = P\{N(t) = 0\} = P_N(0, t) = e^{-\lambda t} u(t) \quad (20.2.17)$$

We will show that the interarrival times  $T_1, T_2, \dots, T_n$  are independent identically distributed (i.i.d.) random variables with density function  $e^{-\lambda t} u(t)$ . The second event

$\{T_2 > t\}$  occurs after the first event  $T_1 = \tau_1$ :

$$P\{T_2 > t | T_1 = \tau_1\} = P\{T_2 > t | S_1 = t_1\} = P\{\text{no events in } (t_1, t_1 + t) | S_1 = t_1\} \quad (20.2.18)$$

From the independence of nonoverlapping intervals (assumption 4) of the Poisson process, Eq. (20.2.18) can be written as

$$P\{T_2 > t | S_1 = t_1\} = P\{T_2 > t\} = P\{\text{no events in } (t_1, t_1 + t)\} = e^{-\lambda t} u(t) \quad (20.2.19)$$

Hence  $T_1$  and  $T_2$  are independent.

In a similar manner, for the  $n$ th arrival time  $T_n$ , we can write

$$P\{T_n > t | T_{n-1} = \tau_{n-1}, \dots, T_1 = \tau_1\} = P\{T_n > t | S_{n-1} = t_{n-1}, \dots, S_1 = t_1\} \quad (20.2.20)$$

Again, from the independent increment property, Eq. (20.2.20) can be simplified to

$$\begin{aligned} P\{T_n > t | S_{n-1} = t_{n-1}, \dots, S_1 = t_1\} &= P\{T_n > t\} = P\{\text{no events in } (t_n, t)\} \\ &= e^{-\lambda t} u(t) \end{aligned} \quad (20.2.21)$$

Thus the interarrival times  $T_1, T_2, \dots, T_n$  are i.i.d. random variables with  $F_T(t) = P(T_n \leq t) = (1 - e^{-\lambda t})u(t)$  and density function  $f_T(t) = \lambda e^{-\lambda t} u(t)$ . As a consequence, the probability density function for the waiting time for the  $n$ th event  $S_n = \sum_{i=1}^n T_i, n \geq 1$  is an  $n$ -fold convolution of  $\lambda e^{-\lambda t} u(t)$ , yielding an Erlang density with  $n$ -degrees of freedom similar to Eq. (7.4.1):

$$f_{S_n}(t) = \frac{(\lambda t)^{n-1} \lambda e^{-\lambda t}}{(n-1)!}, \quad \lambda > 0, t > 0, n: \text{integer} > 0 \quad (20.2.22)$$

**Example 20.2.1** A message sending station sends two independent bitstreams  $X$  and  $Y$  with Poisson rates  $\lambda$  and  $\mu$ . The receiving station receives the composite stream  $W = X + Y$ . We have to find (1) the rate of the composite stream and (2) the probability that stream  $Y$  arrives first. The density functions of the bitstreams  $X$  and  $Y$  are

$$f_X(x) = e^{-\lambda x} \frac{(\lambda x)^k}{k!} u(x) \quad \text{and} \quad f_Y(y) = e^{-\mu y} \frac{(\mu y)^m}{m!} u(y)$$

1. We will find the rate of  $W = X + Y$  using generating functions. The generating functions for  $X$  and  $Y$  are

$$P_X(z) = e^{\lambda(z-1)} \quad \text{and} \quad P_Y(z) = e^{\mu(z-1)}$$

Since  $X$  and  $Y$  are independent, the pdf  $f_W(w)$  is the convolution of  $f_X(x)$  and  $f_Y(y)$ , and the generating function  $P_W(z)$  for  $f_W(w)$  is given by the product  $P_X(z)P_Y(z)$ , or  $P_W(z) = e^{-(\lambda+\mu)(z-1)}$ . Hence

$$f_W(w) = e^{-(\lambda+\mu)w} \frac{[(\lambda+\mu)w]^k}{k!} u(w)$$

2. Let the times of arrival of streams  $X$  and  $Y$  be the random variables  $S$  and  $T$ . From Eq. (20.2.6) the interarrival times are exponentially distributed and hence

$$f_S(s) = \lambda e^{-\lambda s} u(s) \quad \text{and} \quad f_T(t) = \mu e^{-\mu t} u(t)$$

Since the streams are independent, the arrival times are also independent, and the joint density of the arrival times  $S$  and  $T$  is

$$f_{ST}(s, t) = \lambda\mu e^{-(\lambda t + \mu s)}, \quad s \geq 0, t \geq 0$$

The event that stream  $Y$  arrives before stream  $X$  is the same as the arrival time of  $Y$  is before the arrival time of  $X$ . Hence  $P\{Y \leq X\} = P\{T \leq S\}$ . We will find  $P(T \leq S)$ . The region  $\{T \leq S\}$  is shown in Fig. 20.2.4:

$$\begin{aligned} P(T \leq S) &= \int_0^\infty \int_s^\infty \lambda\mu e^{-(\lambda t + \mu s)} dt \, ds = \int_0^\infty \lambda\mu e^{-\mu s} \cdot \frac{e^{-\lambda t}}{-\lambda} \Big|_s^\infty \\ &= \mu \int_0^\infty e^{-(\lambda + \mu)s} ds = \frac{\mu}{\lambda + \mu} \end{aligned}$$

Hence

$$P(Y \leq X) = \frac{\mu}{\lambda + \mu}$$

**Example 20.2.2** A Poisson process  $N(t)$  with rate  $\lambda$  consists of a subprocess  $N_1(t)$  with probability  $p$  and rate  $\lambda_1$  and another subprocess  $N_2(t)$  with probability  $(1 - p)$  and rate  $\lambda_2$ . We have to find the probabilities  $P\{N_1(t) = n\}$  and  $P\{N_2(t) = m\}$  and their rates  $\lambda_1$  and  $\lambda_2$ .

We are given the following:

$$N(t) \sim e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad N_1(t) \sim e^{-\lambda_1 t} \frac{(\lambda_1 t)^k}{k!}, \quad N_2(t) \sim e^{-\lambda_2 t} \frac{(\lambda_2 t)^k}{k!}$$

From the statement of the problem  $N(t) = N_1(t) + N_2(t)$ , and if  $N_1(t) = n$  and  $N_2(t) = m$ , then  $N(t) = n + m$ . The joint probability  $P\{N_1(t) = n, N_2(t) = m\}$  can be given in terms of conditional probabilities as

$$P\{N_1(t) = n, N_2(t) = m\} = P\{[N_1(t) = n, N_2(t) = m] \mid N(t) = m + n\} e^{-\lambda t} \frac{(\lambda t)^{m+n}}{(m+n)!}$$

where  $P\{N(t) = m + n\} = e^{-\lambda t} (\lambda t)^{m+n} / [(m+n)!]$  has been substituted in the equation above. The event  $[N_1(t) = n, N_2(t) = m] \mid N(t) = m + n\}$  can be considered as  $(m+n)$

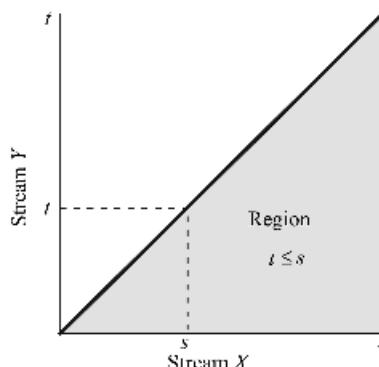


FIGURE 20.2.4

Bernoulli trials with success defined as  $N_1(t) = n$  with probability  $p$  and failure as  $N_2(t) = m$  with probability  $(1 - p)$ . Thus, using the binomial distribution [Eq. (4.2.2)], we have

$$P\{[N_1(t) = n, N_2(t) = m] \mid N(t) = m + n\} = \binom{m+n}{n} p^n (1-p)^m$$

and  $P\{N_1(t) = n, N_2(t) = m\}$  can be given by

$$\begin{aligned} P\{N_1(t) = n, N_2(t) = m\} &= \frac{(m+n)!}{n!m!} p^n (1-p)^m e^{-\lambda t} \frac{(\lambda t)^{m+n}}{(m+n)!} \\ &= e^{-\lambda t p} e^{-\lambda t(1-p)} \frac{(\lambda t p)^n}{n!} \frac{[\lambda t(1-p)]^m}{m!} \end{aligned}$$

The marginal distribution  $P\{N_1(t) = n\}$  can be given as follows:

$$P\{N_1(t) = n\} = e^{-\lambda t p} \frac{(\lambda t p)^n}{n!} \sum_{m=0}^{\infty} e^{-\lambda t(1-p)} \frac{[\lambda t(1-p)]^m}{m!} = e^{-\lambda t p} \frac{(\lambda t p)^n}{n!}$$

Similarly

$$P\{N_2(t) = m\} = e^{-\lambda t(1-p)} \frac{[\lambda t(1-p)]^m}{m!}$$

Hence, the Poisson rates for  $N_1(t)$  and  $N_2(t)$  are  $\lambda_1 = \lambda p$  and  $\lambda_2 = \lambda(1 - p)$ .

Let  $S_{1n}$  and  $S_{2m}$  be the waiting times for the  $n$ th and  $m$ th events of the Poisson subprocesses  $N_1(t)$  and  $N_2(t)$  respectively. We will find the probabilities of (1) the waiting time for the first event in  $N_1(t)$  is greater than the waiting time for the first event in  $N_2(t)$ , or  $P\{S_{11} > S_{21}\}$ , and (2) the waiting time for the second event in  $N_1(t)$  is less than the waiting time for the first event in  $N_2(t)$ , or  $P\{S_{12} < S_{21}\}$ .

From Eq. (20.2.14) the waiting times can be given in terms of the interarrival times, or  $S_{11} = T_{11}$ ,  $S_{21} = T_{21}$  and  $S_{12} = T_{11} + T_{12}$ .

### 1. From Example 20.2.1

$$P\{S_{11} > S_{21}\} = P\{T_{11} > T_{21}\} = \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{\lambda p}{\lambda p + \lambda(1-p)} = p$$

### 2. Since $S_{11} < S_{12}$ , we can write $P\{S_{12} < S_{21}\} = P\{S_{12} < S_{21}, S_{11} < S_{21}\}$ or

$$P\{S_{12} < S_{21}\} = P\{S_{12} < S_{21} \mid S_{11} < S_{21}\} P\{S_{11} < S_{21}\}$$

From Example 20.2.1  $P\{S_{11} < S_{21}\} = \lambda_2 / (\lambda_1 + \lambda_2)$  and

$$\begin{aligned} P\{S_{12} < S_{21} \mid S_{11} < S_{21}\} &= P\{T_{11} + T_{12} < T_{21} \mid T_{11} < T_{21}\} = P\{T_{12} < T_{21}\} \\ &= \frac{\lambda_2}{\lambda_1 + \lambda_2} \end{aligned}$$

Hence

$$P\{S_{12} < S_{11}\} = \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right) = (1-p)^2$$

**Example 20.2.3** Advance booking for a World Series baseball game is open on weekdays only from 10:00 a.m. to 3:00 p.m. During the lunch hour from 12:00 to 1:00 p.m.,

customers arrive at the rate of  $\lambda = 8 \text{ h}^{-1}$  (8 per hour) and for the rest of the time at the rate of  $\lambda = 4\text{h}^{-1}$ . The arrival time  $N(t)$  of customers to the booking counter is a random process given by

$$P\{N(t) = k\} = e^{-\Lambda t} \frac{(\Lambda t)^k}{k!}$$

where the rate  $\Lambda$  is a binary random variable assuming values of 4 and 8 with probabilities 0.2 and 0.8 respectively. We have to find  $E[N(t)]$ ,  $\text{var}[N(t)]$  and determine the variance to mean ratio  $\sigma^2/\mu$ . If this ratio equals 1, we can conclude that  $N(t)$  is Poisson.

We note that the distribution of  $N(t)$  conditioned on either  $\Lambda=8$  or 4 is Poisson. Hence  $E[N(t) | \Lambda = 8] = 8$  and  $E[N(t) | \Lambda = 4] = 4$ . This result is used in the following equation:

$$\begin{aligned} E[N(t)] &= E[N(t) | \Lambda = 8]P(\Lambda = 8) + E[N(t) | \Lambda = 4]P(\Lambda = 4) \\ &= 8 \times 0.2 + 4 \times 0.8 = 1.6 + 3.2 = 4.8 \end{aligned}$$

In a similar manner, we can find the second moment  $E[N^2(t)]$  with conditional second moments  $E[N^2(t) | \Lambda = 8] = \lambda + \mu^2 = 8 + 64 = 72$  and  $E[N^2(t) | \Lambda = 4] = 4 + 16 = 20$ . From the conditional second moments,  $E[N^2(t)]$  can be given as follows:

$$\begin{aligned} E[N^2(t)] &= E[N^2(t) | \Lambda = 8]P(\Lambda = 8) + E[N^2(t) | \Lambda = 4]P(\Lambda = 4) \\ &= 72 \times 0.2 + 20 \times 0.8 = 30.4 \end{aligned}$$

Hence the variance of  $N(t)$  is given by

$$\text{var}(N(t)) = E[N^2(t)] - \{E[N(t)]\}^2 = 30.4 - 4.8^2 = 7.36$$

The variance : mean ratio  $\sigma^2/\mu = 7.36/4.8 = 1.53$ . Since this ratio is not equal to 1,  $N(t)$  is not a Poisson process.

**Example 20.2.4** Customers arrive at a checkout counter at a Poisson rate of  $\lambda$  per minute. The service time is a random variable  $Y$  in minutes. We consider two cases: (1)  $Y$  is a constant  $= k$  minutes and (2)  $Y$  is exponentially distributed with density function  $f_Y(y) = \mu e^{-\mu y} u(y)$ . For both cases we will find the probabilities that (1) a customer does not have to wait and (2) the mean value of a customer's waiting time.

From Eq. (20.2.6) the interarrival time  $T$  of customers is distributed as  $f_T(t) = \lambda e^{-\lambda t} u(t)$ , and from Eq. (20.2.19) the interarrival times are independently distributed. The service time  $Y$  is distributed as  $f_Y(y) = \mu e^{-\mu y} u(y)$ . If the interarrival time  $T$  is greater than the service time  $Y$ , then there will be no waiting time for the customer.

**1.** Service time is a constant:  $Y = k$  minutes

(a)  $P\{\text{no waiting time for the customer}\} = P\{T > k\}$ . Hence

$$P\{T > k\} = \int_k^{\infty} \lambda e^{-\lambda t} dt = \frac{\lambda e^{-\lambda t}}{-\lambda} \Big|_k^{\infty} = e^{-\lambda k}$$

(b) If  $T \leq k$ , then the customer has to wait  $(k - T)$  minutes and the average waiting time is given by

$$E[k - T] = \int_0^k (k - t) \lambda e^{-\lambda t} dt = k - \frac{1}{\lambda} (1 - e^{-\lambda k})$$

2. Service time  $Y$  exponentially distributed:  $f_Y(y) = \mu e^{-\mu y} u(y)$

(a) Again,  $P\{\text{no waiting time for the customer}\} = P\{T > Y\}$  and

$$\begin{aligned} P\{T > Y\} &= \int_0^\infty P\{T > Y \mid Y = t\} f_Y(t) dt \\ &= \int_0^\infty e^{-\lambda t} \mu e^{-\mu t} dt = \frac{\mu}{\lambda + \mu} \end{aligned}$$

(b) Here also, if  $T \leq Y$ , then the customer has to wait  $(Y - T)$  minutes. We have to find  $E[Y - T]$ . In terms of conditional expectation, we can write

$$E[Y - T] = \int_0^\infty E[Y - T \mid Y = y] f_Y(y) dy$$

and from the result 1b (above) the conditional expectation  $E[Y - T \mid Y = y]$  is

$$E[Y - T \mid Y = y] = \int_0^y (y - t) \lambda e^{-\lambda t} dt = y - \frac{1}{\lambda} (1 - e^{-\lambda y})$$

Substituting this equation into the previous equation, we obtain

$$E[Y - T] = \int_0^\infty \left[ y - \frac{1}{\lambda} (1 - e^{-\lambda y}) \right] \mu e^{-\mu y} dy = \frac{\lambda}{\mu(\lambda + \mu)}$$

### Poisson White Noise

The Poisson process given by Eq. (20.2.16) can be further generalized by associating with each event  $T$  a random weight  $A$ . The weights are i.i.d. random variables with density function  $f_A(a)$  and are also independent of the arrival times  $S$ . The weighted Poisson process  $X(t)$  is defined by

$$X(t) = \sum_{i=1}^{\infty} A_i u(t - S_i), \quad t \geq 0 \quad (20.2.23a)$$

which can be simplified from Eq. (20.2.16.b) as follows:

$$X(t) = \sum_{i=1}^{\infty} A_i u(t - S_i) = \sum_{i=1}^{N(t)} A_i u(t - S_i) = \sum_{i=1}^{N(t)} A_i, \quad t \geq 0 \quad (20.2.23b)$$

The pdf  $f_X(x; t)$  can be expressed in terms of the conditional expectation as

$$f_X(x; t) = \sum_{N(t)} f_X[x; t \mid N(t)] P\{N(t)\} \quad (20.2.24)$$

The conditional probability  $f_X(x; t \mid N(t) = i)$  can be evaluated from Eqs. (20.2.23):

$$i = 1: \quad f_X[x; t \mid N(t) = 1] = f_A(a) \quad (20.2.25)$$

$$i = 2: \quad f_X[x; t \mid N(t) = 2] = f_A(a_1 + a_2) \quad (20.2.26)$$

Since the  $A_i$  terms are i.i.d. random variables,  $f_A(a_1 + a_2)$  in Eq. (20.2.26) is a convolution of  $f_A(a)$  with itself; or

$$f_X[x; t \mid N(t) = 2] = f_A(a) * f_A(a)$$

Proceeding in this fashion,  $f_X[x; t \mid N(t) = i]$  is an  $i$ -fold convolution of  $f_A(a)$  and

$$f_X[x; t \mid N(t) = i] = f_A(a) * f_A(a) * \dots * f_A(a) = f_A^{(i)}(a) \quad (20.2.27)$$

In Eq. (20.2.4),  $P\{N(t) = i\}$  is a Poisson distribution given by Eq. (20.10.13):

$$P\{N(t) = i\} = P_N(i, t) = e^{-\lambda t} \frac{(\lambda t)^i}{i!} \quad (20.2.28)$$

Substituting Eqs. (20.2.27) and (20.2.28) in Eq. (20.2.24) we obtain the pdf of  $X(t)$  as

$$f_X(x; t) = e^{-\lambda t} \sum_{i=1}^{\infty} f_A^{(i)}(a) \frac{(\lambda t)^i}{i!} \quad (20.2.29)$$

The mean and autocorrelation of the generalized Poisson process  $X(t)$  can be obtained from the properties of expectations as follows:

*Mean:*

$$\begin{aligned} E[X(t)] &= E\{E[X(t) | N(t) = n]\} \\ &= E\left\{E\left[\sum_{i=1}^{N(t)=n} A_i\right]\right\} = \sum_{n=1}^{\infty} \sum_{i=1}^{N(t)=n} E[A_i] P_N(n, t) \\ &= \sum_{n=1}^{\infty} \mu_A n \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \mu_A e^{-\lambda t} \sum_{n=1}^{\infty} \frac{(\lambda t)^n}{(n-1)!} \\ &= \mu_A \lambda t e^{-\lambda t} = \sum_{m=0}^{\infty} \frac{(\lambda t)^m}{m!} = \mu_A \lambda t = \mu_X(t), \quad t \geq 0 \end{aligned} \quad (20.2.30)$$

*Autocorrelation:*

$$\begin{aligned} E[X(t_1)X(t_2)] &= E\{X(t_1)[X(t_2) - X(t_1) + X(t_1)]\} \\ &= E[X^2(t_1)] + E[X(t_1)]E[X(t_2) - X(t_1)] \\ &= E\{E[X^2(t_1) | N(t_1)]\} \\ &\quad + E[X(t_1) | N(t_1)]E[(X(t_2) - X(t_1)) | N(t_2 - t_1)] \end{aligned} \quad (20.2.31)$$

We will evaluate  $E\{E[X^2(t_1) | N(t_1)]\}$ :

$$\begin{aligned} E\{E[X^2(t_1) | N(t_1) = n]\} &= E\left\{\sum_{i=1}^n \sum_{j=1}^n E[A_i A_j]\right\} \\ &= E\left\{\sum_{i=1}^n E[A_i^2] + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n E[A_i] E[A_j]\right\} \\ &= E\{nE[A^2] + n(n-1)\mu_A^2\} \\ &= \sum_{n=1}^{\infty} E[A^2] n e^{-\lambda t_1} \frac{(\lambda t_1)^n}{n!} + \sum_{n=1}^{\infty} \mu_A^2 n(n-1) e^{-\lambda t_1} \frac{(\lambda t_1)^n}{n!} \\ &= E[A^2] \lambda t_1 e^{-\lambda t_1} \sum_{m=0}^{\infty} \frac{(\lambda t_1)^m}{m!} + \mu_A^2 \lambda^2 t_1^2 e^{-\lambda t_1} \sum_{m=0}^{\infty} \frac{(\lambda t_1)^m}{m!} \\ &= E[A^2] \lambda t_1 + \mu_A^2 \lambda^2 t_1^2 \end{aligned} \quad (20.2.32)$$

Substituting Eqs. (20.2.30) and (20.3.32) in Eq. (20.2.31) and  $E[A^2] = \sigma_A^2 + \mu_A^2$ , we obtain

$$\begin{aligned} E[X(t_1)X(t_2)] &= E[A^2]\lambda(t_1) + \mu_A^2\lambda^2t_1^2 + \mu_A\lambda t_1[\mu_A\lambda t_2 - \mu_A\lambda t_1] \\ &= [\sigma_A^2 + \mu_A^2]\lambda t_1 + \mu_A^2\lambda^2t_1t_2 = R_X(t_1, t_2) \end{aligned} \quad (20.2.33a)$$

Equation (20.2.33a) was obtained on the premise that  $t_1 < t_2$ . We can derive a similar equation for  $t_2 < t_1$ . Hence the autocorrelation function  $R_X(t_1, t_2)$  is given by

$$R_X(t_1, t_2) = [\sigma_A^2 + \mu_A^2]\lambda \min(t_1, t_2) + \mu_A^2\lambda^2t_1t_2, t_1, t_2, \geq 0 \quad (20.2.33b)$$

The autocovariance function  $C_X(t_1, t_2)$  is given by

$$C_X(t_1, t_2) = (\sigma_A^2 + \mu_A^2)\lambda \min(t_1, t_2), t_1, t_2 \geq 0 \quad (20.2.34)$$

**Example 20.2.5** The events of a Poisson process with  $\lambda = \frac{1}{2}$  are weighted with i.i.d. random variables  $\{A_i\}$ , resulting in a weighted Poisson process  $X(t)$ . The weights are uniformly distributed in the interval (0,2). We will find the mean, autocorrelation, and autocovariance of  $X(t)$ . The mean value of the weights  $\mu_A = 1$ . The variance of the weights  $\sigma_A^2 = \frac{1}{3}$ .  $X(t)$  given by Eq. (20.2.23) is shown in Fig. 20.2.5.

*Mean.* From Eq. (20.2.30) the mean value of  $X(t)$  is  $\mu_X(t) = \mu_A\lambda t = 1 \times \frac{1}{2} \times t = t/2$ .

*Autocorrelation.* The second moment of  $A$  is  $E[A^2] = \sigma_A^2 + \mu_A^2 = \frac{1}{3} + 1^2 = \frac{4}{3}$ . Hence, from Eq. (20.2.33), the autocorrelation function is given by

$$R_X(t_1, t_2) = \frac{4}{3}\frac{1}{2}\min(t_1, t_2) + 1^2\left(\frac{1}{2}\right)^2 t_1t_2 = \frac{2}{3}\min(t_1, t_2) + \frac{1}{4}t_1t_2$$

*Autocovariance.* From Eq. (20.2.34) the ACF is  $C_X(t_1, t_2) = \frac{2}{3}\min(t_1, t_2)$ .

### Formal Derivative of Poisson Process

The generalized Poisson process given by Eq. (20.2.23) shown in Fig. 20.2.5 consists of a series of step functions occurring at Poisson-distributed random times with random

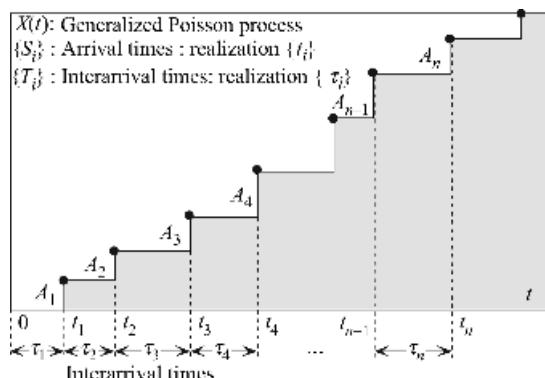


FIGURE 20.2.5

weights. The formal derivative of  $X(t)$  yields a Poisson impulse process  $V(t)$  consisting of a series of Dirac delta functions:

$$\frac{d}{dt}X(t) = \frac{d}{dt} \sum_{i=1}^{\infty} A_i u(t - S_i) = \sum_{i=1}^{\infty} A_i \delta(t - S_i), \quad t \geq 0 \quad (20.2.35)$$

The mean of  $V(t)$  can be calculated as

$$\mu_V(t) = \frac{d\mu_X(t)}{dx} = \frac{d\mu_A \lambda t}{dt} = \mu_A \lambda, \quad t \geq 0 \quad (20.2.36)$$

The autocorrelation function of  $V(t)$  is

$$\begin{aligned} R_V(t_1, t_2) &= \frac{\partial^2 R_X(t_1, t_2)}{\partial t_1 \partial t_2} = \frac{\partial^2}{\partial t_1 \partial t_2} \{ [\sigma_A^2 + \mu_A^2] \lambda \min(t_1, t_2) + \mu_A^2 \lambda^2 t_1 t_2 \} \\ &= [\sigma_A^2 + \mu_A^2] \lambda \delta(t_2 - t_1) + \mu_A^2 \lambda^2 \\ &= R_V(t_2 - t_1) = R_V(\tau) \end{aligned} \quad (20.2.37)$$

Since  $\mu_V$  is independent of time and  $R_V(t_1, t_2)$  is a function of the time difference  $t_2 - t_1$ , we conclude that  $V(t)$  is a stationary process.

The autocovariance function is

$$C_V(t_1, t_2) = [\sigma_A^2 + \mu_A^2] \lambda \delta(t_2 - t_1) = C_V(t_2 - t_1) = C_V(\tau) \quad (20.2.38)$$

The FT  $S_V(\omega)$  of the acf  $C_V(\tau)$  of Eq. (20.2.38) is given by

$$S_V(\omega) = FT \{ [\sigma_A^2 + \mu_A^2] \lambda \delta(t_2 - t_1) \} = \lambda [\sigma_A^2 + \mu_A^2] \quad (20.2.39)$$

This is the psd of the process  $V'(t) = V(t) - \mu_A \lambda$ . Since the spectrum is constant over the entire frequency range, the process  $V'(t)$  is called *Poisson white noise*.

**Example 20.2.6** We will determine the derivative process  $V(t)$  of Example 20.2.5, where  $\lambda = \frac{1}{2}$  and the weight process  $A > 0$  is uniformly distributed in the interval  $(0,2)$ . The process  $V(t)$  is shown in Fig. 20.2.6.

The mean  $\mu_V$  is obtained from Eq. (20.2.36) as  $\mu_V = \mu_A \lambda = 1 \times \frac{1}{2} = \frac{1}{2}$ .

The autocorrelation function from Eq. (20.2.37) is

$$R_V(\tau) = [\sigma_A^2 + \mu_A^2] \delta(t_2 - t_1) + \mu_A^2 \lambda^2 = \frac{4}{3} \delta(\tau) + \frac{1}{4}$$

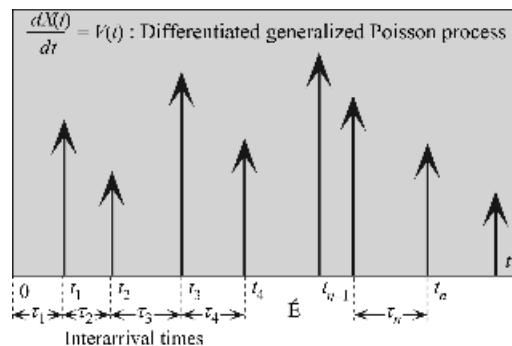


FIGURE 20.2.6

and the autocovariance from Eq. (20.2.38) is  $C_V(\tau) = \frac{4}{3}\delta(\tau)$ . The psd equals  $FT[R_V(\tau)]$  given by  $S_V(\omega) = \frac{4}{3} + (\pi/2)\delta(\omega)$ .

## 20.3 BINOMIAL PROCESS

### Bernoulli Process

In Section 4.1 we discussed Bernoulli trials as repeated functionally and statistically i.i.d. trials, each of which consists of only two events  $s$  and  $f$  with probabilities  $p$  and  $q = (1 - p)$  respectively. Consider a countably infinite sequence  $\{X_n, n = 0, 1, \dots\}$  of random variables defined by

$$X_n = \begin{cases} 1 & \text{for success } s \text{ in the } n\text{th trial} \\ 0 & \text{for failure } f \text{ in the } n\text{th trial} \end{cases}$$

with probabilities

$$\begin{aligned} P(X_n = 1) &= p \\ P(X_n = 0) &= 1 - p \end{aligned} \quad (20.3.1)$$

Thus  $X_n$  represents a Bernoulli process, shown in Fig. 20.3.1.

We will now find the statistics of this process:

*Mean:*

$$E[X_n] = \mu_X = 1 \cdot P(X_n = 1) + 0 \cdot P(X_n = 0) = p \quad (20.3.2)$$

*Variance:*

$$\text{var}[X_n] = \sigma_X^2 = E[X^2] - \mu_X^2 = 1^2 \cdot p + 0^2 \cdot (1 - p) - p^2 = p(1 - p) \quad (20.3.3)$$

*Autocorrelation:*

$$E[X_{n_1}X_{n_2}] = R_X(n_1, n_2) = \begin{cases} 1^2 \cdot p + 0^2 \cdot (1 - p) = p, & n_1 - n_2 = 0 \\ E[X_{n_1}]E[X_{n_2}] = p^2, & n_1 - n_2 \neq 0 \end{cases} \quad (20.3.4)$$

*Autocovariance:*

$$\begin{aligned} C_X(n_1, n_2) &= R_X(n_1, n_2) - \mu_X^2 \\ &= \begin{cases} 1^2 \cdot p + 0^2 \cdot (1 - p) - p^2 = p(1 - p), & n_1 - n_2 = 0 \\ E[X_{n_1}]E[X_{n_2}] = p^2 = 0, & n_1 - n_2 \neq 0 \end{cases} \end{aligned} \quad (20.3.5)$$

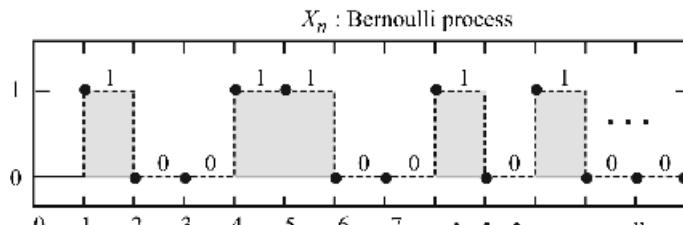


FIGURE 20.3.1

Normalized Autocovariance:

$$\rho_X(n_1, n_2) = \begin{cases} 1, & n_1 - n_2 = 0 \\ 0, & n_1 - n_2 \neq 0 \end{cases} \quad (20.3.6)$$

Since the mean is independent of time  $n$  and the autocorrelation depends only on the time difference  $(n_1 - n_2)$ , the process  $X_n$  is stationary.

### Binomial Process

We now define a new process  $Y_n$  of the sum of the random variables  $\{X_i\}$  in a Bernoulli process as discussed earlier. This is called the *binomial process*, given by

$$Y_n = \sum_{i=1}^n X_i, \quad n > 0 \quad (20.3.7)$$

This process, corresponding to the Bernoulli process of Fig. 20.3.1, is shown in Fig. 20.3.2 and is a discrete-state discrete-time (DSDT) process.

The probability of  $k$  successes in  $n$  trials has already been derived in Eq. (4.2.2) and is given by a binomial distribution:

$$\text{Probability}\{k \text{ successes in any sequence in } n \text{ trials}\} = P(Y_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

The statistics of the binomial process are:

*Mean.* Since the probability of each success is  $p$ , the mean value of  $Y_n$  can be obtained using the linearity of the expectation operator:

$$E[Y_n] = E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = np, \quad n > 0 \quad (20.3.8)$$

*Variance.* From the independence of  $X_i$ , the second moment of  $Y_n$  can be given by

$$\begin{aligned} E[Y_n^2] &= E\left[\sum_{i=1}^n X_i\right]^2 = \sum_{i=1}^n E[X_i^2] + \sum_{i=1}^n \sum_{j=1}^n E[X_i X_j] = np + n(n-1)p^2, \quad n > 0 \\ \sigma_Y^2 &= E[Y_n^2] - \{E[Y_n]\}^2 = np + n(n-1)p^2 - n^2p^2 = np(1-p), \quad n > 0 \end{aligned} \quad (20.3.9)$$

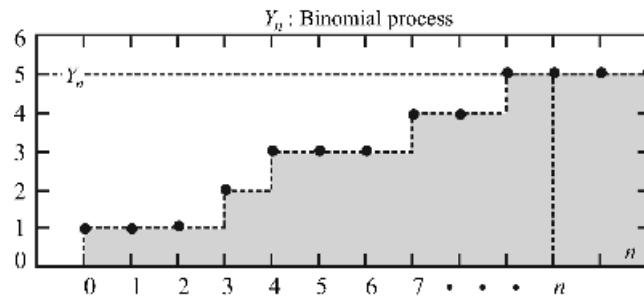


FIGURE 20.3.2

*Autocorrelation:* For  $n_1 \leq n_2$ , we have

$$\begin{aligned}
 E[Y_{n_1} Y_{n_2}] &= R_Y(n_1, n_2) = E\left[\sum_{j=1}^{n_2} \sum_{i=1}^{n_1} X_i X_j\right] \\
 &= E\left[\sum_{i=1}^{n_1} X_i^2\right] + \sum_{j=1}^{n_1} \sum_{\substack{i=1 \\ i \neq j}}^{n_1} E[X_i X_j] + \sum_{j=1}^{n_2-n_1} \sum_{\substack{i=1 \\ i \neq j}}^{n_1} E[X_i X_j] \\
 &= n_1 p + n_1(n_1 - 1)p^2 + n_1(n_2 - n_1)p^2
 \end{aligned}$$

or

$$R_Y(n_1, n_2) = n_1 p(1 - p) + n_1 n_2 p^2, \quad \text{for } n_1 \leq n_2 \quad (20.3.10)$$

Similarly for  $n_2 \leq n_1$  we have

$$R_Y(n_1, n_2) = n_2 p(1 - p) + n_1 n_2 p^2, \quad \text{for } n_2 \leq n_1 \quad (20.3.11)$$

Combining Eqs. (20.3.10) and (20.3.11), we obtain

$$R_Y(n_1, n_2) = p(1 - p) \min(n_1, n_2) + n_1 n_2 p^2, \quad n_1, n_2 > 0 \quad (20.3.12)$$

*Autocovariance:*

$$C_Y(n_1, n_2) = p(1 - p) \min(n_1, n_2), \quad n_1, n_2 > 0 \quad (20.3.13)$$

Unlike the Bernoulli process, the mean of the binomial process is dependent on time  $n$  and the autocorrelation function is dependent on both  $n_1$  and  $n_2$  and hence is a nonstationary process.

## 20.4 INDEPENDENT INCREMENT PROCESS

A random process  $X(t)$  is called an *independent increment process* if for a sequence of times  $t_1 < t_2 < \dots < t_n$  the increments  $X(t_1), \{X(t_2) - X(t_1)\}, \dots, \{X(t_n) - X(t_{n-1})\}$  of the process  $X(t)$  are a sequence of independent random variables:

$$\begin{aligned}
 E\{[X(t_i) - X(t_{i-1})][X(t_{i+1}) - X(t_i)]\} \\
 = E[X(t_i) - X(t_{i-1})] E[X(t_{i+1}) - X(t_i)]
 \end{aligned} \quad (20.4.1)$$

Further, if all order distributions of the increments  $\{X(t) - X(s)\}$ ,  $t < s$  are dependent only on the time difference  $\tau = t - s$ , then  $X(t)$  is called a *stationary independent increment process*.

We will show that a stationary independent increment process is not a stationary process, and its mean value and variance are given by

$$E[X(t)] = \mu_0 + \mu_1 t \quad (20.4.2a)$$

where  $\mu_0 = E[X(0)]$  and  $\mu_1 = E[X(1)] - E[X(0)]$  and

$$\text{var}[X(t)] = \sigma_{X(t)}^2 = \sigma_0^2 + \sigma_1^2 t \quad (20.4.2b)$$

where  $\sigma_0^2 = E[X(0) - \mu_0]^2$  and  $\sigma_1^2 = E[X(1) - \mu_1]^2 - \sigma_0^2$ .

We will verify Eq. (20.4.2a). Defining  $g(t) = E[X(t) - X(0)]$ , for any  $t$  and  $\tau$ , we have

$$\begin{aligned} g(t + \tau) &= E[X(t + \tau) - X(0)] = E\{[X(t + \tau) - X(\tau)] + [X(\tau) - X(0)]\} \\ &= E[X(t + \tau) - X(\tau)] + E[X(\tau) - X(0)] \end{aligned} \quad (20.4.3)$$

In Eq. (20.4.3)

$$E[X(\tau) - X(0)] = g(\tau) \quad \text{by definition} \quad (20.4.4)$$

and from the stationary independent increment property we can write

$$E[X(t + \tau) - X(\tau)] = E[X(t) - X(0)] = g(t) \quad (20.4.5)$$

Substituting Eqs. (20.4.5) and (20.4.4) in Eq. (20.4.3), we obtain

$$g(t + \tau) = g(t) + g(\tau) \quad (20.4.6)$$

Differentiating both sides of Eq. (20.4.6) independently with respect to  $t$  and  $\tau$ , we have

$$\frac{d}{dt}g(t + \tau) = \frac{d}{dt}g(t); \quad \frac{d}{d\tau}g(t + \tau) = \frac{d}{d\tau}g(\tau) \quad (20.4.7)$$

Hence  $g'(t) = g'(\tau)$  and in particular at  $\tau = 1$ ,  $g(\tau)|_{\tau=1} = E[X(1) - X(0)] = k_1$ . Substituting  $k_1$  for  $g'(\tau)$  and solving the differential equation  $g'(t) = k_1$ , we obtain the solution  $g(t) = k_1 t + k_0$ . Substitution of the initial condition  $g(0) = E[X(0) - X(0)] = 0$  in the solution results in  $g(t) = k_1 t$ , or

$$E[X(t) - X(0)] = E[X(1) - X(0)]t \quad (20.4.8)$$

$$E[X(t)] = \mu_0 + \mu_1 t \quad (20.4.2a)$$

thus verifying Eq. (20.4.2a).

In a similar manner we can verify Eq. (20.4.2b). Hence, a random process with stationary independent increments is nonstationary since the mean and variance are functions of time.

### **Uncorrelated and Orthogonal Increment Processes**

A random process  $X(t)$  is called an *uncorrelated increments process* if for a sequence of times  $t_1 < t_2 < t_3 < t_4$  the increments  $\{X(t_2) - X(t_1)\}$  and  $\{X(t_4) - X(t_3)\}$  of the process  $X(t)$  are uncorrelated random variables, or

$$E\{[X(t_2) - X(t_1)][X(t_4) - X(t_3)]\} = E[X(t_2) - X(t_1)]E[X(t_4) - X(t_3)] \quad (20.4.9)$$

Similarly, the random process  $X(t)$  is called an *orthogonal increments process* if

$$E\{[X(t_2) - X(t_1)][X(t_4) - X(t_3)]\} = 0 \quad (20.4.10)$$

The term *independent increment process* implies an uncorrelated increment process, but the converse is not true. If a process has uncorrelated increments, it need not necessarily have independent increments. Also, if a process  $X(t)$  has uncorrelated increments, then the process  $Y(t) = X(t) - E[X(t)]$  has orthogonal increments.

**Example 20.4.1 (Poisson Process)** A Poisson process  $N(t)$  with random arrival times  $\{S_k, k = 1, \dots, n, \dots\}$  and random interarrival times  $\{T_k, k = 1, \dots, n, \dots\}$  is an

independent increment process since

$$\begin{aligned} P\{T_{k-1} \leq s, T_k \leq t\} &= P\{[N(S_{k-1}) - N(S_{k-2}) = 1], [N(S_k) - N(S_{k-1}) = 1]\} \\ &= P[N(S_{k-1}) - N(S_{k-2}) = 1] P[N(S_k) - N(S_{k-1}) = 1] \\ &= (1 - e^{-\lambda t})(1 - e^{-\lambda t})u(s)u(t) \end{aligned}$$

The independent increments are stationary since

$$P[N(t) - N(s) = 1] = \lambda(t - s)e^{-\lambda(t-s)}$$

depends only on the time difference  $(t - s)$  and not the individual times  $t$  and  $s$ . The mean of the Poisson process can be obtained by substituting in Eq. (20.4.2a)

$$\mu_0 = E[N(0)] = \sum_{k=0}^{\infty} k e^{-\lambda \cdot 0} \frac{(\lambda \cdot 0)^k}{k!} = 0 \quad \text{and with} \quad E[N(1)] = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda$$

$\mu_1 = \lambda - 0 = \lambda$  and hence  $E[N(t)] = \lambda t$ , a result that has been obtained previously. The variance of the process can be obtained by substituting in Eq. (20.4.2b)  $\sigma_0^2 = 0$  and

$$\sigma_1^2 = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} - (\lambda t)^2 = \lambda \quad \text{and hence} \quad \text{var}[N(t)] = \lambda t$$

**Example 20.4.2 (Binomial Process)** A binomial process has been defined in Eq. (20.3.7) as

$$Y_n = \sum_{i=1}^n X_i, \quad n > 0$$

where  $n$  is the number of trials and  $X_i$  represents either 1 for success or 0 for failure at count  $i$ . Since each  $X_i$  is independent of  $X_j, j \neq i$ , we conclude that the process  $Y_n$  is an independent increment process. The independent increments are stationary because

$$P[(Y_n - Y_m = k)] = \binom{n-m}{k} p^k (1-p)^{(n-m)-k}, \quad n > m$$

is dependent only on the count difference  $(n - m)$  and not on individual counts  $n$  and  $m$ . In the binomial process  $\mu_0 = E[Y_0] = 0$  by definition and  $E[Y_1] = E[X_1] = p$ . Substitution of  $E[Y_1]$  in  $\mu_1$  results in  $\mu_1 = p - 0 = p$ , and hence, substituting  $\mu_0 = 0, \mu_1 = p$ , and  $t = n$  in Eq. (20.4.2a), we obtain  $E[Y_n] = np$ , a result obtained previously.

In a similar manner the variance of the process  $Y_n$  can be obtained by substituting in Eq. (20.4.2b)  $\sigma_0^2$  and  $\sigma_1^2 = p(1-p)$ , resulting in  $\text{var}[Y_n] = np(1-p)$ , another result obtained previously.

### Elementary Random-Walk Process

The Bernoulli process considered earlier is now modified as shown in Fig. 20.4.1 so that the failure is now represented by  $-1$  instead of  $0$ . Under this modification we define a random variable  $Z_n$

$$Z_n = \begin{cases} 1 & \text{for success in the } n\text{th trial} \\ -1 & \text{for failure in the } n\text{th trial} \end{cases} \quad (20.4.11a)$$

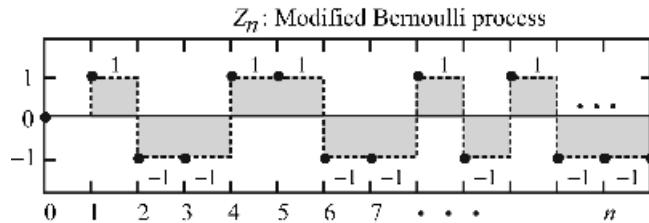


FIGURE 20.4.1

with probabilities

$$\begin{aligned} P(Z_n = 1) &= p \\ P(Z_n = -1) &= 1 - p = q \end{aligned} \quad (20.4.11.b)$$

If  $p = \frac{1}{2}$ , this random process is called a *symmetric random walk*.

The mean and variance of  $Z_n$  are

*Mean:*

$$\mu_Z = E[Z_n] = 1 \cdot P(X_n = 1) - 1 \cdot P(X_n = -1) = p - (1 - p) = 2p - 1 \quad (20.4.12)$$

*Second Moment:*

$$E[Z_n^2] = 1 \cdot p + 1 \cdot (1 - p) = 1 \quad (20.4.13)$$

*Variance:*

$$\sigma_Z^2 = \text{var}[Z_n] = 1^2 \cdot p + (-1)^2 \cdot (1 - p) - (2p - 1)^2 = 4p(1 - p) \quad (20.4.14)$$

In much the same way as we did in Eq. (20.3.7), we define a new process  $W_n$  as the sum of the random variables  $\{Z_i\}$  in the modified Bernoulli process. This is called an *elementary random-walk process*, given by

$$W_n = \sum_{i=1}^n Z_i, \quad n > 0 \quad (20.4.15)$$

This process, shown in Fig. 20.4.2, is also a discrete-state discrete-time (DSDT) process. The probability that there are  $k$  successes and  $(n - k)$  failures in  $n$  trials is again a binomial

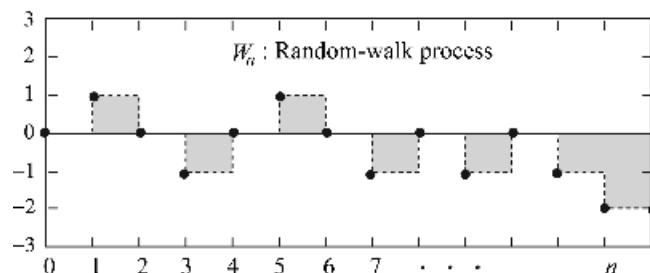


FIGURE 20.4.2

distribution given by

$$\text{Probability}\{k \text{ successes in any sequence in } n \text{ trials}\} = P(W_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

The statistics of this elementary random walk are

*Mean:*

$$\mu_W = E[W_n] = E\left[\sum_{i=1}^n Z_i\right] = \sum_{i=1}^n E[Z_i] = n(2p - 1), \quad n > 0 \quad (20.4.16)$$

*Variance.* From the independence of  $Z_i$ , the second moment of  $W_n$  can be given by

$$\begin{aligned} E[W_n^2] &= E\left[\sum_{i=1}^n Z_i\right]^2 = \sum_{i=1}^n E[Z_i^2] + \sum_{i=1}^n \sum_{j=1}^n E[Z_i Z_j] \\ &= n + n(n-1)(2p-1)^2 \quad n > 0 \end{aligned}$$

Hence

$$\begin{aligned} \sigma_W^2 &= E[W_n^2] - \{E[W_n]\}^2 = n + n(n-1)(2p-1)^2 \\ &\quad - n^2(2p-1)^2 = 4np(1-p), \quad n > 0 \end{aligned} \quad (20.4.17)$$

*Autocorrelation.* For  $n_1 \leq n_2$ , we obtain

$$\begin{aligned} E[W_{n_1} W_{n_2}] &= R_W(n_1, n_2) = E\left[\sum_{j=1}^{n_2} \sum_{i=1}^{n_1} Z_i Z_j\right] \\ &= E\left[\sum_{i=1}^{n_1} Z_i^2\right] + \sum_{j=1}^{n_1} \sum_{\substack{j=1 \\ i \neq j}}^{n_1} E[Z_i Z_j] + \sum_{j=1}^{n_2-n_1} \sum_{\substack{i=1 \\ i \neq j}}^{n_1} E[Z_i Z_j] \end{aligned} \quad (20.4.18)$$

Substituting for the expectation operators from Eqs. (20.4.12) and (20.4.13), we obtain  $E[W_{n_1} W_{n_2}] = 4p(1-p)n_1 + n_1 n_2 (2p-1)^2$  with a similar expression for  $n_2 \leq n_1$  given by  $E[W_{n_1} W_{n_2}] = 4p(1-p)n_2 + n_1 n_2 (2p-1)^2$ . Combining these two equations, we have

$$E[W_{n_1} W_{n_2}] = R_W(n_1, n_2) = 4p(1-p) \min(n_1, n_2) + n_1 n_2 (2p-1)^2 \quad (20.4.19)$$

*Autocovariance:*

$$C_W(n_1, n_2) = 4p(1-p) \min(n_1, n_2) \quad (20.4.20)$$

Since the mean is a function of time  $n$  and the autocorrelation is a function of both  $n_1$  and  $n_2$ , we conclude that the random-walk process is also a nonstationary process.

**Example 20.4.3 (Random-Walk Process)** Since the random-walk process is a takeoff from the Bernoulli process and with distribution function similar to the binomial process, we can conclude that this process is also a stationary independent increment process. We will show that the random-walk process satisfies Eqs. (20.4.2a) and (20.4.2b). In this process  $\mu_0 = E[W_0] = 0$  by definition and

$\mu_1 = E[W_1] = E[Z_1] = 2p - 1$ . Substituting for  $\mu_0 = 0$ ,  $\mu_1 = (2p - 1)$  and  $t = n$  in Eq. (20.4.2a), we obtain  $E[W_n] = \mu_W = n(2p - 1)$ , which is the same as Eq. (20.4.16). In a similar manner,  $\sigma_0^2 = E[W_0 - \mu_0]^2 = 0$  and  $\sigma_1^2 = E[W_1^2] - \mu_1^2 = 1 - (2p - 1)^2 = 4p(1 - p)$ . Hence  $\text{var}[W_n] = 4np(1 - p)$ , which is the same as Eq. (20.4.17).

## 20.5 RANDOM-WALK PROCESS

An elementary random walk has been discussed in the previous section. We will generalize this concept a little further.

A particle is initially at rest at  $i = 0$  with  $W_0 = 0$ . At times  $\{i = 1, 2, \dots, n, \dots\}$  it undergoes jumps  $\{Z_i, i = 1, 2, \dots, n, \dots\}$ , where  $\{Z_i\}$  is a sequence of independent identically distributed random variables with probabilities

$$\left. \begin{array}{l} P(Z_i = +1) = p \\ P(Z_i = -1) = q \\ P(Z_i = 0) = 1 - p - q \end{array} \right\} i = 1, \dots, n, \dots, P(Z_i Z_j) = P(Z_i)P(Z_j), i \neq j \quad (20.5.1)$$

If  $W_n$  is the position of the particle after  $n$  jumps, then it can be written in terms of  $Z_i$  as

$$W_n = \sum_{i=1}^n Z_i, \quad n > 0, \quad \text{and} \quad W_0 = 0 \quad (20.5.2)$$

The random processes  $Z_i$  and  $W_n$  are shown in Fig. 20.5.1.

If the particle tends to move indefinitely in either the positive or negative direction, then this random walk is called *unrestricted*. On the other hand, the particle may be restricted to  $+a$  in the positive direction and  $-b$  in the negative direction where  $a$  and  $b$  are positive, in which case  $a$  and  $b$  are called *absorbing barriers*. Let the position of the particle at time  $n$  be equal to  $k$  where  $k = 0, \pm 1, \pm 2, \dots, \pm n$ . We have to find the probability  $P(W_n = k)$ . We observe that to reach point  $k$ , the particle undergoes  $r_1$  positive jumps,  $r_2$  negative jumps, and  $r_3$  zero jumps with the following constraining conditions:

$$r_1 + r_2 + r_3 = n, \quad k = r_1 - r_2, \quad r_3 = n - r_1 - r_2 \quad (20.5.3)$$

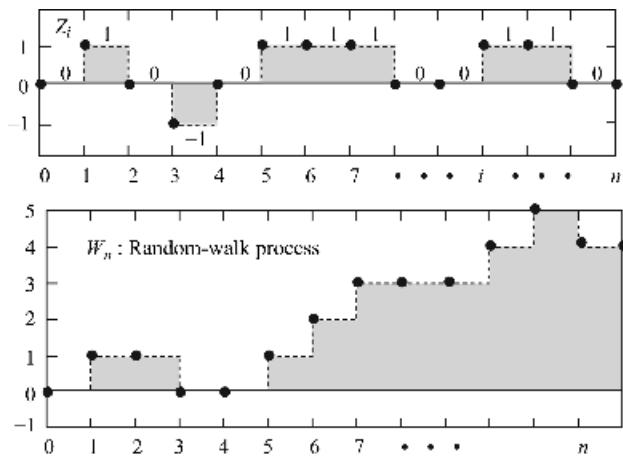


FIGURE 20.5.1

With  $r_1$  as an independent variable  $P(W_{nr_1} = k)$  is a multinomial distribution given by

$$P(W_{nr_1} = k) = \frac{n!}{r_1!r_2!r_3!} p^{r_1} q^{r_2} (1-p-q)^{r_3} \quad (20.5.4)$$

Substituting the constraints of Eq. (20.5.3) in Eq. (20.5.4), we obtain

$$P(W_{nr_1} = k) = \frac{n!}{r_1!(r_1 - k)!(n - 2r_1 + k)!} p^{r_1} q^{r_1 - k} (1-p-q)^{n-2r_1+k} \quad (20.5.5)$$

From Eq. (20.5.5), the constraints for  $r_1$  are  $r_1 \geq k$  and  $r_1 \leq (n+k)/2$ . In other words,

$$k \leq r \leq \left\lfloor \frac{n+k}{2} \right\rfloor \quad (20.5.6)$$

where  $\left\lfloor \frac{n+k}{2} \right\rfloor$  represents the integer closest to but not exceeding  $(n+k)/2$ .

Since the events  $\{W_{nr_1} = k\}$  for each  $r_1$  are mutually exclusive, we can sum over all possible values of  $r_1$  in Eq. (20.5.6) and write  $P(W_n = k)$  as follows:

$$\begin{aligned} P(W_n = k) &= \sum_{r_1} P(W_{nr_1} = k) \\ &= \sum_{r_1=k}^{\lfloor(n+k)/2\rfloor} \frac{n!}{r_1!(r_1 - k)!(n - 2r_1 + k)!} \\ &\quad \times p^{r_1} q^{r_1 - k} (1-p-q)^{n-2r_1+k} \end{aligned} \quad (20.5.7)$$

The probability  $P(W_n = k)$  can be more easily determined by using the generating function (GF) defined in Eq. (11.3.1). The generating function for the jumps  $Z_i$  is

$$\begin{aligned} P_Z(z) = E[z^{Z_i}] &= \sum_k z^k P(Z_i = k) = pz + qz^{-1} + (1-p-q)z^0 \\ &= \frac{pz^2 + (1-p-q)z + q}{z} \end{aligned} \quad (20.5.8)$$

Since  $W_n$  is the sum of  $n$  independent jumps, the GF for  $W_n$  is the  $n$ -fold multiplication of the GF  $P_Z(z)$ , or

$$P_{W_n}(z) = E[z^{W_n}] = [P_Z(z)]^n = \left( \frac{pz^2 + (1-p-q)z + q}{z} \right)^n, \quad n > 0 \quad (20.5.9)$$

and since  $W_0 = 0$ , the GF for  $W_0$  can be defined as follows:

$$P_{W_0}(z) = E[z^{W_0}] = [P_Z(z)]^0 = 1 \quad (20.5.10)$$

With  $P_{W_0}(z) = 1$  from Eq. (20.5.10), we can take the  $s$  transform of  $P_{W_n}(z) = [P_Z(z)]^n$  and write

$$\begin{aligned} G(z, s) &= \sum_{n=0}^{\infty} s^n [P_Z(z)]^n = \frac{1}{1 - sP_Z(z)} \\ &= \frac{z}{z - s(pz^2 + (1-p-q)z + q)} \quad \text{for } |sP_Z(z)| < 1 \end{aligned} \quad (20.5.11)$$

The probability  $P(W_n = k)$  is the coefficient of  $s^k z^k$  in Eq. (20.5.11).

**Example 20.5.1** In a random-walk process  $p = \frac{1}{2}$  and  $q = \frac{3}{10}$ . We will determine the probability  $P(W_n = k)$  of the particle being at  $k = 3$  in  $n = 10$  trials. We will solve this problem using the direct approach of Eq. (20.5.5) and the generating function approach of Eq. (20.5.11).

*Direct Approach.* In Eq. (20.5.5) the known quantities are  $n = 10$ ,  $k = 3$ ,  $p = \frac{1}{2}$ , and  $q = \frac{3}{10}$ . We have to find  $P(W_{10} = 3)$  for all possible values of  $r_1$  the positive jumps. The values of  $r_1$  from Eq. (20.5.6) are 3,4,5,6. From Eq. (20.5.7)  $P(W_{10} = 3)$  is given by

$$P(W_{10} = 3) = \sum_{r_1=3}^6 \frac{10!}{r_1!(r_1-3)!(13-2r_1)!} \frac{1^{r_1}}{2} \frac{3^{r_1-3}}{10} \frac{1^{13-2r_1}}{5}$$

or

$$P(W_{10} = 3) = 0.000192 + 0.00756 + 0.0567 + 0.070875 = 0.135327.$$

*Generating Function Approach.* Substituting the values of  $n = 10$ ,  $p = \frac{1}{2}$ , and  $q = \frac{3}{10}$  in Eq. (20.5.11), we obtain

$$G(z, s) = \frac{z}{z - s\left(\frac{1}{2}z^2 + \frac{1}{5}z + \frac{3}{10}\right)} = \frac{1}{1 - \frac{s(5z^2 + 2z + 3)}{10z}}$$

The required probability  $P(W_{10} = 3)$  is the coefficient of  $s^{10}z^3$  in the power-series expansion of  $G(z, s)$ . The power-series expansion for  $G(z, s)$  in terms of  $s$  is given by

$$G(z, s) = \sum_{k=0}^{\infty} \left[ \frac{s(5z^2 + 2z + 3)}{10z} \right]^k$$

and the coefficient of  $s^{10}$  is

$$\text{Coefficient of } s^{10} = \left[ \frac{5z^2 + 2z + 3}{10z} \right]^{10}$$

This coefficient of  $s^{10}$  will be expanded as a power series in  $z$ . Expanding the preceding equation, the coefficient of  $z^3$  is

$$\text{Coefficient of } z^3 = \frac{135,327}{1,000,000}$$

Hence,  $P(W_{10} = 3) = \frac{135,327}{1,000,000}$ , agreeing with the previous result obtained from the direct method.

### Random Walk with Two Absorbing Barriers a and -b

We will now consider a random walk with two absorbing barriers. The particle starts at an initial value of  $j$  and terminates at either  $a$  or  $-b$  with  $a, b > 0$ . The initial value of the random walk is

$$W_0 = j, -b < j < a$$

It can be shown from heuristic arguments that as  $n \rightarrow \infty$ , the ultimate probability of absorption at either  $a$  or  $-b$  is certain. We will define the probability of absorption as

$$P_a(n, j) = P\{\text{absorption at } a \text{ at time } n \text{ starting from } j\}, \quad n = 1, 2, \dots \quad (20.5.12)$$

The probability  $P_a(n, j)$  also represents the first passage of the particle into  $a$  at time  $n$  starting from  $j$ , without having entered either  $-b$  or  $a$  at any previous times:

$$\begin{aligned} P_a(n, j) &= P\{-b < W_1 < a, \dots, -b < W_{n-1} < a, \\ &W_n = a \mid W_0 = j\}, \quad n = 1, 2, \dots \end{aligned} \quad (20.5.13)$$

At  $n = 0$ , we have the initial probabilities

$$P_a(0, j) = \begin{cases} 1, & \text{if } j = a \text{ (absorption occurs at } a \text{ at 0 when it starts at } a) \\ 0, & \text{if } j \neq a \text{ (absorption cannot occur at } a \text{ at 0 not starting at } a) \end{cases} \quad (20.5.14)$$

and at  $j = a$  and  $j = -b$  the end probabilities for  $n = 1, 2, \dots$ , are

$$\begin{aligned} P_a(n, a) &= 0, \quad n = 1, 2, \dots \\ P_a(n, -b) &= 0, \quad n = 0, 1, 2, \dots \end{aligned} \quad (20.5.15)$$

We define for convenience an event  $A_n$  as

$$A_n = \{\text{absorption at } a \text{ at time } n\} \text{ and } P\{A_n \mid W_0 = j\} = P_a(n, j) \quad (20.5.16)$$

Starting from  $j$  at  $n = 0$ , after the first step at  $n = 1$ , absorption at  $a$  at time  $(n - 1)$  will occur for the cases when the state can be at either  $(j + 1)$  with probability  $p$  or  $(j - 1)$  with probability  $q$ , or remain at  $j$  with probability  $(1 - p - q)$ . Since these are mutually exclusive events, we have

$$\begin{aligned} P\{A_n \mid W_0 = j\} &= pP\{A_{n-1} \mid W_0 = j + 1\} + qP\{A_{n-1} \mid W_0 = j - 1\} \\ &\quad + (1 - p - q)P\{A_{n-1} \mid W_0 = j\} \end{aligned} \quad (20.5.17)$$

Substituting the definition of  $P_a(n, j)$  from Eq. (20.5.16) in Eq. (20.5.17), we obtain

$$\begin{aligned} P_a(n, j) &= pP_a(n - 1, j + 1) + qP_a(n - 1, j - 1) \\ &\quad + (1 - p - q)P_a(n - 1, j), \quad -b < j < a, \quad n = 1, 2, \dots \end{aligned} \quad (20.5.18)$$

Equation (20.5.15) is a difference equation in the two variables  $n$  and  $j$ . We can eliminate the time variable  $n$  by taking the  $z$  transform of  $P_a(n, j)$ , or

$$G_a(z, j) = \sum_{n=0}^{\infty} z^n P_a(n, j) \quad (20.5.19)$$

with the initial probability  $P_a(0, j)$  given by Eq. (20.5.14). Multiplying Eq. (20.5.18) by  $z^n$  and summing over all  $n$ , we obtain

$$G_a(z, j) = z\{pG_a(z, j + 1) + qG_a(z, j - 1) + (1 - p - q)G_a(z, j)\} \quad (20.5.20)$$

Equation (20.5.20) is a second-order difference equation in  $j$ . The end conditions from Eqs. (20.5.14) and (20.5.15) are

$$G_a(z, a) = 1, \quad G_{-b}(z, a) = 0 \quad (20.5.21)$$

We can take another  $z$  transform for the variable  $j$ . Since  $j$  is bounded by  $a$ , we can solve the difference equation [Eq. (20.5.20)] more easily by assuming a solution of the form  $G_a(z, j) = \lambda^j(z)$ . Thus

$$-\lambda^j(z) + z\{p\lambda^{j+1}(z) + q\lambda^{j-1}(z) + (1 - p - q)\lambda^j(z)\} = 0$$

or

$$zp\lambda^2(z) - [1 - z(1 - p - q)]\lambda(z) + zq = 0 \quad (20.5.22)$$

The solutions to the quadratic equation [Eq. (20.5.22)] are

$$\begin{aligned} \lambda_1(z) &= \frac{1 - z(1 - p - q) + \sqrt{[1 - z(1 - p - q)]^2 - 4z^2pq}}{2zp} \\ \lambda_2(z) &= \frac{1 - z(1 - p - q) - \sqrt{[1 - z(1 - p - q)]^2 - 4z^2pq}}{2zp} \end{aligned} \quad (20.5.23)$$

with the relationship between the roots  $\lambda_1(z)\lambda_2(z) = (zq/zp) = (q/p)$  and  $\lambda_1(z) > \lambda_2(z)$ . We will assume that  $z$  is positive and real. If  $\lambda_1(z)$  and  $\lambda_2(z)$  are to be real, then the discriminant in Eq. (20.5.23) must be positive. Hence we have the following inequality

$$[1 - z(1 - p - q)]^2 - 4z^2pq > 0$$

or

$$z^2[(1 - p - q)^2 - 4pq] - 2z(1 - p - q) + 1 > 0 \quad (20.5.24)$$

The factors of the preceding quadratic equation in  $z$  are

$$\begin{aligned} z_1, z_2 &= \frac{2(1 - p - q) \pm \sqrt{4(1 - p - q)^2 - 4[(1 - p - q)^2 - 4pq]}}{2[(1 - p - q)^2 - 4pq]} \\ &= \frac{1 - p - q \pm 2\sqrt{pq}}{(1 - p - q)^2 - 4pq} = \frac{1}{1 - p - q \pm 2\sqrt{pq}} = \frac{1}{1 - (\sqrt{p} \pm \sqrt{q})^2} \end{aligned}$$

or

$$\left[ z_1 - \frac{1}{1 - (\sqrt{p} - \sqrt{q})^2} \right] \left[ z_2 - \frac{1}{1 - (\sqrt{p} + \sqrt{q})^2} \right] > 0 \quad (20.5.25)$$

If  $z$  is positive, the inequality of Eq. (20.5.25) will be satisfied only when both

$$z_1 < \frac{1}{1 - (\sqrt{p} - \sqrt{q})^2} \quad \text{and} \quad z_2 < \frac{1}{1 - (\sqrt{p} + \sqrt{q})^2}$$

Since  $(\sqrt{p} + \sqrt{q})^2 > (\sqrt{p} - \sqrt{q})^2$ , Eq. (20.5.25) is equivalent to

$$0 < z < \frac{1}{1 - (\sqrt{p} + \sqrt{q})^2}$$

The general solution of the difference Eq. (20.5.2) can be given by

$$G_a(z, j) = c_1(z)\lambda_1^j(z) + c_2(z)\lambda_2^j(z) \quad (20.5.26)$$

where  $c_1(z)$  and  $c_2(z)$  are arbitrary functions of  $z$ . Substituting the initial conditions of Eq. (20.5.21) in Eq. (20.5.26), we can solve for  $c_1(z)$  and  $c_2(z)$  as follows:

$$c_1(z) = \frac{-\lambda_1^b(z)\lambda_2^{-b}(z)}{\lambda_2^a(z) - \lambda_2^{-b}(z)\lambda_1^{a+b}(z)} \quad \text{and} \quad c_2(z) = \frac{1}{\lambda_2^a(z) - \lambda_2^{-b}(z)\lambda_1^{a+b}(z)} \quad (20.5.27)$$

Substituting Eq. (20.5.27) into Eq. (20.5.36), we obtain the solution  $G_a(z, j)$  as

$$G_a(z, j) = \frac{-\lambda_1^b(z)\lambda_2^{-b}(z)\lambda_1^j(z) + \lambda_2^j(z)}{\lambda_2^a(z) - \lambda_2^{-b}(z)\lambda_1^{a+b}(z)} = \frac{\lambda_1^{j+b}(z) - \lambda_2^{j+b}(z)}{\lambda_1^{a+b}(z) - \lambda_2^{a+b}(z)} \quad (20.5.28)$$

To obtain the probability of absorption at  $a$  at time  $n$  starting from  $j$ , we have to express Eq. (20.5.8) as a power series in  $z$ , and the coefficient of  $z^n$  will yield  $P_a(n, j)$ . These involved calculations can be found in Refs. 13 and 17. However, the events of *eventual* absorption at  $a$  or  $-b$  are mutually exclusive, and their probabilities can be calculated easily. We will find the probability of eventual absorption at  $a$  starting from  $j$ . Equation (20.5.28) is rewritten as

$$G_a(z, j) = \frac{\lambda_1^{j+b}(z) - \lambda_2^{j+b}(z)}{\lambda_1^{a+b}(z) - \lambda_2^{a+b}(z)} \quad (20.5.29)$$

The probability of eventual absorption at  $a$  starting at  $j$  will be given by the sum of the probabilities  $P_a(n, j)$  over all  $n$  and can be obtained by substituting  $z = 1$  in Eq. (20.5.29):

$$\begin{aligned} P_a(j) &= P\{\text{absorption at } a \text{ starting at } j\} = \sum_{n=0}^{\infty} P_a(n, j) \\ &= G_a(1, j) = \frac{\lambda_1^{j+b}(1) - \lambda_2^{j+b}(1)}{\lambda_1^{a+b}(1) - \lambda_2^{a+b}(1)} \end{aligned} \quad (20.5.30)$$

Substituting  $z = 1$  in the roots  $\lambda_1(z)$  and  $\lambda_2(z)$ , since  $\lambda_1(z) > \lambda_2(z)$ , we have

$$\begin{aligned} p < q, \lambda_1(1) &= \frac{(p+q) - (p-q)}{2p} = \frac{q}{p} > \lambda_2(1) = \frac{(p+q) + (p-q)}{2p} = 1 \\ p > q, \lambda_1(1) &= \frac{(p+q) + (p-q)}{2p} = 1 > \lambda_2(1) = \frac{(p+q) - (p-q)}{2p} = \frac{q}{p} \\ p = q, \lambda_1(1) &= \lambda_2(1) = 1 \end{aligned} \quad (20.5.31)$$

Substituting Eq. (20.5.31) in Eq. (20.5.30), we obtain

$$P_a(j) = G(1, j) = p^{a-j} \frac{p^{j+b} - q^{j+b}}{p^{a+b} - q^{a+b}}, \quad p \neq q \quad (20.5.32a)$$

Using l'Hôpital's rule when  $p = q$ , we have

$$P_a(j) = \frac{j+b}{a+b}, \quad p = q \quad (20.5.32b)$$

In a similar manner we can obtain the solution  $G_{-b}(z, j)$  for the absorption at  $-b$  starting from  $j$ . The result is

$$G_{-b}(z, j) = \frac{\lambda_1^{j-a}(z) - \lambda_2^{j-a}(z)}{\lambda_1^{-(a+b)}(z) - \lambda_2^{-(a+b)}(z)} \quad (20.5.33)$$

The probability of absorption at  $-b$  starting at  $j$  is given by

$$G_{-b}(1, j) = P_{-b}(j) = 1 - P_a(j) = 1 - G_a(1, j)$$

or

$$P_{-b}(0) = G_{-b}(1, 0) \begin{cases} q^{b+j} \frac{p^{a-j} - q^{a-j}}{p^{a+b} - q^{a+b}}, & p \neq q \\ \frac{a-j}{a+b}, & p = q \end{cases} \quad (20.5.34)$$

**Example 20.5.2 (Gambler's Ruin)** This is a classic problem that has wide applications in testing the efficacy of drugs, insurance payoffs, and so on. Two gamblers  $A$  and  $B$  wager on the flip of a biased coin. If heads show up,  $A$  wins 1 dollar from  $B$ , and if tails show up,  $A$  loses one dollar to  $B$ . The probability of heads is  $p$  and tails  $q = 1 - p$ . Initially  $A$  starts with  $k$  dollars and  $B$  with  $N - k$  dollars. The game stops when either  $A$  loses all the money or is ruined or  $A$  wins all the money and  $B$  is ruined. We want to calculate the probabilities of  $A$  either winning all the money or losing all the money.

This problem can be cast in the random-walk framework with absorbing barriers. We will solve this problem from fundamentals and show that we get the same result from using generating functions. Each toss is an i.i.d. random variable  $Z_i$  with probabilities

$$\begin{aligned} P(Z_i = +1) &= p \\ P(Z_i = -1) &= q = (1 - p) \end{aligned} \quad \left\{ \begin{aligned} i &= 1, \dots, n, \dots, \\ P(Z_i Z_j) &= P(Z_i)P(Z_j), \quad i \neq j \end{aligned} \right.$$

and the cumulative sum  $W_n$  is given by Eq. (20.5.2) with initial value  $W_0 = k$ . Or,

$$W_n = \sum_{i=1}^n Z_i, \quad n > 0 \quad \text{and} \quad W_0 = k \quad (20.5.35)$$

The game terminates when  $W_n = N$  or  $W_n = 0$ . The absorbing barriers are either  $N$  or  $0$ . We define an event  $A_n$  as absorption at  $N$  at time  $n$  and the probability of absorption at  $N$  at time  $n$  on condition that the starting state is  $k$ , as follows:

$$P\{A_n | W_0 = k\} = P_N(n, k)$$

Starting from  $k$  at  $n = 0$ , after the first step at  $n = 1$ , absorption at  $N$  at time  $(n - 1)$  will occur for the cases when the state can be at either  $(k + 1)$  with probability  $p$  or  $(k - 1)$  with probability  $q$ . Since these are mutually exclusive events, we have

$$P\{A_n | W_0 = k\} = pP\{A_{n-1} | W_0 = k + 1\} + qP\{A_{n-1} | W_0 = k - 1\}$$

and in terms of the defined probability  $P_N(n, k)$

$$P_N(n, k) = pP_N(n - 1, k + 1) + qP_N(n - 1, k - 1), \quad k = 1, 2, \dots, N - 1$$

with end conditions  $P_N(n, N) = 1$  and  $P_N(n, 0) = 0$ , for  $n = 1, 2, \dots, N - 1$ .

We are interested in the eventual probability of absorption in  $N$  and hence we can eliminate  $n$  from the equation above and write

$$P_N(k) = pP_N(k + 1) + qP_N(k - 1), \quad k = 1, 2, \dots, N - 1$$

with end conditions  $P_N(N) = 1$  and  $P_N(0) = 0$ ,

Since  $p + q = 1$ , the preceding equation can be rewritten as

$$P_N(k)(p + q) = pP_N(k + 1) + qP_N(k - 1)$$

and rearranging terms, we have

$$P_N(k+1) - P_N(k) = \frac{q}{p}[P_N(k) - P_N(k-1)], \quad k = 1, 2, \dots, N-1$$

We solve this equation recursively as follows from the initial condition  $P_N(0) = 0$

$$k = 1: \quad P_N(2) - P_N(1) = \frac{q}{p}[P_N(1) - P_N(0)] = \frac{q}{p}P_N(1)$$

$$k = 2: \quad P_N(3) - P_N(2) = \frac{q}{p}[P_N(2) - P_N(1)] = \left(\frac{q}{p}\right)^2 P_N(1)$$

...

$$k = k-1: \quad P_N(k) - P_N(k-1) = \frac{q}{p}[P_N(k-1) - P_N(k-2)]$$

$$= \left(\frac{q}{p}\right)^{k-1} P_N(1)$$

...

$$k = N-1: \quad P_N(N) - P_N(N-1) = \frac{q}{p}[P_N(N-1) - P_N(N-2)]$$

$$= \left(\frac{q}{p}\right)^{N-1} P_N(1)$$

Adding the first  $k-1$  equations, we obtain

$$P_N(k) - P_N(1) = P_N(1) \left[ \frac{q}{p} + \left(\frac{q}{p}\right)^2 + \dots + \left(\frac{q}{p}\right)^{k-1} \right]$$

or

$$\begin{aligned} P_N(k) &= P_N(1) \left[ 1 + \frac{q}{p} + \left(\frac{q}{p}\right)^2 + \dots + \left(\frac{q}{p}\right)^{k-1} \right] \\ &= P_N(1) \frac{1 - \left(\frac{q}{p}\right)^k}{1 - \frac{q}{p}}, \quad p \neq q \end{aligned}$$

$$P_N(k) = P_N(1)k, \quad p = q$$

Substituting  $k = N$  in the equation above, we can obtain  $P_N(1)$  by applying the boundary condition  $P_N(N) = 1$  as shown below:

$$P_N(1) = \frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N}, \quad p \neq q \quad \text{and} \quad P_N(1) = \frac{1}{N}, \quad p = q$$

Substituting for  $P_N(1)$  in the equation for  $P_N(k)$ , we obtain

$$\begin{aligned} P_N(k) &= \frac{1 - \left(\frac{q}{p}\right)^k}{1 - \frac{q}{p}} \frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N} = \frac{1 - \left(\frac{q}{p}\right)^k}{1 - \left(\frac{q}{p}\right)^N}, \quad p \neq q \\ &= \frac{k}{N}, \quad p = q \end{aligned} \quad (20.5.36)$$

$P_N(k)$  is the probability that  $A$  starting from  $k$  dollars will eventually reach  $N$  dollars. The probability of ruin  $P_0(k)$  of  $A$  is

$$\begin{aligned} P_0(k) &= 1 - P_N(k) = \frac{1 - \left(\frac{p}{q}\right)^{N-k}}{1 - \left(\frac{p}{q}\right)^N}, \quad p \neq q \\ &= \frac{N - k}{N}, \quad p = q \end{aligned} \quad (20.5.37)$$

The probability of ruin  $P_0(k)$  can be easily obtained from Eq. (20.5.34) by substituting for the absorbing barriers  $b = 0$ ,  $a = N$  and  $j = k$ , which results in Eq. (20.5.37).

*Discussion of Results.* We will investigate what happens when  $A$  plays against an infinitely rich adversary  $B$ . If  $p > q$ , then  $(q/p) < 1$ , and as  $N \rightarrow \infty$  in Eq. (20.5.36), we have

$$p > q: \quad \lim_{N \rightarrow \infty} P_N(k) = \lim_{N \rightarrow \infty} \frac{1 - \left(\frac{q}{p}\right)^k}{1 - \left(\frac{q}{p}\right)^N} = 1 - \left(\frac{q}{p}\right)^k > 0 \quad (20.5.38)$$

This results in a finite probability that  $A$  will get infinitely rich. On the other hand, if  $p \leq q$ , then  $(q/p) > 1$ , and as  $N \rightarrow \infty$  in Eq. (20.5.36), we have

$$p < q: \quad \lim_{N \rightarrow \infty} P_N(k) = \lim_{N \rightarrow \infty} \frac{1 - \left(\frac{q}{p}\right)^k}{1 - \left(\frac{q}{p}\right)^N} = \lim_{N \rightarrow \infty} \left(\frac{p}{q}\right)^{N-k} \rightarrow 0 \quad (20.5.39)$$

$$p = q: \quad \lim_{N \rightarrow \infty} P_N(k) = \lim_{N \rightarrow \infty} \frac{k}{N} \rightarrow 0$$

This results in  $A$  getting ruined with probability 1 if  $p \leq q$ . The interesting part is that even if it is a fair game with equal probability of winning, a rich player will always ruin a poorer player. On the other hand, if  $p > q$ ,  $A$  can get infinitely rich even against an infinitely rich adversary. The gambling halls all over the world make sure that the probability of any player winning is always less than 0.5!

**Example 20.5.3** An insurance company earns \$1 per day from premiums with probability  $p$ . However, on each day it pays out a claim of \$2 each day independent of the past with probability  $q = 1 - p$ . When a claim is paid, \$2 is removed from the reserve

of money of  $\$k$ . Hence we can assume that on the  $i$ th day, the income for that day is a random variable  $Z_i$  equal to 1 with probability  $p$  and  $-1$  with probability  $q$  as in Example 20.5.2. The cumulative income for the company is  $W_n$  as in Eq. (20.5.35). If it starts initially with a reserve capital of  $\$k$ , then we have to find the probability that it will eventually run out of money. The solution is a reformulation of Eqs. (20.5.38) and (20.5.39). If  $p > q$ , then the probability that it will not run out of money is given by  $1 - (q/p)^k > 0$ . If  $k$  is large, then it has very high probability of staying in business. On the other hand, if  $p \leq q$ , it will always run out of money as indicated in Eq. (20.5.39).

## 20.6 GAUSSIAN PROCESS

Many of the physically occurring random phenomena can be modeled as a Gaussian process. Further, according to the central limit theorem, random variables characterized individually by arbitrary distributions can be modeled as Gaussian if a sufficient number of them can be added. The distribution of a more commonly used white noise is Gaussian.

We have already stated (in Section 6.4) that a random variable  $X$  is Gaussian-distributed if its density function is given by

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} e^{-(1/2)[(x-\mu_X)/\sigma_X]^2} \quad (20.6.1)$$

where  $\mu_X$  is the mean value and  $\sigma_X^2$  is the variance.

A second-order random process  $X(t)$  is a Gaussian random process if for any collection of times  $\{t_1, t_2, \dots, t_n\}$ , the random variables,  $X_1 = X(t_1), X_2 = X(t_2), \dots, X_n = X(t_n)$  are jointly Gaussian distributed random variables for all  $n$  with the joint density given by

$$f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}_X|} e^{-(1/2)(\mathbf{x}-\boldsymbol{\mu}_X)^T \mathbf{C}_X^{-1} (\mathbf{x}-\boldsymbol{\mu}_X)} \quad (20.6.2)$$

where the mean vector  $\boldsymbol{\mu}_X$  and the covariance matrix  $\mathbf{C}_X$  are given by

$$\boldsymbol{\mu}_X = \begin{bmatrix} \mu_X(t_1) \\ \mu_X(t_2) \\ \vdots \\ \mu_X(t_n) \end{bmatrix}; \quad \mathbf{C}_X = \begin{bmatrix} C_X(t_1) & C_X(t_1, t_2) & \cdots & C_X(t_1, t_n) \\ C_X(t_2, t_1) & C_X(t_2) & \cdots & C_X(t_2, t_n) \\ \vdots & \vdots & & \vdots \\ C_X(t_n, t_1) & C_X(t_n, t_2) & \cdots & C_X(t_n) \end{bmatrix} \quad (20.6.3)$$

### Properties

1. A Gaussian process is completely determined if the mean vector  $\boldsymbol{\mu}_X$  and the covariance matrix  $\mathbf{C}_X$  are known.
2. Sum of Gaussian processes are also Gaussian.
3. If a Gaussian process is weakly stationary, it is also strictly stationary.

### White Gaussian Noise

A white-noise process has been defined in Eq. (19.2.17) as a stationary process whose autocorrelation is a Dirac delta function. The white-noise process that is widely used for

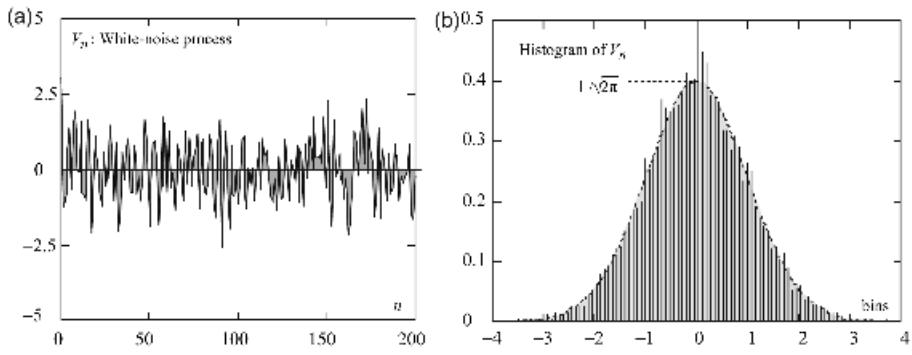


FIGURE 20.6.1

modeling communication problems is the Gaussian process. A standard Gaussian white-noise process  $V(t)$  has zero mean and unit variance with probability density

$$f_V(v) = \frac{1}{\sqrt{2\pi}} e^{-(v^2/2)}, \quad -\infty < v < \infty \quad (20.6.4)$$

The autocorrelation function and power spectral densities are given by

$$R_V(\tau) = \delta(\tau); \quad S_V(\omega) = 1, \quad -\infty < \omega < \infty \quad (20.6.5)$$

Similarly, a standard white Gaussian noise sequence  $\{V_n\}$  has zero mean and unit variance with the sequences coming from a density function given by Eq. (20.6.4). However, its autocorrelation function and power spectral densities are given by

$$R_V(h) = \delta_h; \quad S_V(\omega) = 1, \quad -\infty < \omega < \infty \quad (20.6.6)$$

In this study, 10,000 points of a Gaussian white noise  $\{V_n\}$  was generated from a computer using the Box–Mueller method. The noise sequence was standardized to zero mean and unit variance. This sequence is shown in Fig. 20.6.1a for only 200 points for clarity, and its histogram illustrating the Gaussian nature of the density is shown in Fig. 20.6.1b.

**Example 20.6.1 [AR(1) Process]** In Example 19.2.11 a discrete zero mean stationary random process  $X_i$  was given by

$$X_n = \phi X_{n-1} + V_n, \quad n = 1, \dots, X_0 = V_0$$

where  $\{V_n, n = 0, 1, \dots\}$  is a Gaussian process. This process is called an *autoregressive (1)* process. We will investigate this process assuming that  $V_n$  is standard Gaussian white noise and the coefficient  $\phi = 0.8$ .

We will show that  $X_n$  is also Gaussian-distributed. By defining  $B^k$  as a backward difference operator defined by  $B^k X_n = X_{n-k}$ ,  $k \geq 0$ , we can write the given equation as

$$(1 - \phi B)X_n = V_n$$

or

$$\begin{aligned} X_n &= (1 - \phi B)^{-1} V_n = (1 + \phi B + \phi^2 B^2 + \dots + \phi^{k-1} B^{k-1}) V_n \\ &= V_n + \phi V_{n-1} + \phi^2 V_{n-2} + \dots + \phi^{k-1} V_{n-k+1} + \dots \end{aligned}$$

Since  $X_n$  is a sum of a weighted Gaussian-distributed random sequence, it is also a Gaussian-distributed continuous-state discrete-time (CSDT) random process. The

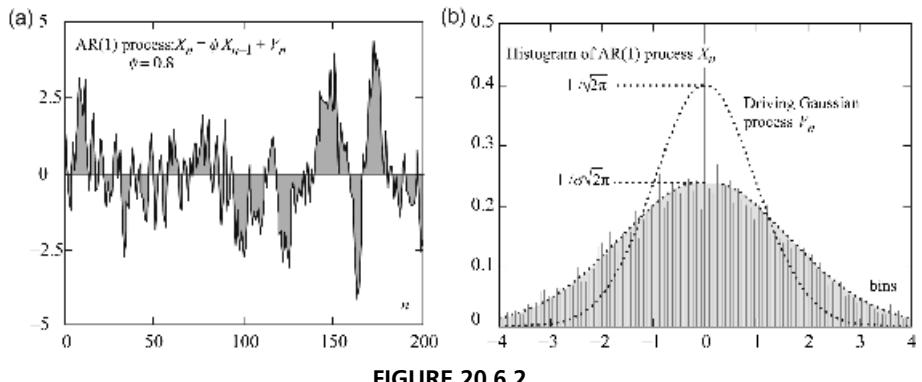


FIGURE 20.6.2

variance of this process as given by Eq. (19.2.47) is  $\sigma_X^2 = \sigma_X^2/(1 - \phi^2) = 1/(1 - 0.64) = 2.778$  with mean  $\mu_X = 0$ .

This process  $X_n$  simulated on a computer with 10,000 points of the Gaussian noise sequence generated previously is shown in Fig. 20.6.2a for only 200 points for clarity. The variance of the simulated process is 2.905, which is very close to the true value of 2.778. The histogram of the data of 10,000 points is shown in Fig. 20.6.2b. On the same histogram plot is shown the true pdf of  $X_n$  and the pdf of true white-noise sequence  $V_n$ .

## 20.7 WIENER PROCESS (BROWNIAN MOTION)

The random motion of a particle in a fluid subject to collisions and influence of other particles is called *Brownian motion*. One of the mathematical models of this motion is the *Wiener process*.

If the following conditions are satisfied,  $W(t)$  is a *Wiener process* with parameter  $\sigma^2$ :

1.  $W(0) = 0$ .
2.  $W(t)$  is a Gaussian process with sample continuous functions.
3.  $E[W(t)] = 0$ .
4. The autocovariance function  $C_W(t, s) = R_W(t, s) = E[W(t), W(s)] = \sigma^2 \min(t, s)$ .
5. The variance of  $W(t)$  from (4) is  $\sigma_W^2(t) = \sigma^2 t$ .

The pdf of the Wiener process is given by

$$f_W(w, t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-(1/2)(w^2/\sigma^2 t)} \quad (20.7.1)$$

The covariance matrix of the random variable vector  $\mathbf{W}^T = [W(t_1), W(t_2), \dots, W(t_n)]^T$  for times  $0 < t_1 < t_2 < \dots < t_n$  can be given from condition 4 as

$$\mathbf{C}_W = E[\mathbf{W}\mathbf{W}^T] = \sigma^2 \begin{bmatrix} t_1 & t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & t_2 & \cdots & t_2 \\ t_1 & t_2 & t_3 & \cdots & t_3 \\ \vdots & \vdots & \vdots & & \vdots \\ t_1 & t_2 & t_3 & \cdots & t_n \end{bmatrix} \quad (20.7.2)$$

and this matrix is positive definite. Since  $W(t)$  is Gaussian, the finite-dimensional density function using Eq. (20.7.2) is given by

$$f_W(w_1, t_1; w_2, t_2; \dots; w_n, t_n) = \frac{1}{(2\pi)^{n/2} |\mathbf{C}_W|^{1/2}} e^{-(1/2)\mathbf{W}^T \mathbf{C}_W^{-1} \mathbf{W}} \quad (20.7.3)$$

The determinant  $|\mathbf{C}_W| = t_1(t_2 - t_1) \cdots (t_{n-1} - t_n)$  and Eq. (20.7.3) can be rewritten as

$$f_W(w_1, t_1; w_2, t_2; \dots; w_n, t_n) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi(t_k - t_{k-1})\sigma^2}} \exp\left\{-\frac{1}{2}\left[\frac{(w_k - w_{k-1})^2}{\sigma^2(t_k - t_{k-1})}\right]\right\} \quad (20.7.4)$$

with  $t_0 = 0 = w_0$ . The density function of Eq. (20.7.4) shows that the sequence

$$\{W_{t_1}, (W_{t_2} - W_{t_1}), \dots, (W_{t_n} - W_{t_{n-1}})\}$$

is a collection of independent random variables for increasing  $\{0 < t_1 < t_2 < \dots < t_n\}$  and the density depends only on the time difference, showing that the Wiener process is a process of stationary independent increments. Hence, it is nonstationary and is a continuous-state continuous-time (CSCT) process.

### Derivation of Wiener Process

We will derive the Wiener process as the limiting form of the elementary random-walk process discussed in Section 20.4, which will be restated in a form where limits can be taken. A particle starts from origin at  $t = 0$ . In a small interval  $\Delta t$  it undergoes a small step  $\pm \Delta w$  with equal probabilities, thus constituting a symmetric random walk. The number  $n$  of Bernoulli trials in the time interval  $(0, t)$  is equal to the greatest integer less than or equal to  $t/\Delta t$ , or

$$n = \left\lfloor \frac{t}{\Delta t} \right\rfloor$$

where  $\lfloor x \rfloor$  represents the greatest integer less than or equal to  $x$ , called the “floor function.”

Similar to the definition in Eq. (20.4.11), the Bernoulli random variable  $Z_i$  is defined by

$$Z_i = \begin{cases} +1 & \text{if the } i\text{th step of } \Delta w \text{ is upward} \\ -1 & \text{if the } i\text{th step of } \Delta w \text{ is downward} \end{cases}$$

with probabilities

$$\begin{aligned} P(Z_i = +1) &= \frac{1}{2} \\ P(Z_i = -1) &= \frac{1}{2} \end{aligned} \quad (20.7.5)$$

The mean and variance of  $Z_i$  are given by  $E[Z_i] = 0$  and  $\text{var}[Z_i] = 1$ .

If the position of the particle at time  $n\Delta t$  is  $W(n\Delta t)$ , then

$$W(n\Delta t) = \Delta w \sum_{i=1}^n Z_i u(t - i\Delta t) \quad (20.7.6)$$

The mean and variance of the process  $W(n\Delta t)$  are

$$E[W(n\Delta t)] = 0 \quad \text{and} \quad \text{var}[W(n\Delta t)] = E[W^2(n\Delta t)] = \Delta w^2 n = \Delta w^2 \left\lfloor \frac{t}{\Delta t} \right\rfloor \quad (20.7.7)$$

We now let  $\Delta w$  and  $\Delta t$  tend to zero in such a way that Eq. (20.7.7) is nontrivial. Hence we have to set

$$\Delta w = \sigma \sqrt{\Delta t} \quad (20.7.8)$$

where  $\sigma$  is a positive constant so that as  $\Delta t \rightarrow \infty$ ,

$$W(t) = \lim_{\Delta t \rightarrow 0} \sigma \sum_{i=1}^{\infty} Z_i u(t - i\Delta t) \sqrt{\Delta t}$$

and

$$\lim_{\Delta t \rightarrow 0} \Delta w^2 \left\lfloor \frac{t}{\Delta t} \right\rfloor \rightarrow \sigma^2 t \quad (20.7.9)$$

$W(t)$  is the desired Wiener process, and a sample function is shown in Fig. 20.7.1.

The statistics of the Wiener process are as follows:

*Mean.* Since  $E[Z_i] = 0$ , we have

$$E[W(t)] = \mu_W(t) = 0 \quad (20.7.10)$$

*Variance.* From Eq. (20.7.7), we obtain

$$\text{var}[W(t)] = E[W^2(t)] = \sigma_W^2(t) = \sigma^2 t \quad (20.7.11)$$

*Autocorrelation.* Since  $E[W(t)] = 0$ ,  $R_W(t_1, t_2) = C_W(t_1, t_2)$ .

For  $t_1 < t_2$ :

$$\begin{aligned} R_W(t_1, t_2) &= C_W(t_1, t_2) = E[W(t_1)W(t_2)] \\ &= E\{[W(t_2) - W(t_1) + W(t_1)]W(t_1)\}. \end{aligned}$$

From linearity of expectations:  $R_W(t_1, t_2) = E\{[W(t_2) - W(t_1)]W(t_1)\} + E[W^2(t_1)]$ .

From independent increments:  $R_W(t_1, t_2) = E[W^2(t_1)] = \sigma^2 t_1$ .

Similarly, for  $t_2 < t_1$ , we have  $R_W(t_1, t_2) = E[W^2(t_2)] = \sigma^2 t_2$ .

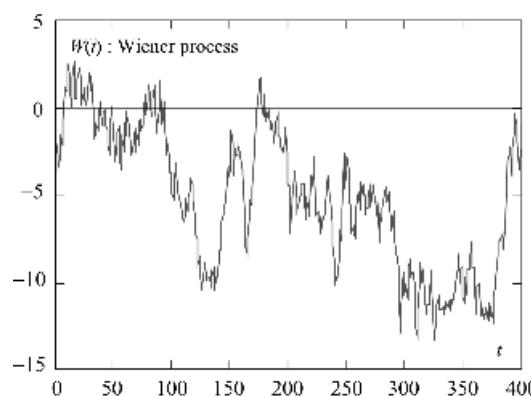


FIGURE 20.7.1

Hence

$$R_W(t_1, t_2) = C_W(t_1, t_2) = \sigma^2 \min(t_1, t_2) \quad (20.7.12)$$

The Wiener process has the following properties:

1. From the properties of the random walk from which  $W(t)$  was constructed, we conclude that for  $(0 < t < s)$  the increments  $W(t) - W(0)$  and  $W(s) - W(t)$  are independent.
2. From Eq. (20.7.1), for  $t_2 < t_1$  the density function of the increment  $W(t_2) - W(t_1)$  is

$$f_W(w_1, t_1; w_2, t_2) = \frac{1}{\sqrt{2\pi(t_2 - t_1)\sigma^2}} e^{-(1/2)[(w_2 - w_1)^2/\sigma^2(t_2 - t_1)]^2} \quad (20.7.13)$$

and since it depends only on the time difference  $(t_2 - t_1)$ ,  $W(t)$  is a stationary independent increment process.

3. Since  $W(t)$  is the limit of a symmetric random walk, it can be conjectured that it is a continuous process for all  $t > 0$ . Indeed it can be shown [35,37] that it is continuous everywhere with probability 1.
4. It can be shown [35,37] that  $W(t)$  is of unbounded variation, and as a result it is differentiable nowhere in the conventional sense.

Property 4 has far-reaching consequences. If we have any function  $g[W(t)]$  of another nonrandom function  $W(t)$ , then  $g[W(t + dt)]$  can be expanded in a Taylor series as

$$g[W(t + dt)] = g[W(t)] + \frac{dg[W(t)]}{dw} dW(t) + \frac{1}{2} \frac{d^2g[W(t)]}{dw^2} dW^2(t) + \dots \quad (20.7.14)$$

Neglecting higher-order terms Eq. (20.7.14), we can write  $dg[W(t)]$  as

$$g[W(t + dt)] - g[W(t)] = dg[W(t)] = \frac{dg[W(t)]}{dw} dW(t) \quad (20.7.15)$$

However, if  $W(t)$  is a Wiener process, the term  $dW^2(t)$  cannot be neglected because from Eq. (20.7.8) it is of the order of  $dt$ . Hence the differential  $dg[W(t)]$  can be written only as

$$\begin{aligned} g[W(t + dt)] - g[W(t)] &\approx \frac{dg[W(t)]}{dw} dW(t) + \frac{1}{2} \frac{d^2g[W(t)]}{dw^2} dW^2(t) \\ &\approx \frac{dg[W(t)]}{dw} dW(t) + \frac{1}{2} \sigma^2 \frac{d^2g[W(t)]}{dw^2} dt \end{aligned} \quad (20.7.16)$$

neglecting only third and higher-order terms. Thus the differential  $dg[W(t)]$  does not obey the ordinary laws of calculus! It is also interesting to note that if  $W(t)$  is a deterministic process, then  $\sigma^2$  is equal to zero and Eq. (20.7.16) reverts to the usual calculus. For more details on this interesting aspect, see Refs. 35 and 37.

### **Formal Derivative of Wiener Process**

We have stated in properties 3 and 4 that the Wiener process is continuous everywhere but differentiable nowhere in the conventional sense. However, we will formally

differentiate it and obtain its autocorrelation function. If  $Z(t) = W(t)$  is the formal derivative of the Wiener process, it can be shown from Eq. (19.2.31b) that the autocorrelation function of  $Z(t) = W(t)$  is given by

$$R_Z(t, s) = R_W(t, s) = E[W(t)W(s)] = \frac{\partial^2 R_W(t, s)}{\partial t \partial s} \quad (20.7.17)$$

and if  $s = t + \tau$ , we obtain Eq. (19.2.31b). We substitute  $R_W(t, s) = \sigma^2 \min(t, s)$  from Eq. (20.7.12) in Eq. (20.7.17) and obtain

$$R_W(t, s) = R_Z(t, s) = E[Z(t)Z(s)] = \frac{\partial^2}{\partial t \partial s} \sigma^2 \min(t, s) = \sigma^2 \delta(t - s) \quad (20.7.18)$$

From Eq. (20.7.18) we conclude that since the autocorrelation of the derivative of the Wiener process yields a Dirac delta function, it exists only in the generalized sense. However, the autocorrelation of  $Z(t) = W(t)$  is a function of the time difference  $(t - s)$ . It is also a zero mean Gaussian process. Hence the formal derivative  $Z(t)$  of a Wiener process is a zero mean stationary Gaussian process with power spectral density  $\sigma^2$  that is constant in the entire frequency range. This is the Gaussian white noise that has been used extensively to model communication and time-series problems. The similarity between the Gaussian and Poisson white noises [Eq. (20.2.39)] is to be noted.

## 20.8 MARKOV PROCESS

It was stated in Section 19.1 that a random process  $X(t)$  is completely specified if the  $n$ th-order joint density given by Eq. (19.1.9) for an ordered set of time functions  $\{t_1 < t_2 < t_3 < \dots < t_n\}$

$$f_X(x_n, \dots, x_1; t_n, \dots, t_1) = \frac{\partial^n}{\partial x_n \dots \partial x_1} F_X(x_n, \dots, x_1; t_n, \dots, t_1) \quad (20.8.1)$$

is specified for all positive integers  $n$ . The joint density given by Eq. (20.8.1) can be decomposed into  $n$  first-order conditional densities for the same set of ordered time functions as shown below:

$$\begin{aligned} & f_X(x_1; t_1) \\ & f_X(x_2; t_2 | x_1; t_1) \\ & f_X(x_3; t_3 | x_2, x_1; t_2, t_1) \\ & f_X(x_{n-1}; t_{n-1} | x_{n-2}, \dots, x_2, x_1; t_{n-2}, \dots, t_2, t_1) \\ & f_X(x_n; t_n | x_{n-1}, \dots, x_2, x_1; t_{n-1}, \dots, t_2, t_1) \end{aligned} \quad (20.8.2)$$

By repeated application of the definition of conditional probability, the  $n$ -dimensional joint density of Eq. (20.8.1) can be obtained from Eq. (20.8.2). Thus, for the random process to be completely determined, we should have  $n \rightarrow \infty$  of these first-order conditional densities. However, the specifications for a class of random processes can be considerably simplified if we can express their conditional densities as

$$f_X(x_n; t_n | x_{n-1}, \dots, x_2, x_1; t_{n-1}, \dots, t_2, t_1) = f_X(x_n; t_n | x_{n-1}; t_{n-1}) \quad (20.8.3)$$

What Eq. (20.8.3) conveys is that the probabilistic behavior of the present given all past history is dependent only on the immediate past. The property exhibited by Eq. (20.8.3) is called the *Markov property*, and the class of processes having this property is known as

*Markov processes.* Using the Markov property the  $n$ -dimensional density can be given as follows:

$$f_X(x_n, \dots, x_1; t_n, \dots, t_1) = f(x_n; t_n | x_{n-1}; t_{n-1}) f_X(x_{n-1}, \dots, x_1; t_{n-1}, \dots, t_1) \quad (20.8.4)$$

By applying the Markov property repeatedly to lower-order density functions, we can write the  $n$ -dimensional density as

$$f_X(x_n, \dots, x_1; t_n, \dots, t_1) = f_X(x_1; t_1) \prod_{i=2}^n f(x_i; t_i | x_{i-1}; t_{i-1}) \quad (20.8.5)$$

Hence all finite-dimensional densities of a Markov process are completely determined by the initial marginal density  $f_X(x_1; t_1)$  and the set of first-order conditional densities,  $\{f(x_i; t_i | x_{i-1}; t_{i-1}), i = 2, 3, \dots\}$ . The conditional probability density  $f(x_i; t_i | x_{i-1}; t_{i-1})$  is called the *transition probability density*. For a discrete-time Markov process, the equation corresponding to Eq. (20.8.3) can be written as follows:

$$P\{X(t_n) \leq x_n | X(t_{n-1}) \leq x_{n-1}, \dots, X(t_1) \leq x_1\} = P\{X(t_n) \leq x_n | X(t_{n-1}) \leq x_{n-1}\} \quad (20.8.6)$$

What we have shown is that given a Markov process, its  $n$ -dimensional density can be expressed by the initial marginal density and a set of first-order conditional densities. However, to construct a Markov process from the given first-order conditional densities, the following consistency conditions have to be satisfied:

$$1. \quad f_X(x_2; t_2) = \int_{-\infty}^{\infty} f_X(x_2; t_2 | x_1; t_1) f_X(x_1; t_1) dx_2, t_1 < t_2 \quad (20.8.7)$$

This equation is the usual equation that relates a joint density to its marginal density.

$$2. \quad f_X(x_3; t_3 | x_1; t_1) = \int_{-\infty}^{\infty} f_X(x_3; t_3 | x_2; t_2) f_X(x_2; t_2 | x_1; t_1) dx_2, t_1 < t_2 < t_3 \quad (20.8.8)$$

This equation is called the *Chapman–Kolmogorov* equation, which all Markov processes must satisfy. We will show that Markov processes satisfy this equation. The joint pdf  $f_X(x_3, x_1; t_3, t_1)$  with  $t_1 < t_2 < t_3$  can be written as

$$f_X(x_3, x_1; t_3, t_1) = \int_{-\infty}^{\infty} f_X(x_3, x_2, x_1; t_3, t_2, t_1) dx_2 \quad (20.8.9)$$

Using the property of conditional expectation twice, we can rewrite Eq. (20.8.9) as,

$$\begin{aligned} f_X(x_3, x_1; t_3, t_1) &= \int_{-\infty}^{\infty} f_X(x_3; t_3 | x_2, x_1; t_2, t_1) f_X(x_2, x_1; t_2, t_1) dx_2 \\ &= \int_{-\infty}^{\infty} f_X(x_3; t_3 | x_2, x_1; t_2, t_1) f_X(x_2; t_2 | x_1; t_1) f_X(x_1; t_1) dx_2 \end{aligned} \quad (20.8.10)$$

Using the Markov property  $f_X(x_3; t_3 | x_2, x_1; t_2, t_1) = f_X(x_3; t_3 | x_2; t_2)$  in Eq. (20.8.10), we obtain

$$f_X(x_3, x_1; t_3, t_1) = f_X(x_1; t_1) \int_{-\infty}^{\infty} f_X(x_3; t_3 | x_2; t_2) f_X(x_2; t_2 | x_1; t_1) dx_2 \quad (20.8.11)$$

Dividing Eq. (20.8.11) throughout by  $f_X(x_1; t_1)$ , for  $t_1 < t_2 < t_3$ , we have

$$\begin{aligned} \frac{f_X(x_3, x_1; t_3, t_1)}{f_X(x_1; t_1)} &= f_X(x_3, t_3 | x_1; t_1) \\ &= \int_{-\infty}^{\infty} f_X(x_3; t_3 | x_2; t_2) f_X(x_2; t_2 | x_1; t_1) dx_2 \end{aligned} \quad (20.8.8)$$

which is the desired Chapman–Kolmogorov (C–K) equation.

This derivation also holds good for discrete-time processes, and the result is

$$\begin{aligned} P\{X(t_3) \leq x_3 | X(t_1) \leq x_1\} &= \sum_{x_2} P\{X(t_3) \leq x_3 | X(t_2) \leq x_2\} \\ &\quad \times P\{X(t_2) \leq x_2 | X(t_1) \leq x_1\} \end{aligned} \quad (20.8.12)$$

The state  $\{x_2, t_2\}$  represents the intermediate state, and operation of the C–K equation is shown in Fig. 20.8.1.

The C–K equation is a sufficient condition for a Markov process in the sense that non-Markovian processes may also satisfy this equation.

We can also define a Markov process in terms of conditional expectations. If  $g(\cdot)$  is any function with  $\max_x |g(x)| < \infty$ , then  $X(t)$  is a Markov process for an ordered sequence  $\{t_1 < t_2 < \dots < t_n\}$  if and only if

$$E\{g[X(t_n)] | X(t_{n-1}), X(t_{n-2}), \dots, X(t_1)\} = E\{g[X(t_n)] | X(t_{n-1})\} \quad (20.8.13)$$

If the conditional probability  $P\{X(t_n) \leq x_n | X(t_{n-1}) \leq x_{n-1}\}$  of a Markov process is dependent only on the time difference  $(t_n - t_{n-1})$ , it is called a *homogeneous* process. In other words, if  $(t_2 - t_1) = (t_n - t_{n-1})$ , then homogeneity implies

$$P\{X(t_n) \leq x_n | X(t_{n-1}) \leq x_{n-1}\} = P\{X(t_2) \leq x_2 | X(t_1) \leq x_1\}$$

Generally, a homogenous Markov process may not be stationary since  $P\{X(t_1) \leq x_1\}$  may depend on  $t_1$ .

We have shown earlier that a stationary independent increment process is always non-stationary in light of Eqs. (20.4.2a) and (20.4.2b). We will now show that this process is always a Markov process. Let  $X(t)$  be a stationary independent increment process with

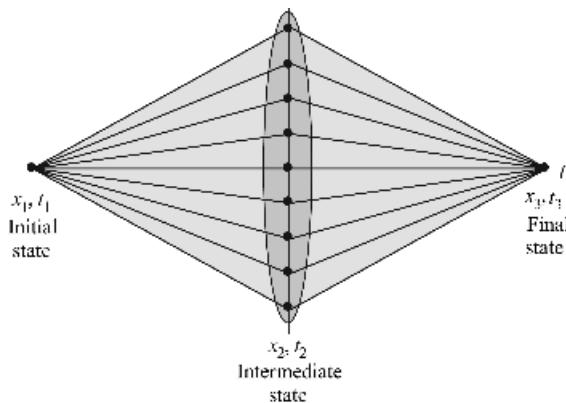


FIGURE 20.8.1

$t_0 = 0$  and  $X(t_0) = x_0 = 0$ . For all  $t_i < t_j$ , the events

$$\{X(t_i) = x_i, X(t_j) = x_j\} \quad \text{and} \quad \{X(t_i - t_0) = x_i - x_0, X(t_j - t_i) = x_j - x_i\}$$

are equal, and hence their joint densities are also equal. As a consequence

$$f_X(x_j; t_j, x_i; t_i) = f_X(x_j - x_i; t_j - t_i, x_i - x_0; t_i - t_0), \quad i < j \quad (20.8.14)$$

With  $x_0 = t_0 = 0$  and using the independent increment property, Eq. (20.8.14) can be rewritten after dropping the time difference  $(t_j - t_i)$ , due to stationarity:

$$f_X(x_j; t_j, x_i; t_i) = f_X(x_j - x_i) f_X(x_i; t_i), \quad i < j \quad (20.8.15)$$

Analogous to Eq. (20.8.15), for the sequence  $\{t_n < t_{n-1} < \dots < t_1 < t_0 = 0\}$ , we can write

$$\begin{aligned} f_X(x_n, \dots, x_1, x_0; t_n, \dots, t_1, t_0) &= f_X(x_n - x_{n-1}, \dots, x_2 - x_1, x_1 - x_0; \\ &\quad \times t_n - t_{n-1}, \dots, t_2 - t_1, t_1 - t_0) \\ &= f_X(x_n - x_{n-1}) \cdots f_X(x_2 - x_1) f_X(x_1; t_1) \end{aligned} \quad (20.8.16)$$

From Eq. (20.8.15) we can substitute

$$f_X(x_k - x_{k-1}) = \frac{f_X(x_k, x_{k-1}; t_k, t_{k-1})}{f_X(x_{k-1}; t_{k-1})}, \quad k = n, n-1, \dots, 2$$

in Eq. (20.8.16) and write

$$f_X(x_n, \dots, x_1; t_n, \dots, t_1) = \frac{f_X(x_n, x_{n-1}; t_n, t_{n-1}) \cdots f_X(x_2, x_1; t_2, t_1)}{f_X(x_{n-1}; t_{n-1}) \cdots f_X(x_2; t_2)} \quad (20.8.17)$$

Equation (20.8.17) expresses the  $n$ th-order joint density of a stationary independent increment process  $X(t)$  in terms of the second- and first-order densities. From the definition of conditional density, we can write

$$f_X(x_n; t_n | x_{n-1}, \dots, x_1; t_{n-1}, \dots, t_1) = \frac{f_X(x_n, x_{n-1}, \dots, x_1; t_n, t_{n-1}, \dots, t_1)}{f_X(x_{n-1}, \dots, x_1; t_{n-1}, \dots, t_1)} \quad (20.8.18)$$

Applying Eq. (20.8.17) to both numerator and denominator of Eq. (20.8.18), we obtain

$$\begin{aligned} &f_X(x_n; t_n | x_{n-1}, \dots, x_1; t_{n-1}, \dots, t_1) \\ &= \frac{f_X(x_n, x_{n-1}; t_n, t_{n-1}) \cdots f_X(x_2, x_1; t_2, t_1)}{f_X(x_{n-1}; t_{n-1}) \cdots f_X(x_2; t_2)} \\ &= \frac{f_X(x_n, x_{n-1}; t_n, t_{n-1})}{f_X(x_{n-1}; t_{n-1})} = f_X(x_n; t_n | x_{n-1}; t_{n-1}) \end{aligned} \quad (20.8.19)$$

which is the defining equation of a Markov process. However, the converse is not true, or, not every Markov process is a stationary independent increment process.

### Summary of Properties of Markov Processes

1.  $f_X(x_n; t_n | x_{n-1}; t_{n-1}, \dots, x_1; t_1) = f_X(x_n; t_n | x_{n-1}; t_{n-1})$  (20.8.20)

Applying the chain rule for conditional expectations from Eq. (20.8.5), we can write

$$\begin{aligned} f_X(x_n; t_n | x_{n-1}; t_{n-1}, \dots, x_1; t_1) &= \frac{f_X(x_n; t_n, x_{n-1}; t_{n-1}, \dots, x_1; t_1)}{f_X(x_{n-1}; t_{n-1}, \dots, x_1; t_1)} \\ &= f_X(x_n; t_n | x_{n-1}; t_{n-1}) \end{aligned} \quad (20.8.21)$$

2. This rule can be generalized to any sequence of times  $\{t_{n_k} > t_{n_{k-1}} > \dots > t_{n_1}\}$ :

$$f_X(x_{n_k}; t_{n_k} | x_{n_{k-1}}; t_{n_{k-1}}, \dots, x_{n_1}; t_{n_1}) = f_X(x_{n_k}; t_{n_k} | x_{n_{k-1}}; t_{n_{k-1}})$$

3. A time-reversed Markov process is also a Markov process:

$$f_X(x_n; t_n | x_{n+1}; t_{n+1}, \dots, x_{n+k}; t_{n+k}) = f_X(x_n; t_n | x_{n+1}; t_{n+1}) \quad (20.8.22)$$

4. Given the present information, the past and the future are conditionally independent. Or, for  $k < m < n$ , we obtain

$$f_X(x_n; t_n, x_k; t_k | x_m; t_m) = f_X(x_n; t_n | x_m; t_m) f(x_k; t_k | x_m; t_m) \quad (20.8.23)$$

We will show that this result is true:

$$\begin{aligned} f_X(x_n; t_n, x_k; t_k | x_m; t_m) &= \frac{f_X(x_n; t_n, x_m; t_m, x_k; t_k)}{f_X(x_m; t_m)} \\ &= \frac{f_X(x_n; t_n | x_m; t_m) f_X(x_m; t_m | x_k; t_k) f_X(x_k; t_k)}{f_X(x_m; t_m)} \quad (\text{chain rule}) \end{aligned}$$

Substituting for

$$f_X(x_m; t_m | x_k; t_k) = \frac{f_X(x_k; t_k | x_m; t_m) f_X(x_m; t_m)}{f_X(x_k; t_k)}$$

in the preceding equation, we obtain Eq. (20.8.23).

5. Taking conditional expectations in property 3, we have

$$\begin{aligned} E[X(t_n)X(t_k) | X(t_m) = x_m] &= E[X(t_n) | X(t_m) = x_m] \\ &= E[X(t_k) | X(t_m) = x_m] \end{aligned} \quad (20.8.24)$$

6. The future, conditioned on the past history upto the present, is the future conditioned on the present:

$$E\{X(t_{n+1}) | X(t_n), X(t_{n-1}), \dots, X(t_1)\} = E\{X(t_{n+1}) | X(t_n)\} \quad (20.8.25)$$

This equation is a restatement of Eq. (20.8.20).

7. If  $t_1 < t_2 < t_3$ , then

$$E[X(t_3) | X(t_1)] = E\{E[X(t_3) | X(t_2)] | X(t_1)\} \quad (20.8.26)$$

This equation is a restatement of the C-K property in terms of conditional expectations.

**Example 20.8.1 (Poisson Process)** The Poisson process  $N(t)$  is an independent increment process from Eq. (20.2.10.). It is also a stationary independent increment process, because the probability of the increment,  $N(t_2) - N(t_1)$ , given by

$$P\{N(t_2) - N(t_1) = k_2 - k_1\} = \frac{[\lambda(t_2 - t_1)]^{k_2 - k_1}}{(k_2 - k_1)!} e^{-\lambda(t_2 - t_1)}$$

is dependent only on time difference  $(t_2 - t_1)$ . From the result of Eq. (20.8.19), we can conclude that the Poisson process is a homogeneous Markov process. However, it will be instructive to show the Markov property from the definition of Eq. (20.8.20). The second-order probability function is given by

$$P\{N(t_2) = k_2, N(t_1) = k_1\} = P\{N(t_2) = k_2 | N(t_1) = k_1\} P\{N(t_1) = k_1\}$$

The conditional probability  $N(t_2) = k_2$  given that  $N(t_1) = k_1$  is the probability that there are  $k_2 - k_1$  events occurring in the time interval  $t_2 - t_1$ :

$$\begin{aligned} P\{N(t_2) = k_2 | N(t_1) = k_1\} &= P\{N(t_2) - N(t_1) = k_2 - k_1\} \\ &= \frac{[\lambda(t_2 - t_1)]^{k_2 - k_1}}{(k_2 - k_1)!} e^{-\lambda(t_2 - t_1)}, \quad k_2 \geq k_1 \end{aligned}$$

Hence the second-order probability function  $P\{N(t_2) = k_2, N(t_1) = k_1\}$  is given by

$$P\{N(t_2) = k_2, N(t_1) = k_1\} = \frac{[\lambda t_1]^{k_1} [\lambda(t_2 - t_1)]^{k_2 - k_1}}{k_1!(k_2 - k_1)!} e^{-\lambda t_2}, \quad k_2 \geq k_1$$

Proceeding in a similar manner, we can obtain the third-order probability function as

$$\begin{aligned} P\{N(t_3) = k_3, N(t_2) = k_2, N(t_1) = k_1\} &= \frac{[\lambda t_1]^{k_1} [\lambda(t_2 - t_1)]^{k_2 - k_1} [\lambda(t_3 - t_2)]^{k_3 - k_2}}{k_1!(k_2 - k_1)!(k_3 - k_2)!} e^{-\lambda t_3}, \quad k_3 \geq k_2 \geq k_1 \end{aligned}$$

The second-order probability function  $P\{N(t_3) = k_3, N(t_2) = k_2\}$  can be evaluated as

$$\begin{aligned} P\{N(t_3) = k_3, N(t_2) = k_2\} &= P\{N(t_3) - N(t_2) = k_3 - k_2\} P\{N(t_2) = k_2\} \\ &= \frac{[\lambda t_2]^{k_2} [\lambda(t_3 - t_2)]^{k_3 - k_2}}{k_2!(k_3 - k_2)!} e^{-\lambda t_3}, \quad k_3 \geq k_2 \end{aligned}$$

Combining all these equations, we have

$$\begin{aligned}
 P\{N(t_3) = k_3 \mid N(t_2) = k_2\} &= \frac{P\{N(t_3) = k_3, N(t_2) = k_2\}}{P\{N(t_2) = k_2\}} \\
 &= \frac{\frac{[\lambda t_2]^{k_2} [\lambda(t_3 - t_2)]^{k_3 - k_2}}{k_2!(k_3 - k_2)!} e^{-\lambda t_3}}{\frac{[\lambda t_2]^{k_2}}{k_2!} e^{-\lambda t_2}} = \frac{[\lambda(t_3 - t_2)]^{k_3 - k_2}}{(k_3 - k_2)!} e^{-\lambda(t_3 - t_2)} \\
 P\{N(t_3) = k_3 \mid N(t_2) = k_2, N(t_1) = k_1\} &= \frac{P\{N(t_3) = k_3, N(t_2) = k_2, N(t_1) = k_1\}}{P\{N(t_2) = k_2, N(t_1) = k_1\}} \\
 &= \frac{\frac{[\lambda t_1]^{k_1} [\lambda(t_2 - t_1)]^{k_2 - k_1} [\lambda(t_3 - t_2)]^{k_3 - k_2}}{k_1!(k_2 - k_1)!(k_3 - k_2)!} e^{-\lambda t_3}}{\frac{[\lambda t_1]^{k_1} [\lambda(t_2 - t_1)]^{k_2 - k_1}}{k_1!(k_2 - k_1)!} e^{-\lambda t_2}} \\
 &= \frac{[\lambda(t_3 - t_2)]^{k_3 - k_2}}{(k_3 - k_2)!} e^{-\lambda(t_3 - t_2)}
 \end{aligned}$$

Hence  $P\{N(t_3) = k_3 \mid N(t_2) = k_2, N(t_1) = k_1\} = P\{N(t_3) = k_3 \mid N(t_2) = k_2\}$ , showing that  $N(t)$  is indeed a Markov process. Since the conditional density

$$P\{N(t_2) = k_2 \mid N(t_1) = k_1\} = \frac{[\lambda(t_2 - t_1)]^{k_2 - k_1}}{(k_2 - k_1)!} e^{-\lambda(t_2 - t_1)}, \quad k_2 \geq k_1$$

is a function of the time difference  $(t_2 - t_1)$ , we conclude further that the Poisson process is also homogeneous Markov process.

**Example 20.8.2 (Wiener Process)** The  $n$ -dimensional density function of a Wiener process as given by Eq. (20.7.4) is

$$f_W(w_1, w_2, \dots, w_n; t_1, t_2, \dots, t_n) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi(t_k - t_{k-1})\sigma^2}} \exp \left\{ -\frac{1}{2} \left[ \frac{(w_k - w_{k-1})^2}{\sigma^2(t_k - t_{k-1})} \right] \right\}$$

with  $t_0 = 0 = w_0$  and the sequence

$$\{W_{t_1}, (W_{t_2} - W_{t_1}), \dots, (W_{t_n} - W_{t_{n-1}})\}$$

is a collection of independent random variables for increasing  $\{0 < t_1 < t_2 < \dots < t_n\}$ . We have to show that

$$f_W(w_n | w_{n-1}, \dots, w_1) = f_W(w_n | w_{n-1})$$

or

$$\frac{f_W(w_n, w_{n-1}, \dots, w_1)}{f_W(w_{n-1}, \dots, w_1)} = \frac{f_W(w_n, w_{n-1})}{f_W(w_{n-1})}$$

$$\begin{aligned} \frac{f_W(w_n, w_{n-1}, \dots, w_1)}{f_W(w_{n-1}, \dots, w_1)} &= \frac{\prod_{k=1}^n \frac{1}{\sqrt{2\pi(t_k - t_{k-1})\sigma^2}} \exp\left\{-\frac{1}{2}\left[\frac{(w_k - w_{k-1})^2}{\sigma^2(t_k - t_{k-1})}\right]\right\}}{\prod_{k=1}^{n-1} \frac{1}{\sqrt{2\pi(t_k - t_{k-1})\sigma^2}} \exp\left\{-\frac{1}{2}\left[\frac{(w_k - w_{k-1})^2}{\sigma^2(t_k - t_{k-1})}\right]\right\}} \\ &= \frac{1}{\sqrt{2\pi(t_n - t_{n-1})\sigma^2}} \exp\left\{-\frac{1}{2}\left[\frac{(w_n - w_{n-1})^2}{\sigma^2(t_n - t_{n-1})}\right]\right\} \\ &\quad \frac{1}{2\pi(t_n - t_{n-1})(t_{n-1} - t_{n-2})\sigma^4} \exp\left\{-\frac{1}{2}\left[\frac{(w_n - w_{n-1})^2}{\sigma^2(t_n - t_{n-1})}\right]\right\} \\ \frac{f_W(w_n, w_{n-1})}{f_W(w_{n-1})} &= \frac{\times \exp\left\{-\frac{1}{2}\left[\frac{(w_{n-1} - w_{n-2})^2}{\sigma^2(t_{n-1} - t_{n-2})}\right]\right\}}{\frac{1}{\sqrt{2\pi(t_{n-1} - t_{n-2})\sigma^2}} \exp\left\{-\frac{1}{2}\left[\frac{(w_{n-1} - w_{n-2})^2}{\sigma^2(t_{n-1} - t_{n-2})}\right]\right\}} \\ &= \frac{1}{\sqrt{2\pi(t_n - t_{n-1})\sigma^2}} \exp\left\{-\frac{1}{2}\left[\frac{(w_n - w_{n-1})^2}{\sigma^2(t_n - t_{n-1})}\right]\right\} \end{aligned}$$

which shows that  $W(t)$  is a homogeneous Markov process with stationary independent increments.

**Example 20.8.3 (AR Process)** In Example 20.6.1 an autoregressive(1) process was given by

$$X_n = \phi X_{n-1} + V_n, \quad n = 1, \dots, X_0 = V_0$$

where  $\{V_n, n = 0, 1, \dots\}$  is a Gaussian process with zero mean and variance  $\sigma_V^2$ . The conditional density  $f_X(x_n | X_{n-1} = x_{n-1})$  is also Gaussian-distributed with mean  $\phi x_{n-1}$  and variance  $\sigma_V^2$ . This is a Markov process since the conditional density function depends only on the  $X_{n-1}$  and not any of the previous values. Thus

$$f_X(x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = f_X(x_n | X_{n-1} = x_{n-1})$$

**Example 20.8.4 (Polya Urn Model)** Initially at  $n = 0$ , an urn contains  $b$  blue balls and  $r$  red balls. In each trial, a ball is drawn at random and its color recorded. It is then replaced with  $c$  balls of the same color. Let  $X_n$  be the random variable representing the color of the ball at the  $n$ th draw. We will also assume that  $X_k = 1$  if the ball selected at the  $k$ th trial is

blue and  $X_k = 0$  if it is red. The probability of  $X_1 = 1$  is

$$P\{X_1 = 1\} = \frac{b}{b+r}$$

The probability of  $X_2 = 1$  conditioned on  $X_1 = 1$  is

$$P\{X_2 = 1 | X_1 = 1\} = \frac{b+c}{b+c+r}$$

Hence, the joint probability  $P\{X_2 = 1, X_1 = 1\}$

$$P\{X_2 = 1, X_1 = 1\} = \frac{b}{b+r} \cdot \frac{b+c}{b+c+r}$$

The probability of  $X_3 = 1$  conditioned on  $X_2 = 1$  and  $X_1 = 1$  is

$$P\{X_3 = 1 | X_2 = 1, X_1 = 1\} = \frac{b+2c}{b+2c+r}$$

and the joint probability  $P\{X_3 = 1, X_2 = 1, X_1 = 1\}$  is

$$P\{X_3 = 1, X_2 = 1, X_1 = 1\} = \frac{b+2c}{b+2c+r} \cdot \frac{b+c}{b+c+r} \cdot \frac{b}{b+r}$$

The realization  $\{X_6 = 0, X_5 = 1, X_4 = 0, X_3 = 0, X_2 = 1, X_1 = 1\}$  can be interpreted as the event of drawing 3 blue balls in 6 trials, and the probability of this event is  $p(3, 6)$ . Thus

$$\begin{aligned} P\{X_6 = 0, X_5 = 1, X_4 = 0, X_3 = 0, X_2 = 1, X_1 = 1\} &= p(3, 6) \\ &= \frac{r+2c}{b+5c+r} \cdot \frac{b+2c}{b+4c+r} \cdot \frac{r+c}{b+3c+r} \cdot \frac{r}{b+2c+r} \cdot \frac{b+c}{b+c+r} \cdot \frac{b}{b+r} \end{aligned}$$

Extending these results, the probability of a sequence of  $k$  blue balls and  $(n - k)$  red balls defined by  $p(k, n)$  can be determined as follows:

$$\begin{aligned} p(k, n) &= P\{X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_1 = x_1\} \\ &= \frac{b(b+c)(b+2c) \cdots [b + (k-1)c]r(r+c)(r+2c) \cdots [r + (n-k-1)c]}{(b+r)(b+c+r)(b+2c+r) \cdots [b+r+(n-1)c]} \end{aligned}$$

This random sequence is an example of a discrete-state discrete-time (DSDT) random process. It is nonstationary because the probability is dependent on time,  $n$ .

From the discussion above, we have

$$P\{X_3 = 1 | X_2 = 1\} = \frac{b+c}{b+c+r}$$

but

$$P\{X_3 = 1 | X_2 = 1, X_1 = 1\} = \frac{b+2c}{b+2c+r}$$

Since  $P\{X_3 = 1 | X_2 = 1, X_1 = 1\} \neq P\{X_3 = 1 | X_2 = 1\}$ , the sequence  $\{X_n\}$  is not a Markov process.

## 20.9 MARKOV CHAINS

When the state of a Markov process is discrete, we call it a *Markov chain* (MC). If the time is also discrete, it is called a *discrete-time Markov chain*. The binomial process (Example 20.4.2) and the random-walk process (Example 20.4.3) are illustrations of a discrete Markov chain. If the time is continuous, it is called a *continuous-time Markov chain*. The Poisson process discussed in detail in Section 20.2 is an example of a continuous-time Markov chain. These processes have been extensively studied, and many physical phenomena are modeled by them. We will now discuss discrete Markov chains. For continuous Markov chains, see Refs. 13, 30, and 54.

### Discrete Markov Chain

A random process  $\{X_n, n = 0, 1, \dots\}$  with discrete state space  $S$  is a *Markov chain* if the following probability relation holds for each  $\{j, i_n, \dots, i_0\} \in S$ :

$$P\{X_{n+m} = j | X_n = i_n, \dots, X_0 = i_0\} = P\{X_{n+m} = j | X_n = i_n\} = p_{i_n j}^{(m)}(n) \quad (20.9.1)$$

The quantity  $P\{X_{n+m} = j | X_n = i_n\} = p_{i_n j}^{(m)}(n)$  is the conditional probability of transition from state  $i_n$  at time  $n$  to state  $j$  at time  $n+m$ . It is referred to as the *m-step transition probability*, which in general will be a function of time  $n$ . However, if the transition from state  $i$  to state  $j$  is dependent only on the time difference  $m = n+m-n$  and not on the current time  $n$ , then the transition probability is *stationary* and the Markov chain is *homogeneous*, or

$$P\{X_{n+m} = j | X_n = i\} = P\{X_m = j | X_0 = i\} = p_{ij}^{(m)} \quad (20.9.2)$$

The one-step transition probability of a homogeneous Markov chain  $p_{ij}^{(1)} = p_{ij}$  can be defined as

$$p_{ij} = P\{X_{n+1} = j | X_n = i\} = P\{X_1 = j | X_0 = i\} \quad (20.9.3)$$

If the number of states  $N$  is finite, the Markov chain is finite; otherwise it is infinite. The one-step probabilities  $\{p_{ij}\}$  can be expressed as either a transition matrix  $\mathbf{P}$  or a state transition flow diagram.

To make these ideas clear, we will consider the modified Bernoulli process given in Eq. (20.4.11), where the random variable  $X_n$  at time  $n$  is in state 1 with probability  $p_0^{(n)}$  or in state  $-1$  with probability  $p_1^{(n)}$  as given below:

$$\begin{aligned} P(X_n = 1) &= p_0^{(n)} \\ P(X_n = -1) &= p_1^{(n)} \end{aligned} \quad (20.9.4)$$

Further, if the particle is in state 1 at time  $n$ , then the probability of transition to  $-1$  is  $\alpha$  at time  $(n+1)$ , and if it is in state  $-1$  at  $n$ , then the probability of transition to 1 is  $\beta$  at the time  $(n+1)$ . We are considering a two-state homogeneous Markov chain. The transition matrix  $\mathbf{P}$  is shown in Eq. (20.9.5):

$$\mathbf{P} = \begin{Bmatrix} 1 & -1 \\ -1 & 1 \end{Bmatrix} \quad (20.9.5)$$

The state transition flow diagram is shown in Fig. 20.9.1.

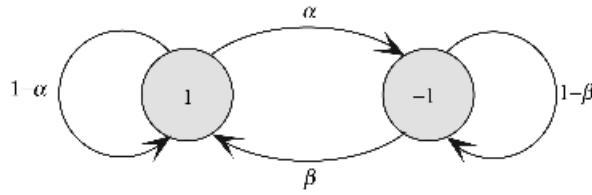


FIGURE 20.9.1

If there are  $N+1$  states and the transition probability from state  $i$  to state  $j$  is  $p_{ij}$ , then the one-step transition matrix  $\mathbf{P}$  is given by

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & \cdots & j & \cdots & N \\ 0 & p_{00} & p_{01} & \cdots & p_{0j} & \cdots & p_{0N} \\ 1 & p_{10} & p_{11} & \cdots & p_{1j} & \cdots & p_{1N} \\ \vdots & \vdots & & & \vdots & & \vdots \\ i & p_{i0} & p_{i1} & \cdots & p_{ij} & \cdots & p_{iN} \\ \vdots & \vdots & & & \vdots & & \vdots \\ N & p_{N0} & p_{N1} & \cdots & p_{Nj} & \cdots & p_{NN} \end{bmatrix} \quad (20.9.6)$$

The matrix  $\mathbf{P}$  is called a *stochastic matrix* and has the property that the row sum equals 1:

$$\sum_{j=0}^N p_{ij} = 1, \quad i = 0, 1, \dots, N \quad (20.9.7)$$

The  $m$ -step transition probability for all  $i, j$  defined in Eq. (20.9.2) can also be given in matrix form as follows:

$$\mathbf{P}^{(m)} = \begin{bmatrix} 0 & 1 & \cdots & j & \cdots & N \\ 0 & p_{00}^{(m)} & p_{01}^{(m)} & \cdots & p_{0j}^{(m)} & \cdots & p_{0N}^{(m)} \\ 1 & p_{10}^{(m)} & p_{11}^{(m)} & \cdots & p_{1j}^{(m)} & \cdots & p_{1N}^{(m)} \\ \vdots & \vdots & & & \vdots & & \vdots \\ i & p_{i0}^{(m)} & p_{i1}^{(m)} & \cdots & p_{ij}^{(m)} & \cdots & p_{iN}^{(m)} \\ \vdots & \vdots & & & \vdots & & \vdots \\ N & p_{N0}^{(m)} & p_{N1}^{(m)} & \cdots & p_{Nj}^{(m)} & \cdots & p_{NN}^{(m)} \end{bmatrix} \quad (20.9.8)$$

Let us consider the two-step transition from state  $i$  at time 0 through an intermediate state  $k$  at time  $m-1$  to state  $j$  at time  $m$ . From the property of conditional expectations we can write

$$\begin{aligned} P\{X_m = j, X_0 = i\} &= \sum_k P\{X_m = j, X_{m-1} = k, X_0 = i\} \\ &= \sum_k P\{X_m = j | X_{m-1} = k, X_0 = i\} \\ &\quad \times P\{X_{m-1} = k | X_0 = i\} P\{X_0 = i\} \end{aligned} \quad (20.9.9)$$

Dividing Eq. (20.9.9) throughout by  $P\{X_0 = i\}$  and using the Markov property,  $P\{X_m = j | X_{m-1} = k, X_0 = i\} = P\{X_m = j | X_{m-1} = k\}$ , we obtain the following equation

for all  $i, j$ :

$$\begin{aligned} P\{X_m = j \mid X_0 = i\} &= p_{ij}^{(m)} = \sum_k P\{X_m = j \mid X_{m-1} = k\} P\{X_{m-1} = k \mid X_0 = i\} \\ &= \sum_k p_{kj} p_{ik}^{(m-1)} \end{aligned} \quad (20.9.10)$$

This equation is a restatement of the Chapman–Kolmogorov equation [Eq. (20.8.12)]. Similar equations can be written for  $p_{ij}^{(m-1)}$  in terms of  $p_{kj}$  and  $p_{ik}^{(m-2)}$ . Thus, in matrix form

$$\mathbf{P}^{(m)} = \mathbf{P}\mathbf{P}^{(m-1)} \quad \text{or} \quad \mathbf{P}^{(m)} = \mathbf{P}^m \quad (20.9.11)$$

Eq. (20.9.11) is an important result that shows the  $m$ -step transition matrix  $\mathbf{P}^{(m)}$  as the  $m$ th power of the one-step transition matrix  $\mathbf{P}$ .

### State Occupancy Probabilities

Let the row vector  $\mathbf{p}^{(n)} = [p_0^{(n)}, p_1^{(n)}, \dots, p_N^{(n)}]$  represent the state occupancy probabilities at time  $n$  and the row vector  $\mathbf{p}^{(0)} = [p_0^{(0)}, p_1^{(0)}, \dots, p_N^{(0)}]$  represent the initial probability of the states at time 0. Since  $\mathbf{p}$  is the state transition probability matrix, we can obtain the state occupancy probabilities  $\mathbf{p}^{(n)}$  of any state at time  $n$  from a knowledge of the probability of the state at time  $n-1$  from the following compact matrix relation:

$$\mathbf{p}^{(n)} = \mathbf{p}^{(n-1)}\mathbf{P} \quad (20.9.12)$$

This equation can be solved recursively, and the probabilities of the states at time  $n$  can be related to the initial probabilities of the states at time 0 as given below:

$$\mathbf{p}^{(n)} = \mathbf{p}^{(n-1)}\mathbf{P} = \mathbf{p}^{(n-2)}\mathbf{P}^2 = \dots = \mathbf{p}^{(0)}\mathbf{P}^n \quad (20.9.13)$$

The next logical question that may arise is whether after a sufficiently long time a statistical equilibrium can be established in the sense that the occupancy probabilities of the states are independent of the initial probabilities. To answer that question, we set  $\lim_{n \rightarrow \infty} \mathbf{p}^{(n)} = \boldsymbol{\pi}$  in Eq. (20.9.12) and obtain

$$\lim_{n \rightarrow \infty} \mathbf{p}^{(n)} = \lim_{n \rightarrow \infty} \mathbf{p}^{(n-1)}\mathbf{P} \quad \text{or} \quad \boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P} \quad \text{or} \quad \boldsymbol{\pi}(\mathbf{I} - \mathbf{P}) = 0 \quad (20.9.14)$$

The matrix equation  $\boldsymbol{\pi}(\mathbf{I} - \mathbf{P})$  is a homogeneous system of linear equations and has a solution only if the determinant of  $(\mathbf{I} - \mathbf{P})$  is equal to 0. This is very similar to the eigenvector [Eq. (16.2.2)]. Solution of Eq. (20.9.14) will give the row vector  $\boldsymbol{\pi}$  only upto a multipli- cative constant, and it can be made unique by the normalization equation  $\sum_{i=0}^N \pi_i = 1$ . By utilizing the techniques of Chapter 16, we can obtain the solution for a finite Markov chain by diagonalizing the matrix  $\mathbf{P}$ .

The eigenvalues of the matrix  $\mathbf{P}$  are obtained by forming the characteristic equation  $|\lambda\mathbf{I} - \mathbf{P}| = \lambda^N + a_1\lambda^{N-1} + \dots + a_N = 0$  and solving for the  $N$  eigenvalues. If we substitute  $\lambda = 1$  in the eigenequation  $\lambda\mathbf{I} - \mathbf{P} = \mathbf{0}$ , we obtain  $\mathbf{I} - \mathbf{P} = \mathbf{0}$ . Since the row sum of the stochastic matrix is 1 and all the terms are positive, we conclude that  $\lambda = 1$  is always an eigenvalue of the stochastic matrix  $\mathbf{P}$  and all the other eigenvalues will be less than 1. We can diagonalize the matrix  $\mathbf{P}$  assuming that the eigenvalues are distinct and write

$$\mathbf{M}^{-1}\mathbf{P}\mathbf{M} = \boldsymbol{\Lambda} \quad \text{or} \quad \mathbf{P} = \mathbf{M}\boldsymbol{\Lambda}\mathbf{M}^{-1} \quad (20.9.15)$$

where  $\mathbf{M}$  is the modal matrix of eigenvectors and  $\mathbf{\Lambda}$  is a diagonal matrix of distinct eigenvalues of  $\mathbf{P}$ . From Eq. (20.9.15) we can write the  $\lim_{n \rightarrow \infty} \mathbf{P}^n$  as follows:

$$\Pi = \lim_{n \rightarrow \infty} \mathbf{P}^n = \lim_{n \rightarrow \infty} [\mathbf{M} \mathbf{\Lambda} \mathbf{M}^{-1}]^n = \mathbf{M} \lim_{n \rightarrow \infty} \mathbf{\Lambda}^n \mathbf{M}^{-1} \quad (20.9.16)$$

As  $n \rightarrow \infty$ , since all except one of the eigenvalues in  $\mathbf{\Lambda}$  are less than 1, they all tend to 0 as shown below:

$$\lim_{n \rightarrow \infty} \mathbf{\Lambda}^n = \lim_{n \rightarrow \infty} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_N \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \quad (20.9.17)$$

It can be shown that for a stochastic matrix with distinct eigenvalues, the eigenvector corresponding to the eigenvalue 1 is given by  $\mathbf{m} = [1/\sqrt{N}, 1/\sqrt{N}, \dots, 1/\sqrt{N}]^T$ .

Substituting Eq. (20.9.17) in Eq. (20.9.16) and expanding, we obtain

$$\begin{aligned} \Pi &= \lim_{n \rightarrow \infty} \mathbf{P}^n = \mathbf{P}^\infty \\ &= \begin{bmatrix} \frac{1}{\sqrt{N}} & m_{01} & \cdots & m_{0N} \\ \frac{1}{\sqrt{N}} & m_{11} & \cdots & m_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{N}} & m_{N1} & \cdots & m_{NN} \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} m'_{00} & m'_{01} & \cdots & m'_{0N} \\ m'_{10} & m'_{11} & \cdots & m'_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ m'_{N0} & m'_{N1} & \cdots & m'_{NN} \end{bmatrix} \end{aligned} \quad (20.9.18)$$

where the primes on  $m$  denote the elements of the inverse matrix  $\mathbf{M}^{-1} = \mathbf{M}'$  of the modal matrix  $\mathbf{M}$ . Equation (20.9.18) can be simplified as

$$\Pi = \begin{bmatrix} \frac{1}{\sqrt{N}} \\ \frac{1}{\sqrt{N}} \\ \vdots \\ \frac{1}{\sqrt{N}} \end{bmatrix} \begin{bmatrix} m'_{00} & m'_{01} & \cdots & m'_{0N} \end{bmatrix} \begin{bmatrix} \frac{m'_{00}}{\sqrt{N}} & \frac{m'_{01}}{\sqrt{N}} & \cdots & \frac{m'_{0N}}{\sqrt{N}} \\ \frac{m'_{10}}{\sqrt{N}} & \frac{m'_{11}}{\sqrt{N}} & \cdots & \frac{m'_{1N}}{\sqrt{N}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{m'_{N0}}{\sqrt{N}} & \frac{m'_{N1}}{\sqrt{N}} & \cdots & \frac{m'_{NN}}{\sqrt{N}} \end{bmatrix} \quad (20.9.19a)$$

or

$$\Pi = \begin{bmatrix} \pi_0 & \pi_1 & \cdots & \pi_N \\ \pi_0 & \pi_1 & \cdots & \pi_N \\ \vdots & \vdots & \ddots & \vdots \\ \pi_0 & \pi_1 & \cdots & \pi_N \end{bmatrix} \quad (20.9.19b)$$

is the steady-state probability transition matrix. Note that in the  $\Pi$  matrix each column has the same elements. Substitution of Eq. (20.9.19b) in Eq. (20.9.14) results in

$$\lim_{n \rightarrow \infty} \mathbf{p}^{(n)} = \mathbf{p}^{(0)} \mathbf{P}^\infty = \mathbf{p}^{(0)} \Pi = [\pi_0 \ \pi_1 \ \cdots \ \pi_N] = \pi \quad (20.9.20)$$

This equation shows that the limiting state occupancy probabilities are independent of the initial state probability row vector  $\mathbf{p}^{(0)}$ .

**Example 20.9.1** A two-state Markov chain with probability transition matrix  $\mathbf{P}$  as in Eq. (20.9.5) is shown below along with the initial probability row vector  $\mathbf{p}^{(0)}$ :

$$\mathbf{P} = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix}; \quad \mathbf{p}^{(0)} = \begin{bmatrix} p_0^{(0)} & p_1^{(0)} \end{bmatrix} = [p \quad 1 - p]$$

We want to find the equilibrium probability distributions of the occupancy of states 1,  $-1$  and show that they are independent of the initial occupancy  $p_0^{(0)}$  and  $p_1^{(0)}$ .

The eigenvalues of  $\mathbf{P}$  are obtained by solving the characteristic equation given by

$$|\lambda \mathbf{I} - \mathbf{P}| = \begin{vmatrix} \lambda - 1 + \alpha & -\alpha \\ -\beta & \lambda - 1 + \beta \end{vmatrix} = (\lambda - 1)(\lambda - 1 - \alpha + \beta) = 0$$

and the eigenvalues  $\lambda_1 = 1$  and  $\lambda_2 = 1 - \alpha - \beta$ . The unnormalized eigenvector  $\psi_1$  corresponding to  $\lambda_1 = 1$  is obtained from the eigenequation

$$\begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} \begin{bmatrix} \psi_{21} \\ \psi_{22} \end{bmatrix} = 0 \quad \text{or} \quad \begin{bmatrix} \psi_{11} \\ \psi_{12} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and the unnormalized eigenvector  $\psi_2$  corresponding to  $\lambda_2 = 1 - \alpha - \beta$  is obtained from

$$\begin{bmatrix} -\beta & -\alpha \\ -\beta & -\alpha \end{bmatrix} \begin{bmatrix} \psi_{21} \\ \psi_{22} \end{bmatrix} = 0 \quad \text{or} \quad \begin{bmatrix} \psi_{21} \\ \psi_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ -\frac{\beta}{\alpha} \end{bmatrix}$$

The matrices of unnormalized eigenvectors  $\mathbf{M}_U$  and its inverse  $\mathbf{M}_U^{-1}$  are

$$\mathbf{M}_U = \begin{bmatrix} 1 & 1 \\ 1 - \frac{\beta}{\alpha} & 0 \end{bmatrix} \quad \mathbf{M}_U^{-1} = \begin{bmatrix} \frac{\beta}{\beta + \alpha} & \frac{\alpha}{\beta + \alpha} \\ \frac{\alpha}{\beta + \alpha} & -\frac{\alpha}{\beta + \alpha} \end{bmatrix}$$

The diagonal eigenvalue matrix is  $\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 1 - \alpha - \beta \end{bmatrix}$ , and since  $1 - \alpha - \beta < 1$ , the limit as  $n \rightarrow \infty$  of  $\Lambda^n$  is given by  $\lim_{n \rightarrow \infty} \Lambda^{(n)} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ . Hence, as  $n \rightarrow \infty$ ,  $\mathbf{P}^n$  is given by

$$\lim_{n \rightarrow \infty} \mathbf{P}^{(n)} = \begin{bmatrix} 1 & 1 \\ 1 & -\frac{\beta}{\alpha} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{\beta}{\beta + \alpha} & \frac{\alpha}{\beta + \alpha} \\ \frac{\alpha}{\beta + \alpha} & -\frac{\alpha}{\beta + \alpha} \end{bmatrix} = \begin{bmatrix} \frac{\beta}{\beta + \alpha} & \frac{\alpha}{\beta + \alpha} \\ \frac{\alpha}{\beta + \alpha} & -\frac{\alpha}{\beta + \alpha} \end{bmatrix}$$

and

$$\pi = \mathbf{p}^{(\infty)} = \mathbf{p}^{(0)} \mathbf{P}^{\infty} = [p \quad 1 - p] \begin{bmatrix} \frac{\beta}{\beta + \alpha} & \frac{\alpha}{\beta + \alpha} \\ \frac{\beta}{\beta + \alpha} & \frac{\alpha}{\beta + \alpha} \end{bmatrix} = \begin{bmatrix} \frac{\beta}{\beta + \alpha} & \frac{\alpha}{\beta + \alpha} \\ \frac{\alpha}{\beta + \alpha} & \frac{\alpha}{\beta + \alpha} \end{bmatrix}$$

which is independent of the initial probability vector  $\mathbf{p}^{(0)} = [p \quad 1 - p]$ .

**Example 20.9.2** A book representative visits three universities  $a$ ,  $b$ , and  $c$  each week. However, there is a finite probability that she may stay over for another week in any of the three places. The transition probability matrix  $\mathbf{P}$  and the initial probability row

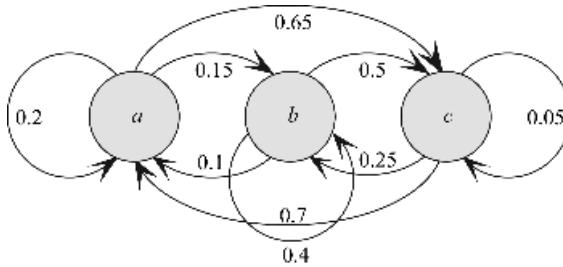


FIGURE 20.9.2

vector  $\mathbf{p}^{(0)}$  of being in states  $a$ ,  $b$ , and  $c$  at time 0 are shown below:

$$\mathbf{P} = \begin{bmatrix} a & b & c \\ 0.2 & 0.15 & 0.65 \\ 0.1 & 0.4 & 0.5 \\ 0.7 & 0.25 & 0.05 \end{bmatrix} \quad \mathbf{p}^{(0)} = \begin{bmatrix} p_a^{(0)} \\ p_b^{(0)} \\ p_c^{(0)} \end{bmatrix}^T = \begin{bmatrix} 0.3 \\ 0.45 \\ 0.25 \end{bmatrix}^T$$

We have to find (1) the probabilities of the representative being in the universities  $a$ ,  $b$ , and  $c$  after 3 weeks and (2) the steady-state probability of being in any of the universities. The state transition flow diagram for this three-state Markov chain is shown in Fig. 20.9.2.

1. We can apply Eq. (20.9.14) to obtain the probabilities that she will be in the various universities after 3 weeks. Thus

$$\begin{aligned} \mathbf{p}^{(3)} &= \begin{bmatrix} p_a^{(3)} \\ p_b^{(3)} \\ p_c^{(3)} \end{bmatrix}^T = \mathbf{p}^{(0)} \mathbf{P}^3 = \begin{bmatrix} 0.3 \\ 0.45 \\ 0.25 \end{bmatrix}^T \cdot \begin{bmatrix} 0.2 & 0.15 & 0.65 \\ 0.1 & 0.4 & 0.5 \\ 0.7 & 0.25 & 0.05 \end{bmatrix}^3 \\ &= \begin{bmatrix} 0.3 \\ 0.45 \\ 0.25 \end{bmatrix}^T \cdot \begin{bmatrix} 0.293 & 0.237 & 0.47 \\ 0.315 & 0.254 & 0.431 \\ 0.47 & 0.262 & 0.268 \end{bmatrix} = \begin{bmatrix} 0.347 \\ 0.251 \\ 0.402 \end{bmatrix} \end{aligned}$$

2. The eigenvalues of the matrix  $\mathbf{P}$  are  $\lambda_1 = 1$ ,  $\lambda_2 = 0.23197$ ,  $\lambda_3 = -0.58197$ .

The modal matrix  $\mathbf{M}$  is:

$$\mathbf{M} = \begin{bmatrix} 0.57735 & -0.37197 & -0.56601 \\ 0.57735 & -0.89737 & -0.32757 \\ 0.57735 & -0.22574 & 0.75652 \end{bmatrix}$$

and

$$\mathbf{P}^\infty = \mathbf{M} \Lambda^\infty \mathbf{M}^{-1}$$

$$\begin{aligned} &= \begin{bmatrix} 0.57735 & -0.37197 & -0.56601 \\ 0.57735 & -0.89737 & -0.32757 \\ 0.57735 & -0.22574 & 0.75652 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.63437 & 0.43479 & 0.66288 \\ -0.65635 & 0.80072 & -0.14437 \\ -0.67997 & -0.0929 & 0.77287 \end{bmatrix} \\ &= \begin{bmatrix} 0.36626 & 0.25103 & 0.38272 \\ 0.36626 & 0.25103 & 0.38272 \\ 0.36626 & 0.25103 & 0.38272 \end{bmatrix} \end{aligned}$$

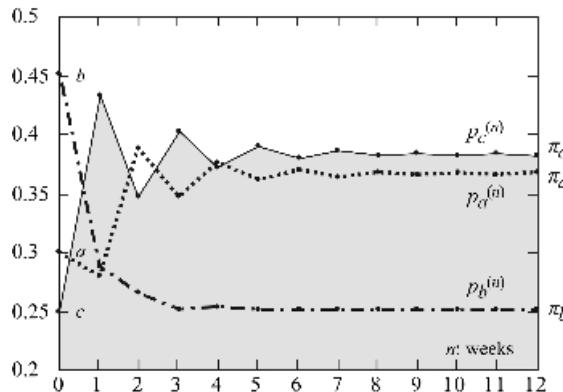


FIGURE 20.9.3

and the steady-state probability  $\pi = \mathbf{p}^{(\infty)} = \mathbf{p}^{(0)}\mathbf{P} = [0.36626 \ 0.25103 \ 0.38272]$ . The evolution of state occupancy probabilities  $\mathbf{p}^{(n)}$  for all three states are shown in Fig. 20.9.3 for  $n = 1, 2, \dots, 12$ . From the figure it is seen that steady-state equilibrium has been obtained in about 9 weeks.

**Example 20.9.3** A new automobile can be classified into five states: 1 = excellent, 2 = good, 3 = fair, 4 = poor, 5 = scrap. The transition probabilities of going from one state to another in each year is given by the transition probability matrix  $\mathbf{P}$ . Clearly, the initial probability row vector  $\mathbf{p}$  will be given by  $\mathbf{p} = [1 \ 0 \ 0 \ 0 \ 0]$ . At the end of the first year the probabilities of the automobile being in excellent condition (state 1) = 0.6, in good condition (state 2) = 0.25, in fair condition (state 3) = 0.1 and in poor condition (state 4) = 0.05. Similarly, if the automobile is in scrap condition (state 5), the probabilities of it continuing to be in scrap condition = 0.85, in poor condition = 0.1, and in fair condition = 0.05. We will find the probabilities of the condition of the car in (1) 5 years and (2) steady-state equilibrium for the state occupancy probabilities:

$$\mathbf{P} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0.6 & 0.25 & 0.1 & 0.05 & 0 \\ 0.2 & 0.45 & 0.2 & 0.1 & 0.05 \\ 0.05 & 0.15 & 0.4 & 0.25 & 0.15 \\ 0 & 0.05 & 0.15 & 0.35 & 0.45 \\ 0 & 0 & 0.05 & 0.1 & 0.85 \end{bmatrix}$$

The state transition flow diagram is shown in Fig. 20.9.4.

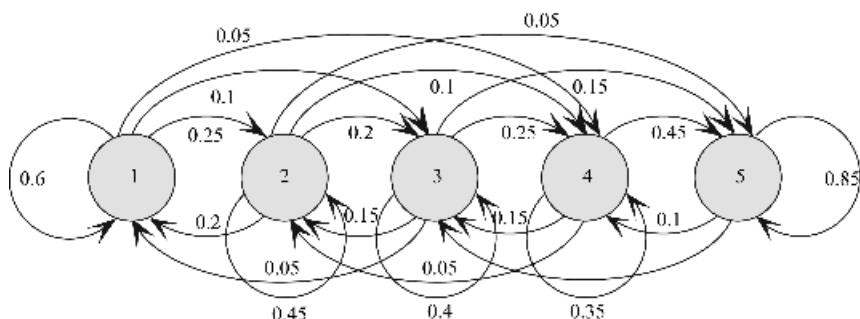


FIGURE 20.9.4

1. From Eq. (20.9.13) the state probabilities after  $n = 5$  years is given by  $\mathbf{p}^{(5)} = \mathbf{p}^{(0)}\mathbf{P}^5$ , where  $\mathbf{p}^{(0)}$  is the row vector  $\mathbf{p}^{(0)} = [1 \ 0 \ 0 \ 0 \ 0]$ . Hence

$$\begin{aligned}\mathbf{p}^{(5)} &= \mathbf{p}^{(0)}\mathbf{P}^5 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} 0.2034 & 0.1984 & 0.1736 & 0.1502 & 0.2744 \\ 0.1506 & 0.1618 & 0.1641 & 0.1571 & 0.3664 \\ 0.0863 & 0.1090 & 0.1447 & 0.1615 & 0.4985 \\ 0.0419 & 0.0656 & 0.1215 & 0.1572 & 0.6138 \\ 0.0204 & 0.0415 & 0.1061 & 0.1518 & 0.6802 \end{bmatrix} \\ &= \begin{bmatrix} 0.2034 \\ 0.1984 \\ 0.1736 \\ 0.1502 \\ 0.2744 \end{bmatrix}^T\end{aligned}$$

Thus, after 5 years there is a probability of 0.2034 that the car is still in excellent condition and a probability of 0.2744 that the car is in scrap condition. It is also interesting that with the given transition probabilities, there is a better chance that the car is in excellent condition than in either good, fair, or poor condition. We will now check what happens to these probabilities after a long period of time has elapsed.

2. We follow the procedure of finding the eigenvalues and eigenvectors of the transition matrix  $\mathbf{P}$ . The eigenvalue matrix  $\Lambda$ , the corresponding modal matrix of  $\mathbf{P}$ , and its inverse are as follows:

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.4529 & 0 & 0 & 0 \\ 0 & 0 & 0.2417 & 0 & 0 \\ 0 & 0 & 0 & 0.7999 & 0 \\ 0 & 0 & 0 & 0 & 0.1555 \end{bmatrix} \quad \Lambda^\infty = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{M} = \begin{bmatrix} 0.4472 & 0.7139 & -0.4491 & 0.7891 & 0.1265 \\ 0.4472 & -0.1188 & 0.8065 & 0.5471 & -0.4190 \\ 0.4472 & -0.6081 & -0.2789 & 0.2238 & 0.7403 \\ 0.4472 & -0.2899 & -0.2566 & -0.0293 & -0.5099 \\ 0.4472 & 0.1496 & 0.0651 & -0.1647 & 0.0201 \end{bmatrix}$$

$$\mathbf{M}^{-1} = \begin{bmatrix} 0.1114 & 0.1555 & 0.2695 & 0.3446 & 1.3551 \\ 0.4614 & -0.1919 & -0.5895 & -0.5495 & 0.8695 \\ -0.3828 & 0.7069 & -0.0846 & -0.7773 & 0.5378 \\ 0.5697 & 0.5102 & 0.2363 & 0.0127 & -1.3288 \\ -0.0048 & -0.1395 & 0.6005 & -0.9562 & 0.4999 \end{bmatrix}$$

Hence from Eq. (20.9.18) we have

$$\Pi = \mathbf{P}^\infty = \mathbf{M} \boldsymbol{\Lambda}^\infty \mathbf{M}^{-1} = \begin{bmatrix} 0.0499 & 0.0695 & 0.1205 & 0.1541 & 0.6060 \\ 0.0499 & 0.0695 & 0.1205 & 0.1541 & 0.6060 \\ 0.0499 & 0.0695 & 0.1205 & 0.1541 & 0.6060 \\ 0.0499 & 0.0695 & 0.1205 & 0.1541 & 0.6060 \\ 0.0499 & 0.0695 & 0.1205 & 0.1541 & 0.6060 \end{bmatrix}$$

and the steady-state probabilities row vector  $\boldsymbol{\pi}$  is

$$\boldsymbol{\pi} = \mathbf{p}^{(0)} \Pi = [0.0499 \ 0.0695 \ 0.1205 \ 0.1541 \ 0.6060]$$

These are more reasonable numbers since the long-term probabilities of being in various states are excellent 4.99%, good 6.95%, fair 12.05%, poor 15.41%, and finally scrap 60.6%. The conclusion is that there is 60.6% chance that the automobile will be a scrap in the long run with the given probability transition matrix.

### First-Passage Probabilities and Classification of States

A Markov chain is called *irreducible* if every state can be reached from every other state in a finite number of steps. In this case, the probability of going from every state  $i$  to every other state  $j$  in the state space  $S$  of the chain will be given by the probability,  $p_{ij}^{(m)} > 0$ , where  $m$  is a finite integer. In this case, the states are called *intercommunicating*. On the other hand, if  $E$  is a proper subset of  $S$ ,  $E \subset S$ , and if  $i$  and  $j$  are states in  $E$ , then the subset  $E$  is *closed* if

$$\sum_{j \in E} p_{ij} = 1 \quad \text{for all } i \in E$$

In other words, no transition is possible from a state  $i \in E$  to a state  $k \in E^c$ . In this case the Markov chain is called *reducible*. If there are no other proper subsets of  $E$  that are closed, then it is called an *irreducible sub-Markov chain*. If  $i \in E$  is the only state, then it is called an *absorbing state*, in which case  $p_{ii} = 1$ .

Let us consider that the Markov chain is initially in state  $i$  at time  $n$ . We define the probability of *first return* to  $i$  at time  $n+m$  as  $f_{ii}^{(m)}$ , or

$$f_{ii}^{(m)} = P\{X_{n+1} \neq i, \dots, X_{n+m-1} \neq i, X_{n+m} = i \mid X_n = i\} \quad (20.9.21)$$

with one-step transition  $f_{ii}^{(1)} = P\{X_{n+1} = i \mid X_n = i\} = p_{ii}^{(1)} = p_{ii}$  similar to Eq. (20.9.2).

For a homogeneous Markov chain, Eq. (20.9.21) can also be written as

$$f_{ii}^{(m)} = P\{X_1 \neq i, \dots, X_{m-1} \neq i, X_m = i \mid X_0 = i\} \quad (20.9.22)$$

with one-step transition  $f_{ii}^{(1)} = P\{X_1 = i \mid X_0 = i\} = p_{ii}^{(1)} = p_{ii}$ .

In a similar manner we can also define the probability of *first passage* from state  $i$  at time  $n$  to state  $j$  at time  $n+m$  as  $f_{ij}^{(m)}$ , or

$$f_{ij}^{(m)} = P\{X_{n+1} \neq j, \dots, X_{n+m-1} \neq j, X_{n+m} = j \mid X_n = i\} \quad (20.9.23)$$

with one-step transition  $f_{ij}^{(1)} = P\{X_{n+1} = j | X_n = i\} = p_{ij}^{(1)} = p_{ij}$  in conformity with Eq. (20.9.2). For a homogeneous Markov chain, Eq. (20.9.23) is equivalent to

$$f_{ij}^{(m)} = P\{X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j | X_0 = i\} \quad (20.9.24)$$

with the corresponding one-step probability  $f_{ij}^{(1)} = P\{X_1 = j | X_0 = i\} = p_{ij}^{(1)} = p_{ij}$ .

The probabilities of eventual first return to state  $i$  or the first passage to state  $j$  starting from state  $i$  are given by

$$f_i = \sum_{m=1}^{\infty} f_{ii}^{(m)} \quad f_{ij} = \sum_{m=1}^{\infty} f_{ij}^{(m)} \quad (20.9.25)$$

It is now possible to have more classifications of the Markov chains depending on the nature of  $f_i$ . If  $f_i = 1$ , the state  $i$  is called *recurrent*; if  $f_i < 1$ , the state  $i$  is called *transient*. If the chain starts at time 0 in state  $i$  and returns to state  $i$  only at times  $\alpha, 2\alpha, \dots$ , with period  $\alpha > 1$ , then the state  $i$  is called *periodic* with period  $\alpha$ . If  $\alpha = 1$ , then  $i$  is *aperiodic*.

If  $f_i = 1$  and  $f_{ij} = 1$ , we can define a *mean recurrence time* and a *mean first-passage time* from state  $i$  to state  $j$  as follows:

$$\mu_i = \sum_{m=1}^{\infty} mf_i \quad \mu_{ij} = \sum_{m=1}^{\infty} mf_{ij} \quad (20.9.26)$$

With this knowledge we can classify the states even further. If the mean recurrence time  $\mu_i$  is finite, then the state is called *positive recurrent*; if it is infinite, it is called *null recurrent*. Finally, a state  $i$  is *ergodic* if it is aperiodic and positive recurrent, or  $\alpha = 1, f_i = 1$  and  $\mu_i < \infty$ . A finite-state aperiodic irreducible Markov chain is always ergodic, and the occupancy probability of the states as  $n \rightarrow \infty$  will tend to an equilibrium distribution independent of the initial probabilities as in Examples 20.9.1–20.9.3. In transient states the equilibrium distribution will be 0.

**Example 20.9.4** The transition matrix  $\mathbf{P}$  of a four-state Markov chain is given below:

$$\mathbf{P} = \begin{pmatrix} & 1 & 2 & 3 & 4 \\ 1 & 1 & 0 & 0 & 0 \\ 2 & 0 & 0.4 & 0.6 & 0 \\ 3 & 0 & 0.8 & 0.2 & 0 \\ 4 & 0.3 & 0 & 0.2 & 0.5 \end{pmatrix}$$

and the state transition flow diagram is shown in Fig. 20.9.5.

We can obtain the classification of the states by inspection of the transition flow diagram of Fig. 20.9.5. From the flow diagram state 1 is absorbing because  $p_{11} = 1$ , and once the transition occurs to state 1, it never gets out of it. It is also a recurrent state. States 2 and 3 are recurring states because the transition occurs only between these two states. If the transition occurs out of state 4, it can never get back in. Hence, state 4 is a transient state. We will determine the equilibrium probabilities  $\mathbf{\Pi}$  from

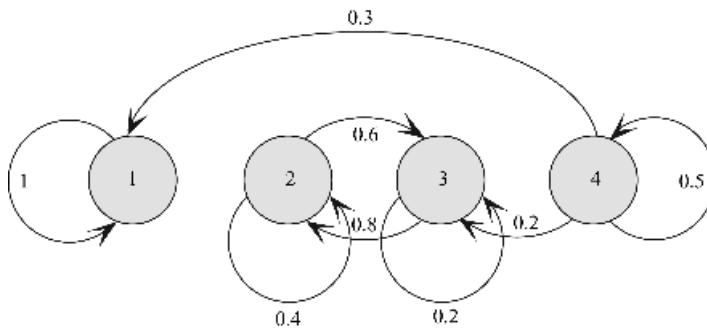


FIGURE 20.9.5

Eq. (20.9.18). The diagonal eigenvalue matrix  $\Lambda$  is given by

$$\Lambda = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -0.4 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \lim_{n \rightarrow \infty} \Lambda^n = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The presence of two 1s for eigenvalues indicates the presence of sub-Markov chains. The modal matrix and its inverse are

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 0 & 0.8575 \\ 0 & 0.6804 & -0.5907 & 0 \\ 0 & 0.6804 & 0.7877 & 0 \\ 1 & 0.2722 & -0.1750 & 0.5145 \end{bmatrix}$$

$$\mathbf{M}^{-1} = \begin{bmatrix} -0.6 & -0.3556 & -0.0444 & 1 \\ 0 & 0.8398 & 0.6299 & 0 \\ 0 & -0.7255 & 0.7255 & 0 \\ 1.1622 & 0 & 0 & 0 \end{bmatrix}$$

and the equilibrium probability  $\Pi$  is given by

$$\Pi = \mathbf{M} \lim_{n \rightarrow \infty} \Lambda^n \mathbf{M}^{-1} = \begin{cases} 1 & \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \\ 2 & \begin{bmatrix} 0 & 0.5714 & 0.4286 & 0 \end{bmatrix} \\ 3 & \begin{bmatrix} 0 & 0.5714 & 0.4286 & 0 \end{bmatrix} \\ 4 & \begin{bmatrix} 0.6 & 0.2286 & 0.1714 & 0 \end{bmatrix} \end{cases}$$

State 4 is transient since  $p_{4j} = 0, j = 1, \dots, 4$ . State 1 is absorbing since  $p_{11} = 1$ . States 2 and 3 are recurrent with  $p_{23} = 0.4286$  and  $p_{32} = 0.5714$ .

**Example 20.9.5** We will slightly change the transition probabilities by introducing a path between states 2 and 1 by inserting  $p_{21} = 0.02$  and modifying  $p_{23} = 0.58$ :

$$\mathbf{P} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 0 & 0 & 0 \\ 2 & 0.02 & 0.4 & 0.58 & 0 \\ 3 & 0 & 0.8 & 0.2 & 0 \\ 4 & 0.3 & 0 & 0.2 & 0.5 \end{pmatrix}$$

The state transition flow diagram is shown in Fig. 20.9.6.

This small change makes a big difference in the properties of the Markov chain. From the flow diagram, the only state that is recurrent is 1. The other states (2–4) are transient because everything flows into state 1 directly or indirectly. This will be confirmed from the equilibrium probabilities  $\mathbf{\Pi}$  of the states. The diagonal eigenvalue matrix  $\mathbf{\Lambda}$  and the  $\lim_{n \rightarrow \infty}$  are given by

$$\mathbf{\Lambda} = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.9885 & 0 & 0 \\ 0 & 0 & -0.3885 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \lim_{n \rightarrow \infty} \mathbf{\Lambda}^n = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Since there is only one eigenvalue of 1, there are no sub-Markov chains.

The modal matrix and its inverse are

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 0 & 0.5 \\ 0 & 0.6739 & -0.5830 & 0.5 \\ 0 & 0.6837 & 0.7926 & 0.5 \\ 1 & 0.28 & -0.1784 & 0.5 \end{bmatrix} \quad \mathbf{M}^{-1} = \begin{bmatrix} -0.5853 & -0.3687 & -0.0461 & 1 \\ -1.4748 & 0.8497 & 0.6251 & 0 \\ 0.0106 & -0.7330 & 0.7225 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix}$$

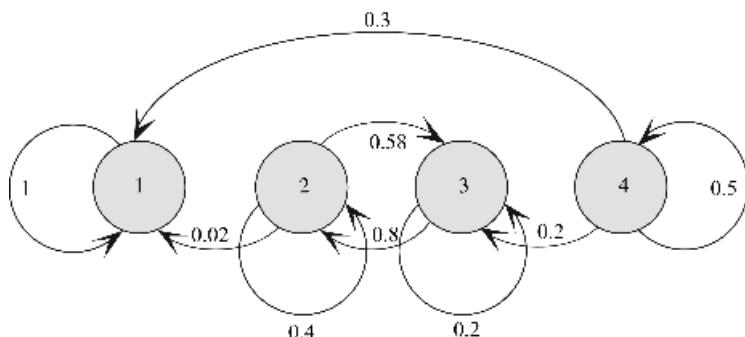


FIGURE 20.9.6

and the equilibrium probability  $\mathbf{\Pi}$  is given by

$$\mathbf{\Pi} = \mathbf{M} \lim_{n \rightarrow \infty} \mathbf{\Lambda}^n \mathbf{M}^{-1} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 & 0 \end{bmatrix}$$

The equilibrium probability matrix shows that states 2–4 are transient because the probabilities are zero. State 1 is the absorbing recurrent state.

**Example 20.9.6** A state transition matrix  $\mathbf{P}$  for a seven-state Markov chain is given by

$$\mathbf{P} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0.7 & 0.3 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0.7 & 0.3 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0.4 & 0.3 & 0.3 & 0 & 0 \\ 5 & 0 & 0 & 0.5 & 0.4 & 0.1 & 0 & 0 \\ 6 & 0.2 & 0 & 0.1 & 0.1 & 0.1 & 0.2 & 0.3 \\ 7 & 0 & 0 & 0.3 & 0.2 & 0.4 & 0 & 0.1 \end{bmatrix}$$

The corresponding state transition flow diagram is shown in Fig. 20.9.7.

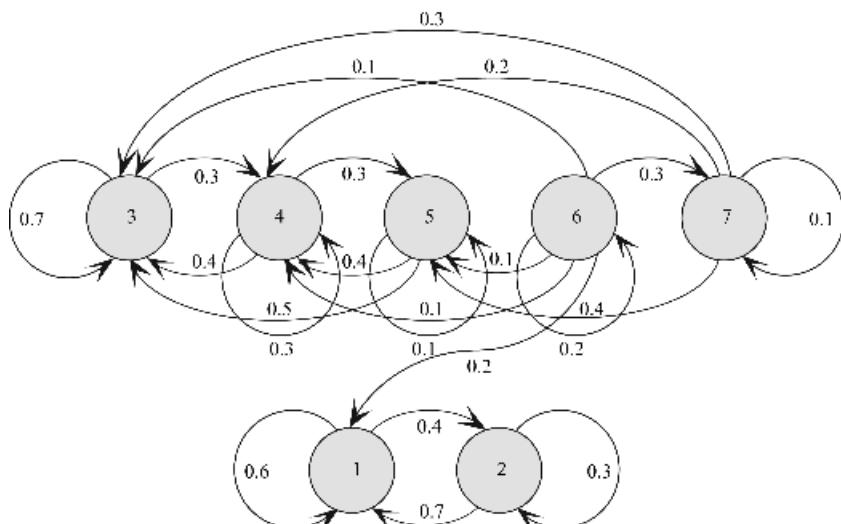


FIGURE 20.9.7

We will classify the states by inspection from Fig. 20.9.7. The flow diagram shows that states 1 and 2 form a sub-Markov chain since there are no paths out of that pair. Hence they form recurring states. States 3–5 form another sub-Markov chain, and there are no paths leading out of these three states. Hence they form another set of recurring states. All paths lead out of states 6 and 7 and hence are transient states. We can confirm these findings by computing the equilibrium probability  $\mathbf{\Pi}$ .

The diagonal eigenvalue matrix  $\mathbf{\Lambda}$  and the  $\lim_{n \rightarrow \infty} \mathbf{\Lambda}^n$  are given by

$$\mathbf{\Lambda} = \begin{bmatrix} 0.2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.2303 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.1303 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.1 \end{bmatrix}$$

$$\lim_{n \rightarrow \infty} \mathbf{\Lambda}^n = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Since there are two eigenvalues of 1, we conclude that there is a sub-Markov chain. The modal matrix  $\mathbf{M}$  and its inverse  $\mathbf{M}^{-1}$  are given by

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0.6963 & -0.4710 \\ 0 & 0 & 0 & 0 & 0 & 0.6963 & 0.8243 \\ 0 & 0 & 0.4682 & -0.028 & -0.1665 & 0 & 0 \\ 0 & 0 & 0.4682 & 0.0438 & 0.4607 & 0 & 0 \\ 0 & 0 & 0.4682 & 0.0271 & -0.4388 & 0 & 0 \\ 1 & 0.9487 & 0.3511 & 0.9946 & -0.4821 & 0.1741 & 0.3140 \\ 0 & -0.3162 & 0.4682 & 0.0861 & 0.5789 & 0 & 0 \end{bmatrix}$$

$$\mathbf{M}^{-1} = \begin{bmatrix} 0.0833 & -0.3333 & 18.25 & -8.75 & -13.25 & 1 & 3 \\ 0 & 0 & -2.1082 & 4.2164 & 1.0541 & 0 & -3.1623 \\ 0 & 0 & 1.2521 & 0.6629 & 0.2210 & 0 & 0 \\ 0 & 0 & -16.6314 & 5.0356 & 11.5958 & 0 & 0 \\ 0.9139 & 0.5222 & 0 & 0 & 0 & 0 & 0 \\ -0.7720 & 0.7720 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The equilibrium probability  $\mathbf{\Pi}$  is given by

$$\mathbf{\Pi} = \mathbf{M} \lim_{n \rightarrow \infty} \mathbf{\Lambda}^n \mathbf{M}^{-1} = \left( \begin{array}{cccccc|cc} & & & & & & & 6 & 7 \\ & 1 & 2 & | & 3 & 4 & 5 & | & \\ & 0.6364 & 0.3636 & | & 0 & 0 & 0 & | & 0 & 0 \\ & 0.6364 & 0.3636 & | & 0 & 0 & 0 & | & 0 & 0 \\ & 0 & 0 & | & 0.5862 & 0.3103 & 0.1035 & | & 0 & 0 \\ & 0 & 0 & | & 0.5862 & 0.3103 & 0.1035 & | & 0 & 0 \\ & 0 & 0 & | & 0.5862 & 0.3103 & 0.1035 & | & 0 & 0 \\ & 0.1591 & 0.0909 & | & 0.4396 & 0.2328 & 0.0776 & | & 0 & 0 \\ & 0 & 0 & | & 0.5862 & 0.3103 & 0.1035 & | & 0 & 0 \end{array} \right)$$

The matrix  $\mathbf{\Pi}$  shows that there are two closed sets of recurring states {1,2} and {3,4,5} with equilibrium probabilities. States 6 and 7 have zero probabilities, thus confirming that they are transient states. An interesting point to note is that there is no direct transition from state 6 to state 2 in the state transition matrix but there is a direct transition from state 6 to state 2 in the equilibrium probability distribution.

### Properties of Markov Chains

1. Let  $X = \{X_n: n = 0, 1, \dots, N\}$  be a Markov chain with state space  $S = \{i, j, k, \dots\}$ . The  $m$ -step transition probability  $p_{ij}^{(m)}(n)$  is the probability of  $X_{n+1} = j$  given that  $X_{n=i}$ , or

$$P\{X_{n+m} = j|X_n = i\} = P\{X_{n+m} = j|X_n = i\} = p_{ij}^{(m)}(n)$$

The Markov chain is homogeneous when

$$P\{X_{n+m} = j|X_n = i\} = P\{X_m = j|X_0 = i\} = p_{ij}^{(m)}$$

The transition probabilities are independent of time  $n$ .

2. If  $\mathbf{P}$  is the matrix of one-step transition probabilities, then  $p_{ij}$  is the  $(i, j)$ th element in  $\mathbf{P}$  as shown below:

$$\mathbf{P} = \left( \begin{array}{cccccc|cc} & 0 & 1 & \dots & j & \dots & N & \\ & p_{00} & p_{01} & \dots & p_{0j} & \dots & p_{0N} & \\ & p_{10} & p_{11} & \dots & p_{1j} & \dots & p_{1N} & \\ & \vdots & \vdots & & \vdots & & \vdots & \\ i & p_{i0} & p_{i1} & \dots & p_{ij} & \dots & p_{iN} & \\ & \vdots & \vdots & & \vdots & & \vdots & \\ N & p_{N0} & p_{N1} & \dots & p_{Nj} & \dots & p_{NN} & \end{array} \right)$$

$\mathbf{P}$  is called the *stochastic matrix* or *Markov matrix* with row sum equal to 1, or

$$\sum_{j=0}^N p_{ij} = 1, \quad i = 0, 1, \dots, N$$

3. The Markov chain satisfies the Chapman–Kolmogorov equation

$$p_{ij}^{(m)} = \sum_k P\{X_m = j|X_{m-1} = k\} P\{X_{m-1} = k|X_0 = i\} = \sum_k p_{kj} p_{ik}^{(m-1)}$$

4. Let the row vector of the initial probability of the states of a Markov chain at time 0 be  $\mathbf{p}^{(0)} = [p_0^{(0)}, p_1^{(0)}, \dots, p_N^{(0)}]$ . Then the row vector of state occupancy probabilities at time  $n$ ,  $\mathbf{p}^{(n)} = [p_0^{(n)}, p_1^{(n)}, \dots, p_N^{(n)}]$ , is given by

$$\mathbf{p}^{(n)} = \mathbf{p}^{(n-1)} \mathbf{P} = \mathbf{p}^{(0)} \mathbf{P}^n$$

5. A Markov chain with Markov matrix  $\mathbf{P}$  is called *irreducible* if every state can be reached from every other state in a finite number of steps. The long-term statistical equilibrium occupancy probabilities  $\boldsymbol{\pi}$  of the states can be obtained from solving

$$\boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{P} \quad \text{or} \quad \boldsymbol{\pi}(\mathbf{I} - \mathbf{P}) = 0$$

6. Only one eigenvalue of an irreducible Markov chain is 1, and the absolute values of all other eigenvalues are less than 1.
7. Starting from a state  $i$  at time  $n$ , the probability of *first return* to  $i$  at time  $n+m$  as  $f_{ii}^{(m)}$  is defined by

$$f_{ii}^{(m)} = P\{X_{n+1} \neq i, \dots, X_{n+m-1} \neq i, X_{n+m} = i | X_n = i\}$$

and starting from a state  $i$  at time  $n$  the probability of *first passage* to  $j$  at time  $n+m$  as  $f_{ij}^{(m)}$  is defined by

$$f_{ij}^{(m)} = P\{X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j | X_0 = i\}$$

The probabilities of eventual first return to state  $i$  or first passage to state  $j$  are

$$f_i = \sum_{m=1}^{\infty} f_{ii}^{(m)} \quad f_{ij} = \sum_{m=1}^{\infty} f_{ij}^{(m)}$$

If  $f_i$  or  $f_{ij}$  is equal to 1, then the state  $i$  or  $j$  is certain, and both are called *recurrent*. If they are less than 1, then the states are called *transient*.

8. The mean recurrence time and the mean first passage time can be defined as

$$\mu_i = \sum_{m=1}^{\infty} m f_i \quad \mu_{ij} = \sum_{m=1}^{\infty} m f_{ij}$$

If  $\mu_i < \infty$ , then the state  $i$  is positive recurrent, and if  $\mu_i = \infty$ , then  $i$  is null recurrent and similar definitions apply to state  $j$  if  $\mu_{ij} < \infty$ , or  $\mu_{ij} = \infty$

9. In a Markov chain with transition matrix  $\mathbf{P}$ , if  $E$  is a proper subset of the state space  $S$ , or if  $E \subset S$ , then  $E$  is closed if  $\sum_{j \in E} p_{ij} = 1$  for all  $i \in E$ . The states in  $E$  form a sub-Markov chain in which case there will be two eigenvalues of 1.
10. A closed set of states  $E \subset S$  containing no proper subsets that are closed is called *irreducible*. If the number of states within an irreducible set is finite, then each state in  $E$  is recurrent. For any irreducible set, every state communicates with every other state.
11. If all the states of a Markov chain are irreducible, then a steady-state row vector defined by  $\boldsymbol{\pi}$  can be determined from  $\boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{P}$ , or  $\boldsymbol{\pi}(\mathbf{I} - \mathbf{P}) = 0$ .

## 20.10 MARTINGALE PROCESS

The meaning of a *martingale* in both British and American dictionaries refers to a strap preventing a horse from rearing. This has no relevance in the mathematical sense,

where it conveys the sense of a fair game. It is a stochastic process where the best estimate of the future value conditioned on the past, including the present, is the present value. It is a process without trend and hence unpredictable. Martingale concepts are widely used in engineering and stochastic finance. Many filtering problems in engineering such as the Kalman filter [27] can be cast in a martingale framework. The option pricing model developed by Black and Scholes [7] in 1973 for pricing stock options can be formulated within the framework of martingale theory. Just like a Markov process, conditional expectations play a key role in martingale processes. For detailed study of martingales, see Refs. 16, 26, 35, 37, and 40.

A random sequence  $\{Y_n, n \geq 0\}$  is a *discrete martingale* with respect to the sequence  $\{X_m, m \geq n\}$  if the following conditions are satisfied:

- (1)  $E|Y_n| < \infty$
  - (2)  $E[Y_n|X_m, X_{m-1}, \dots, X_0] = Y_m, m \leq n$
- (20.10.1)

If the second condition in Eq. (20.10.1) is modified as

$$(2a) \quad E[Y_n|X_m, X_{m-1}, \dots, X_0] \geq Y_m, m \leq n \quad (20.10.2)$$

then the random sequence  $\{Y_n, n \geq 0\}$  is a *submartingale* with respect to the sequence  $\{X_m, m \geq 0\}$  and if modified as

$$(2b) \quad E[Y_n|X_m, X_{m-1}, \dots, X_0] \leq Y_m, m \geq n \quad (20.10.3)$$

then the random sequence  $\{Y_n, n \geq 0\}$  is a *supermartingale* with respect to the sequence  $\{X_m, m \geq 0\}$ .

A random process  $\{Y(t), t \geq 0\}$  is a *continuous martingale* with respect to the process  $\{X(s), 0 \leq s < t\}$  if the following conditions are satisfied:

- (3)  $E|Y(t)| < \infty$
  - (4)  $E[Y(t)|X(s), 0 \leq s < t] = Y(s)$
- (20.10.4)

If the second condition in Eq. (20.10.1) is modified as

$$(4a) \quad E[Y(t)|X(s), 0 \leq s < t] \leq Y(s) \quad (20.10.5)$$

then the random process  $\{Y(t), t \geq 0\}$  is a *continuous submartingale* with respect to the process  $\{X(s), 0 \leq s < t\}$  and if modified as

$$(4b) \quad E[Y(t)|X(s), 0 \leq s < t] \geq Y(s) \quad (20.10.6)$$

then the random process  $\{Y(t), t \geq 0\}$  is a *continuous supermartingale* with respect to the underlying process  $\{X(s), 0 \leq s < t\}$ .

We can explain the concept of a martingale by a simple example. A player tosses a fair coin  $p = q = \frac{1}{2}$  and wins a dollar if heads come up and loses a dollar if tails come up. This is the modified Bernoulli process described by Eq. (20.4.11). Let  $W_n = Z_1 + Z_2 + \dots + Z_n$  be his fortune at the end of  $n$  tosses. The average fortune of the player at the  $(n+1)$ st toss is the same as his current fortune  $W_n$ , and this is not affected by how he arrived at his current fortune, that is,  $E[w_{n+1}|Z_n, Z_{n-1}, \dots, Z_1] = W_n$ , thus capturing the notion that the game is fair. In the same spirit we can think of the game being biased in favor of the player (submartingale) or biased against the player (supermartingale).

Unless otherwise stated, we will assume that the condition  $E|Y_n| < \infty$  or  $E|Y(t)| < \infty$  is always satisfied in the following examples.

**Example 20.10.1 Random-Walk Process** Let  $\{Z_i\}$  be the modified Bernoulli process described by Eq. (20.4.11) with  $\mu_Z = p - q$  and  $\sigma_Z^2 = 4pq$ . The corresponding random-walk process is a modified form of Eq. (20.4.15):

$$W_n = \sum_{i=1}^n Z_i = \sum_{i=1}^{n-1} Z_i + Z_n = W_{n-1} + Z_n \quad (20.10.7)$$

The mean value of  $W_n$  from Eq. (20.4.16) is  $\mu_W = n(p - q)$ , and the variance of  $W_n$  from Eq. (20.4.17) is  $\sigma_W^2 = 4npq$ . The conditional expectation  $E[W_{n+1}|Z_n, Z_{n-1}, \dots, Z_1]$  is given by

$$E[W_{n+1}|Z_n, Z_{n-1}, \dots, Z_1] = E[Z_{n+1} + W_n|Z_n, Z_{n-1}, \dots, Z] = (p - q) + W_n \quad (20.10.8)$$

If we now define  $Y_n = W_n - n(p - q)$  in Eq. (20.10.5), we can write

$$E[Y_{n+1}|Z_n, Z_{n-1}, \dots, Z_1] = E[Z_{n+1} - (p - q) + Y_n|Z_n, Z_{n-1}, \dots, Z] = Y_n \quad (20.10.9)$$

Hence  $Y_n$  is a discrete martingale with respect to the sequence  $\{Z_k, 1 \leq k \leq n\}$ .

**Example 20.10.2** The characteristic function of  $W_n$  in Eq. (20.10.4) is  $\Phi_{W_n}(\omega) = E[e^{j\omega W_n}]$ , and we define

$$Y_n = \frac{e^{j\omega W_n}}{\Phi_{W_n}(\omega)} \quad \text{and} \quad E[Y_n] = 1$$

For  $m \leq n$ , we can write for the following conditional expectation:

$$E[Y_n|W_m, \dots, W_1] = E\left[\frac{e^{j\omega(W_n - W_m)} e^{j\omega W_m}}{\Phi_{W_n}(\omega)} \middle| W_m, \dots, W_1\right]$$

Because of the independent increment property,  $(W_n - W_m)$  is independent of  $W_m$  and hence  $E[e^{j\omega(W_n - W_m)}|W_m, \dots, W_1] = E[e^{j\omega(W_n - W_m)}]$  and the preceding equation can be written as

$$E[Y_n|W_m, \dots, W_1] = \frac{E[e^{j\omega(W_n - W_m)}]}{\Phi_{W_{n-m}}(\omega)} \frac{e^{j\omega W_m}}{\Phi_{W_m}(\omega)} = \frac{e^{j\omega W_m}}{\Phi_{W_m}(\omega)} = Y_m, \quad m \leq n$$

Hence  $Y_n$  is a martingale with respect to the sequence  $\{W_m, \dots, W_1\}$ .

**Example 20.10.3** The sequence  $\{Z_i\}$  of Example 20.10.2 is now a set of i.i.d. random variables with  $p = q = \frac{1}{2}$  with variance  $\sigma_Z^2$ . Here the mean value  $\mu_Z = 0$ . We form the sequence

$$Y_n = \left( \sum_{i=1}^n Z_i \right)^2 - n\sigma_Z^2$$

Since  $E(\sum_{i=1}^n Z_i)^2 = n\sigma_Z^2$ , we have  $E[Y_n] = 0$ . For  $m \leq n$ , we can write

$$\begin{aligned} E[Y_n|Z_m, \dots, Z_1] &= E\left\{ \left[ \left( \sum_{i=m}^{n-m} Z_i \right)^2 - \sigma_Z^2(n-m) \right] \right. \\ &\quad \left. + \left[ \left( \sum_{i=1}^m Z_i \right)^2 - m\sigma_Z^2 \right] \middle| Z_m, \dots, Z_1 \right\} \end{aligned}$$

Since  $E(\sum_{i=m}^{n-m} Z_i)^2 - \sigma_Z^2(n-m)$  is independent of the sequence  $\{Z_m, \dots, Z_1\}$  we have

$$E \left[ \left( \sum_{i=m}^{n-m} Z_i \right)^2 - \sigma_Z^2(n-m) \middle| Z_m, \dots, Z_1 \right] = E \left[ \left( \sum_{i=m}^{n-m} Z_i \right)^2 - \sigma_Z^2(n-m) \right] = 0$$

and hence

$$E[Y_n | Z_m, \dots, Z_1] = E \left[ \left( \sum_{i=1}^m Z_i \right)^2 - m\sigma_Z^2 \middle| Z_m, \dots, Z_1 \right] = \left( \sum_{i=1}^m Z_i \right)^2 - m\sigma_Z^2 = Y_m$$

and  $Y_n$  is a discrete martingale with respect to the sequence  $\{Z_k, 1 \leq k \leq n\}$ .

**Example 20.10.4 (Poisson Martingale)** The Poisson process  $\{N(t), t \geq 0\}$  with mean  $\mu_N(t) = \lambda t$  is given by the probability function

$$P\{N(t) = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

Let  $Y(t) = N(t) - \lambda t$ . We will show that  $Y(t)$  is a martingale. For  $s \leq t$  we can write

$$E[Y(t) | Y(s), s \leq t] = E[Y(t) - Y(s) + y(s) | Y(s), s \leq t]$$

Since  $Y(t)$  is an independent increment process,  $Y(t) - Y(s)$  is independent of  $Y(s)$ . Hence

$$E[Y(t) - Y(s) | Y(s), s \leq t] = E[Y(t) - Y(s), s \leq t] = 0$$

Hence we have

$$E[Y(t) | Y(s)] = Y(s), \quad s \leq t$$

showing that  $Y(t)$  is a martingale with respect to the process  $\{Y(s), s \leq t\}$ .

**Example 20.10.5 (Wiener Martingale)** The Wiener process  $\{W(t), t \geq 0\}$  has been defined in Section 20.7 as a zero mean independent increment process with density function given by Eq. (20.7.1):

$$f_W(w, t) = \frac{1}{\sqrt{2\pi\sigma t}} e^{-(1/2)(w^2/\sigma^2 t)}$$

Using the same procedure as in previous examples, we can write

$$E[W(t) | W(s), s \leq t] = E[W(t) - W(s) + W(s) | W(s), s \leq t]$$

Since  $W(t)$  is an independent increment process,  $W(t) - W(s)$  is independent of  $W(s)$ . Hence

$$E[W(t) - W(s) | W(s), s \leq t] = E[W(t) - W(s), s \leq t] = 0$$

Hence we have

$$E[W(t) | W(s)] = W(s), \quad s \leq t$$

showing that  $W(t)$  is a martingale with respect to the process  $\{W(s), s \leq t\}$ .

In all the five examples given above the underlying random process was an independent increment process. The question arises as to whether all independent increment processes with zero mean are martingales. We will show that this is indeed the case. If

$X(t)$  is an independent increment process, then from Eq. (20.4.1), we have

$$\begin{aligned} E\{[X(t_i) - X(t_{i-1})][X(t_{i+1}) - X(t_i)]\} \\ = E[X(t_i) - X(t_{i-1})]E[X(t_{i+1}) - X(t_i)] \end{aligned} \quad (20.4.1)$$

If the mean of  $X(t)$  is  $\mu_X(t)$ , then we can define a zero mean independent increment process  $Y(t) = X(t) - \mu_X(t)$  and write the conditional expectation  $E[Y(t)|Y(s)]$  for  $0 \leq s \leq t$  as follows:

$$E[Y(t)|Y(s)] = E[Y(t) - Y(s)|Y(s)] + E[Y(s)|Y(s)], \quad 0 \leq s \leq t \quad (20.10.10)$$

Using the independent increment property, we obtain

$$E[Y(t) - Y(s)|Y(s)] = E[Y(t) - Y(s)] = 0, \quad 0 \leq s \leq t \quad (20.10.11)$$

Substituting Eq. (20.10.11) in Eq. (20.10.10), we obtain

$$E[Y(t) - Y(s)|Y(s)] = E[Y(t) - Y(s)] = 0, \quad 0 \leq s \leq t \quad (20.10.12)$$

thus showing that a zero mean independent increment process is a martingale with respect to the process  $\{Y(s), 0 \leq s \leq t\}$ . However, not all martingales are independent increment processes, as shown in the following example.

**Example 20.10.6 (Likelihood Ratio Martingale)** The likelihood ratio under the two hypotheses  $H_0$  and  $H_1$  in Eq. (18.6.17) is modified as

$$L_X(n) = \frac{f_{\mathbf{X}}(X_n, \dots, X_1 | H_1)}{f_{\mathbf{X}}(X_n, \dots, X_1 | H_0)}$$

We will show that under the hypothesis  $H_0$ , the likelihood ratio  $L_X(n)$  is a martingale with respect to the sequence  $\{X_n, \dots, X_1\}$ . The conditional expectation  $E[L_X(n+1)|X_n, \dots, X_1|H_0]$  can be written as

$$\begin{aligned} E[L_X(n+1)|X_n, \dots, X_1|H_0] \\ = \int_{-\infty}^{\infty} \frac{f_{\mathbf{X}}(x_{n+1}, x_n, \dots, x_1 | H_1)}{f_{\mathbf{X}}(x_{n+1}, x_n, \dots, x_1 | H_0)} f_{\mathbf{X}}(x_{n+1}, x_n, \dots, x_1 | H_0) dx_{n+1} \end{aligned}$$

From the definition of conditional probabilities, we can substitute the following in the equation above

$$f_{\mathbf{X}}(x_{n+1}, x_n, \dots, x_1 | H_0) = f_{\mathbf{X}}(x_{n+1} | x_n, \dots, x_1 | H_0) f_{\mathbf{X}}(x_n, \dots, x_1 | H_0)$$

and write

$$\begin{aligned} E[L_X(n+1)|X_n, \dots, X_1|H_0] \\ = \int_{-\infty}^{\infty} \frac{f_{\mathbf{X}}(x_{n+1}, x_n, \dots, x_1 | H_1)}{f_{\mathbf{X}}(x_n, \dots, x_1 | H_0)} \frac{f_{\mathbf{X}}(x_{n+1} | x_n, \dots, x_1 | H_0)}{f_{\mathbf{X}}(x_{n+1} | x_n, \dots, x_1 | H_0)} dx_{n+1} \\ = \frac{1}{f_{\mathbf{X}}(x_n, \dots, x_1 | H_0)} \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_{n+1}, x_n, \dots, x_1 | H_1) dx_{n+1} = L_X(n) \end{aligned}$$

thus showing that  $L_X(n)$  is a martingale with respect to the sequence  $\{X_n, \dots, X_1\}$  under the hypothesis  $H_0$ .

**Example 20.10.7** The Wiener process  $\{W(t), t \geq 0\}$  has mean 0 and variance  $\sigma^2 t$ . We form the function  $Y(t)$  with  $W(t)$  as

$$Y(t) = e^{kW(t) - k^2(\sigma^2 t/2)}, \quad t > 0$$

where  $k$  is an arbitrary constant. We will show that  $Y(t)$  is a martingale with respect to the Wiener process  $\{W(t), t \geq 0\}$ . We form the conditional expectation with  $s \leq t$ :

$$\begin{aligned} E[Y(t) | W(s), 0 \leq s < t] &= e^{-k^2(\sigma^2 t/2)} E[e^{kW(t)} | W(s)] \\ &= e^{-k^2(\sigma^2 t/2)} E[e^{k[W(t) - W(s)]} e^{W(s)} | W(s)] \end{aligned}$$

Using the independent increment property  $E[e^{k[W(t) - W(s)]} | W(s)] = E[e^{k[W(t) - W(s)]}]$  in the preceding equation, we find that

$$E[Y(t) | W(s), 0 \leq s < t] = e^{kW(s)} e^{-k^2(\sigma^2 t/2)} E[e^{k[W(t) - W(s)]}]$$

But  $E[e^{k[W(t) - W(s)]}]$  is the moment generating function (MGF) of the process  $W(t) - W(s)$ , which has zero mean and variance  $\sigma^2(t - s)$ . From Example 11.5.2 the MGF is given by  $E[e^{k[W(t) - W(s)]}] = e^{k^2[\sigma^2(t-s)]/2}$ , and substituting this equation, we obtain

$$E[Y(t) | W(s), 0 \leq s < t] = e^{kW(s)} e^{-k^2(\sigma^2 t/2)} e^{k^2[\sigma^2(t-s)]/2} = e^{kW(s)} e^{-k^2(\sigma^2 s/2)} = Y(s)$$

thus showing the required result.

**Example 20.10.8** The random process is given by  $Y(t) = W^2(t)$ , where  $W(t)$  is a Wiener process with variance  $\sigma^2 t$ . We will show that  $Y(t)$  is a submartingale.

We form the conditional expectation with  $s < t$ :

$$E[Y(t) | W(s)] = E[W^2(t) - W^2(s) - W^2(s) | W(s)], \quad 0 \leq s < t$$

With  $0 \leq s < t$  and, using the independent increment property, this equation can be written as

$$E[Y(t) | W(s)] = E[W^2(t) - W^2(s)] - W^2(s) = \sigma^2(t - s) - W^2(s) > Y(s)$$

showing that  $Y(t)$  is indeed a submartingale. However, we can show that  $V(t) = W^2(t) - \sigma^2 t$  is a martingale as shown below:

$$E[V(t) | W(s)] = E[W^2(t) - \sigma^2 t - W^2(s) - \sigma^2 s] + V(s) = V(s), \quad 0 \leq s < t$$

Thus,  $V(t)$  is a martingale with respect to the Wiener process  $\{W(t), t \geq 0\}$ .

**Example 20.10.9** If  $\{Y_n, n \geq 0\}$  is a martingale with respect to the sequence  $\{X_m, m \geq 0\}$  and  $\phi(\cdot)$  is some convex function, then  $\phi(Y_n)$  is submartingale with respect to  $\{X_m, m \geq 0\}$  as shown below:

From Jensen's inequality [Eq. (14.6.3),  $E[h(X)] \geq h(E[X])$ ], we can write

$$E[\phi(Y_n) | X_m] \geq \phi\{E[Y_n | X_m]\} = \phi(Y_m)$$

Showing that  $\phi(Y_n)$  is a submartingale.

As particular cases the following convex functions of a martingale  $Y_n$  are all submartingales: (1)  $|Y_n|$ , (2)  $Y_n^2$ , (3)  $|Y_n|^\alpha$ . Similarly, convex functions of a continuous martingale  $Y(t)$  are submartingales as shown below:

1. Poisson martingale  $N(t) - \lambda t$ .  $[N(t) - \lambda t]^2$  is a submartingale.
2. Wiener martingale  $W(t)$ . We have already seen in Example 20.10.8 that  $W^2(t)$  is a submartingale.
3.  $W^2(t) - \sigma^2 t$  is a martingale.  $[W^2(t) - \sigma^2 t]^2$  is a submartingale.

A martingale process is based on expected values and a Markov process on conditional probabilities, or

$$\begin{aligned} E[X(t) | X(k), 0 \leq k \leq s] &= X(s) && \text{martingale} \\ P[X(t) | X(k), 0 \leq k \leq s] &= P[X(t) | X(s)] && \text{Markov} \end{aligned}$$

Hence a martingale process need not be a Markov process or vice versa. However, all independent increment processes are both martingales and Markov processes.

### Decomposition of Submartingales

Doob [16] and Meyer [41] have shown that under certain regularity conditions any submartingale can be uniquely decomposed into a martingale and an increasing predictable process. Heuristically, a process  $A(t)$  or  $A_n$  is *predictable* if it can be completely determined from a knowledge of either  $\{X(s), 0 \leq s < t\}$  or  $\{X_k, 0 < k \leq n\}$ . All deterministic functions are predictable. The Doob–Meyer decomposition states that if  $Y_t$  is a submartingale with respect to an underlying random process  $\{X_t, t \geq 0\}$ , then the unique decomposition is given by

$$Y_t = M_t + A_t \quad (20.10.13)$$

where  $M_t$  is a martingale with respect to the same underlying random process  $\{X_t, t \geq 0\}$  and  $A_t$  is an increasing predictable process. It is the decomposition aspect of the submartingales that gave rise to an elegant proof [19] for the nonlinear filtering problem of which the Kalman linear filtering problem is a special case. Complete details can be found in Refs. 26, 35, and 37.

## 20.11 PERIODIC RANDOM PROCESS

The continuity of the sample paths of a periodic random processes is of concern to us for expanding them in Fourier series. Among the different types of continuity the mean-square continuity is of importance. A random process  $X(t)$  is *continuous in the mean-square sense* if the following condition is satisfied:

$$\lim_{s \rightarrow t} E|X(t) - X(s)|^2 \rightarrow 0 \quad (20.11.1)$$

The mean-square continuity of the random process is tied to the autocorrelation  $R_X(t,s)$  of the process. We will show that a random process  $X(t)$  is mean-square continuous if and only if  $R_X(t,s)$  is continuous at the diagonal point  $s = t = \tau$ :

$$E|X(t) - X(s)|^2 = R_X(t, t) - 2R_X(t, s) + R_X(s, s) \quad (20.11.2)$$

Since  $R_X(t,s)$  is continuous at  $s = t$ , by assumption the righthand side is zero at  $s = t$ , thus showing the necessity. Conversely

$$\begin{aligned} R_X(t, \tau) - R_X(s, s) &= E[X(t)X(\tau)] - E[X_s^2] \\ &= E\{[X(t) - X(s)]X(\tau)\} + E\{[X(\tau) - X(s)]X(s)\} \end{aligned}$$

From the preceding equation, using Schwarz's inequality of Eq. (14.5.13), we can form

$$\begin{aligned} R_X(t, \tau) - R_X(s, s) &\leq \sqrt{E[|X(t) - X(s)|^2]E[X^2(\tau)]} + \sqrt{E[|X(\tau) - X(s)|^2]E[X^2(s)]} \end{aligned} \quad (20.11.3)$$

From the assumption of mean-square continuity of  $X(t)$  the righthand side goes to zero as  $\tau, s \rightarrow t$  and hence  $R_X(t,s)$  is continuous at  $s = t$ .

A random process  $X(t)$  is *periodic in the wide sense* with period  $T_c$  if

$$\begin{aligned} \mu_X(t + kT_c) &= \mu_X(t) \quad \text{for all } t \text{ and integer } k \\ R_X(t, s) &= R_X(t + kT_c, s) = R_X(t, s + kT_c) \quad \text{for all } t, s, \text{ and integer } k \end{aligned} \quad (20.11.4)$$

The autocorrelation function  $R_X(t,s)$  is periodic in both arguments  $t$  and  $s$  and is shown in Fig. 20.11.1 in terms of constant-value contours.

If  $X(t)$  is a stationary process, then with  $\tau = t - s$ , Eq. (20.11.4) simplifies to

$$R_X(\tau) = R_X(\tau + kT_c) \quad \text{for all } \tau \text{ and integer } k \quad (20.11.5)$$

and the mean-square continuity of  $X(t)$  guarantees that  $R_X(\tau)$  is uniformly continuous in  $\tau$ . Such wide-sense *periodic stationary processes* are shown in Examples 19.2.1–19.2.3. As a consequence of the continuity, the Fourier series of  $R_X(\tau)$  given by

$$R_X(\tau) = \sum_{n=-\infty}^{\infty} R_n e^{j n \omega_c (\tau - t)}, \quad \tau = s - t \quad (20.11.6)$$

converges uniformly to  $R_X(\tau)$ .

As a consequence of the uniform continuity of  $R_X(\tau)$ , the zero mean stationary periodic random process  $X(t)$  with fundamental frequency  $\omega_c = 2\pi/T_c$  can be represented in

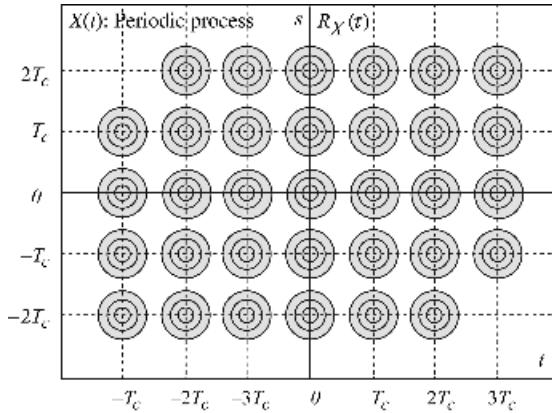


FIGURE 20.11.1

the mean-square sense by a Fourier series

$$X(t) = \sum_{n=-\infty}^{\infty} X_n e^{jn\omega_c t}, \quad X_0 = 0 \quad \text{and} \quad X_n = \frac{1}{T_c} \int_0^{T_c} X(t) e^{-jn\omega_c t} dt \quad (20.11.7)$$

The Fourier coefficients  $\{X_n\}$  are in general complex random variables, and we will investigate their properties. Since  $X(t)$  is a zero mean process,  $E[X_n] = 0$ . The cross-correlation between two of the coefficients can be calculated using Eq. (20.11.7):

$$\begin{aligned} E[X_m X_n^*] &= \frac{1}{T_c^2} \int_0^{T_c} \int_0^{T_c} E[X(s)X(t)] e^{-j\omega_c(ms-nt)} ds dt \\ &= \frac{1}{T_c^2} \int_0^{T_c} \int_0^{T_c} R_X(s-t) e^{-j\omega_c(ms-nt)} dt ds \end{aligned} \quad (20.11.8)$$

Substituting for  $R_X(s-t)$  from Eq. (20.11.6) in Eq. (20.11.8), we obtain

$$\begin{aligned} E[X_m X_n^*] &= \frac{1}{T_c^2} \int_0^{T_c} \int_0^{T_c} \sum_{k=-\infty}^{\infty} R_k e^{jk\omega_c(s-t)} e^{-j\omega_c(ms-nt)} dt ds \\ &= \frac{1}{T_c^2} \sum_{k=-\infty}^{\infty} R_k \int_0^{T_c} e^{j\omega s(k-m)} ds \int_0^{T_c} e^{j\omega t(n-k)} dt \end{aligned} \quad (20.11.9)$$

The complex exponentials  $\{e^{jn\omega_c s}\}$  and  $\{e^{jn\omega_c t}\}$  are orthogonal sets and hence

$$\begin{aligned} \frac{1}{T_c} \int_0^{T_c} e^{j\omega s(k-m)} ds &= \delta_{km} = \begin{cases} 1, & k = m \\ 0, & k \neq m \end{cases} : \frac{1}{T_c} \int_0^{T_c} e^{j\omega t(n-k)} dt \\ &= \delta_{nk} = \begin{cases} 1, & n = k \\ 0, & n \neq k \end{cases} \end{aligned} \quad (20.11.10)$$

Substituting Eq. (20.11.10) into Eq. (20.11.9), we obtain the result

$$E[X_m X_n^*] = \frac{1}{T_c^2} \sum_{k=-\infty}^{\infty} R_k \delta_{km} \delta_{nk} = R_m \delta_{mn} = \begin{cases} R_m, & n = m \\ 0, & n \neq m \end{cases}$$

or

$$E[X_m^2] = R_m \quad (20.11.11)$$

Thus, the Fourier coefficients of a zero mean periodic stationary random process  $X(t)$  are orthogonal and uncorrelated. If it is a Gaussian process, then these coefficients are also independent.

**Example 20.11.1** The autocorrelation of a zero mean white Gaussian noise  $W(t)$  is given by

$$E[W(s)W(t)] = R_W(t-s) = \sigma^2 \delta(t-s)$$

where  $\sigma^2$  is the variance parameter. The process may be considered periodic with period  $T_c \rightarrow \infty$ . Substituting for  $R_W(t - s)$  in Eq. (20.11.8), we obtain

$$E[W_m W_n^*] = \lim_{T_c \rightarrow \infty} \frac{\sigma^2}{T_c^2} \int_0^{T_c} \int_0^{T_c} \delta(s - t) e^{-j\omega_c(ms - nt)} dt ds = \lim_{T_c \rightarrow \infty} \frac{\sigma^2}{T_c^2} \delta_{mn}$$

or

$$E[W_m^2] = \lim_{T_c \rightarrow \infty} \frac{\sigma^2}{T_c^2} = \sigma_W^2$$

and the Fourier coefficients  $W_m$  and  $W_n$  are independent for any  $m$  and  $n$  provided  $m \neq n$ .

### Cyclostationary Process

A random process  $X(t)$  is called *strictly cyclostationary* [21] if the joint distribution function for any set of samples at times  $t_1, t_2, \dots, t_n$  is the same for any integer multiples of some period  $T$ , that is, if the cdf and pdf satisfy the equations

$$\begin{aligned} F_X(x_1, \dots, x_n; t_1, \dots, t_n) &= F_X(x_1, \dots, x_n; t_1 + kT, \dots, t_n + kT) \\ f_X(x_1, \dots, x_n; t_1, \dots, t_n) &= f_X(x_1, \dots, x_n; t_1 + kT, \dots, t_n + kT) \end{aligned} \quad (20.11.12)$$

for any  $n$  and  $k$ . This is somewhat restrictive, and hence we define a *wide-sense cyclostationary process* if the mean  $\mu_X(t)$  and autocorrelation  $R_X(t_1, t_2)$  are periodic:

$$\begin{aligned} \mu_X(t + kT) &= \mu_X(t) \\ R_X(t_1 + kT, t_2 + kT) &= R_X(t_1, t_2) \end{aligned} \quad (20.11.13)$$

The autocorrelation function of a wide-sense cyclostationary process is shown in Fig. 20.11.2. The difference between Figs. 20.11.2 and 20.11.1 should be noted.

**Example 20.11.2** A random process  $X(t)$  is defined by  $X(t) = A \cos(\omega_c t) + B \sin(\omega_c t)$ , where  $A$  and  $B$  are two zero mean independent Gaussian random variables with variances

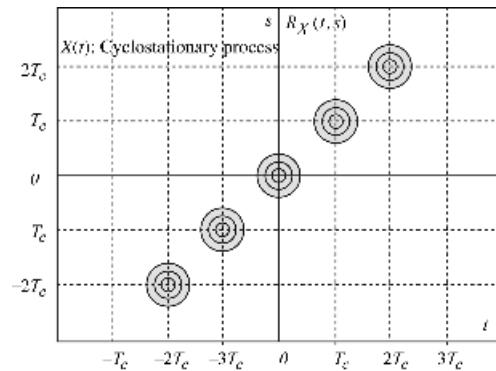


FIGURE 20.11.2

$\sigma_A^2$  and  $\sigma_B^2$  respectively. The mean value of this process is

$$E[X(t)] = \mu_X(t) = E[A]\cos(\omega_c t) + E[B]\sin(\omega_c t) = 0, \quad E[A] = E[B] = 0$$

The autocorrelation function  $R_X(t,s)$  is given by

$$\begin{aligned} R_X(t,s) &= E[X(t)X(s)] = E[[A \cos(\omega_c t) + B \sin(\omega_c t)][A \cos(\omega_c s) + B \sin(\omega_c s)]] \\ &= E[A^2]\cos(\omega_c t)\cos(\omega_c s) + E[B^2]\sin(\omega_c t)\sin(\omega_c s) \\ &\quad + E[AB][\cos(\omega_c t)\sin(\omega_c s) + \sin(\omega_c t)\cos(\omega_c s)] \end{aligned}$$

Since  $A$  and  $B$  are zero mean and independent,  $E[A^2] = \sigma_A^2$ ,  $E[B^2] = \sigma_B^2$ , and  $E[AB] = E[A]E[B] = 0$  and hence

$$\begin{aligned} R_X(t,s) &= \sigma_A^2 \frac{1}{2}[\cos(\omega_c(s-t)) + \cos(\omega_c(s+t))] \\ &\quad + \sigma_B^2 \frac{1}{2}[\cos(\omega_c(s-t)) - \cos(\omega_c(s+t))] \\ &= \frac{\sigma_A^2 + \sigma_B^2}{2} \cos(\omega_c(s-t)) + \frac{\sigma_A^2 - \sigma_B^2}{2} \cos(\omega_c(s+t)) \end{aligned}$$

Since the autocorrelation function  $R_X(s-t, s+t)$  is dependent on both  $t$  and  $s$ , the process is not stationary. However, with  $\mu_X(t) = 0$  and  $T_c = 2\pi/\omega_c$ , we have  $R_X(s-t, s+t) = R_X(s-t+kT_c, s+t+kT_c)$  in the preceding equation. Hence the process is wide-sense cyclostationary.

**Example 20.11.3** A pulse-amplitude-modulated random process is of the form

$$X(t) = \sum_{k=-\infty}^{\infty} A_k p(t - kT_c)$$

where  $\{A_k\}$  is a stationary digital sequence of random variables with mean  $\mu_A$  and autocorrelation function  $R_A(h) = E[A_k A_{k+h}]$ .  $p(t)$  is a deterministic pulse of duration  $T_c$ . We want to find the mean and the autocorrelation function of  $X(t)$ :

*Mean:*

$$\mu_X(t): \quad E[X(t)] = \mu_X(t) = \sum_{k=-\infty}^{\infty} E[A_k]p(t - kT_c) = \mu_A \sum_{k=-\infty}^{\infty} p(t - kT_c)$$

and the mean is a periodic function with period  $T_c$ .

*Autocorrelation:*

$$\begin{aligned} E[X(t)X(s)] &= R_X(t,s) = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E[A_k A_j]p(t - kT_c)p(s - jT_c) \\ &= \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} R_A(k, j)p(t - kT_c)p(s - jT_c) \\ &= R_X(t - kT_c, s - kT_c), k = \pm 1, \pm 2, \dots \end{aligned}$$

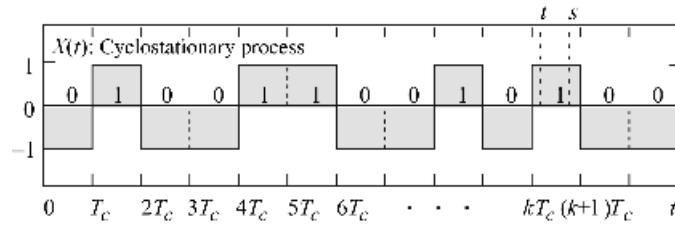


FIGURE 20.11.3

Since both conditions of Eq. (20.11.13) are satisfied, we conclude that  $X(t)$  is a cyclostationary process.

We will make the result more concrete by assuming that  $p(t)$  is a rectangular pulse of unit height and width  $T_c$ . The random variable sequence  $\{A_k\}$  is an independent set of equiprobable  $+1$  or  $-1$  representing the symbols  $1$  or  $0$  respectively. Since  $A_k$  is equiprobable,  $\mu_A = 0$ , and we obtain  $\mu_X(t) = 0$ . A sample waveform of  $X(t)$  corresponding to sending symbols  $\{0100110010100\}$  is shown in Fig. 20.11.3.

Since the  $A_k$  values are independent with zero mean, we have

$$E[X(t)X(s)] = R_X(t, s) = \sum_{k=-\infty}^{\infty} E[A_k^2]p(t - kT_c)p(s - kT_c)$$

With  $E[A_k^2] = 1.1 \cdot \frac{1}{2} + (-1) \cdot (-1) \cdot \frac{1}{2} = 1$ , we obtain

$$R_X(t, s) = \begin{cases} 1, & \begin{cases} kT_c \leq t < (k+1)T_c \\ kT_c \leq s < (k+1)T_c \end{cases} \\ 0, & \text{otherwise} \end{cases}$$

The function  $R_X(t, s)$  is shown in Fig. 20.11.4, which corresponds to Fig. 20.11.2, showing that  $X(t)$  is a cyclostationary process.

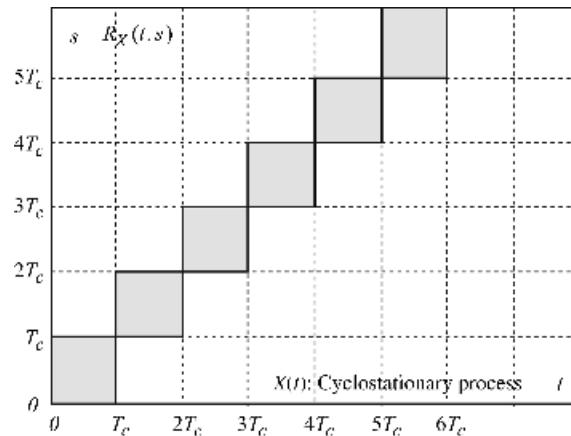


FIGURE 20.11.4

## 20.12 APERIODIC RANDOM PROCESS (KARHUNEN–LOEVE EXPANSION)

In the last section the coefficients of the Fourier expansion of a mean-square continuous periodic random process were found to be orthogonal random variables. We can attempt a similar Fourier expansion in the interval  $[a,b]$  for a mean-square continuous aperiodic random process  $X(t)$  as

$$X(t) = \sum_{n=1}^{\infty} X_n \phi_n(t), \quad a \leq t \leq b \quad (20.12.1)$$

where the sequence  $\{\phi_n(t)\}$  is an orthonormal set in the interval  $[a,b]$  defined by

$$\int_a^b \phi_n(t) \phi_m^*(t) dt = \begin{cases} 1, & m = n \\ 0, & m \neq n \end{cases} \quad (20.12.2)$$

and the generalized Fourier coefficients (FCs)  $X_n$  are given by

$$X_n = \int_a^b X(t) \phi_n^*(t) dt, \quad n = 0, 1, \dots \quad (20.12.3)$$

If the sequence  $\{\phi_n(t)\}$  is any orthonormal set, then the coefficients  $\{X_n\}$  will not be orthogonal, unlike the periodic random processes as shown in the following example.

**Example 20.12.1** Let  $X(t)$ ,  $0 \leq t \leq T$ , be a zero mean stationary Gaussian random process with continuous ACF  $R_X(\tau)$ . A Fourier series expansion of  $X(t)$  is given by

$$X(t) = \lim_{N \rightarrow \infty} \sum_{n=-N}^N X_n e^{jn\omega_0 t}, \quad \omega_0 = \frac{2\pi}{T}$$

The Fourier coefficient  $X_n$  is given by

$$X_n = \frac{1}{T} \int_{-T/2}^{T/2} X(t) e^{-j n \omega_0 t} dt \quad \text{with} \quad E[X_n] = 0, \quad E[X(t)] = 0$$

We will show that the random variables  $\{X_n\}$  are not orthogonal.  $E[X_n X_m^*]$  can be written as

$$E[X_n X_m^*] = \frac{1}{T^2} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} E[X(t) X(s)] e^{-j \omega_0 (nt - ms)} dt ds$$

Substituting  $m = n - k$  in this equation, we obtain

$$E[X_n X_m^*] = \frac{1}{T^2} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} R_X(t - s) e^{-j \omega_0 n(t-s)} e^{-j \omega_0 k s} dt ds$$

Substituting  $\tau = t - s$  in this equation, yields

$$E[X_n X_m^*] = \frac{1}{T^2} \int_{-T/2}^{T/2} \left[ \int_{-T/2-s}^{T/2-s} R_X(\tau) e^{-j \omega_0 n \tau} d\tau \right] e^{-j \omega_0 k s} ds \quad (A)$$

We will define the quantity in Eq. (A):

$$\int_{-T/2-s}^{T/2-s} R_X(\tau) e^{-j\omega_0 n \tau} d\tau = S_X(n, s) \quad \text{and} \quad \lim_{T \rightarrow \infty} S_X(n, s) = S_X(n\omega_0)$$

Substituting the preceding equation in Eq. (A), we find that

$$E[X_n X_m^*] = \frac{1}{T^2} \int_{-T/2}^{T/2} S_X(n, s) e^{-j\omega_0 ks} ds \neq 0$$

Hence we have shown that if  $X(t)$  is expanded in any orthonormal set, the resulting Fourier coefficients  $\{X_n\}$  are not orthogonal. However as  $T \rightarrow \infty$ , we see that  $S_X(n, s)$  tends to  $S_X(n\omega_0)$  and hence

$$\lim_{T \rightarrow \infty} \frac{1}{T^2} \int_{-T/2}^{T/2} S_X(n\omega_0) e^{j\omega_0 ks} ds \rightarrow 0$$

Thus, the Fourier coefficients are orthogonal as  $T \rightarrow \infty$ .

We are interested in finding the necessary and sufficient conditions for that orthonormal set that will yield an orthogonal Fourier coefficient set. We first multiply Eq. (20.12.1) by  $X_m^*$  and take expectations

$$E[X(t) X_m^*] = \sum_{n=1}^{\infty} E[X_n X_m^*] \phi_n(t), \quad a \leq t \leq b \quad (20.12.4)$$

Under the orthogonality condition of the FCs we can substitute  $E[X_n X_m^*] = 0, n \neq m$  in Eq. (20.12.4) and obtain

$$E[X(t) X_m^*] = E|X_m|^2 \phi_m(t), \quad a \leq t \leq b \quad (20.12.5)$$

We can obtain a second equation by multiplying the following equation for  $X_m^*$

$$X_m^* = \int_a^b X^*(s) \phi_m(s) ds$$

by  $X(t)$  and taking expectations. Or, for  $a \leq t \leq b$

$$E[X(t) X_m^*] = \int_a^b E[X(t) X^*(s)] \phi_m(s) ds = \int_a^b R_X(t, s) \phi_m(s) ds \quad (20.12.6)$$

Comparing Eqs. (20.12.5) and (20.12.6), we can write

$$E[X(t) X_m^*] = \int_a^b R_X(t, s) \phi_m(s) ds = E|X_m|^2 \phi_m(t), \quad a \leq t \leq b \quad (20.12.7)$$

Defining the signal energy  $E|X_n|^2 = \lambda_n$ , Eq. (20.12.7) can be rewritten as

$$\int_a^b R_X(t, s) \phi_n(s) ds = \lambda_n \phi_n(t), \quad a \leq t \leq b$$

or

$$\int_a^b R_X(t, s) \phi(s) ds = \lambda \phi(t), \quad a \leq t \leq b \quad (20.12.8)$$

This equation which is called the *Karhunen–Loeve* (K-L) *integral equation*, which gives the necessary and sufficient conditions for the random process  $X(t)$  to be expanded as a series in terms of the orthonormal set  $\{\phi_m(t)\}$  such that the generalized Fourier coefficients  $X_m$ , called *Karhunen–Loeve* (K-L) *coefficients*, are orthogonal. The expansion of Eq. (20.12.1) is called the *Karhunen–Loeve* (K-L) *expansion*. The orthonormal functions  $\{\phi_m(t)\}$  are called the *eigenfunctions* and  $\{\lambda_m\}$  are called the *eigenvalues* of the K-L integral equation. It is to be particularly noted that the process  $X(t)$  can be nonstationary.

*Summary:*

K-L equation satisfied  $\Rightarrow$  Fourier coefficients  $\{X_n\}$  are orthogonal

Fourier coefficients  $\{X_n\}$  are orthogonal  $\Rightarrow$  K-L equation is satisfied

**Example 20.12.2** We are given a white-noise process with autocorrelation function  $R_V(t, s) = \sigma^2 \delta(t - s)$ . Even though this process violates the second-order condition in the sense that the autocorrelation function is not finite, we can formally find the orthonormal set such that the Fourier coefficients are orthogonal in the interval  $[-a, a]$ . Substituting for  $R_X(t, s)$  in Eq. (20.12.8), we obtain

$$\int_a^b \sigma^2 \delta(t - s) \phi_n(s) ds = \lambda_n \phi_n(t)$$

or

$$\sigma^2 \phi_n(t) = \lambda_n \phi_n(t) \quad \text{and} \quad \sigma^2 = \lambda_n, \quad -a \leq t \leq a$$

This result conveys that *any* orthonormal set  $\{\phi_n(t)\}$  with eigenvalues  $\sigma^2$  will generate K-L coefficients  $V_n$  for the white-noise process in the interval  $[-a, a]$  with  $E[V_n^2] = \lambda_n = \sigma^2$ .

**Example 20.12.3** A random process  $Y(t)$  not necessarily stationary is given by

$$Y(t) = X(t) + V(t)$$

where  $V(t)$  is a zero mean white Gaussian noise uncorrelated with  $X(t)$ . We want to find the eigenfunctions  $\{\phi_n(t)\}$  in the interval  $[-a, a]$  that will give K-L coefficients. The autocorrelation function of  $Y(t)$  is

$$R_Y(t, s) = R_X(t, s) + \sigma^2 \delta(t - s)$$

The eigenfunctions  $\{\phi_X(t)\}$  and  $\{\phi_Y(t)\}$  for representing  $X(t)$  and  $Y(t)$  can be found by applying the K-L integral equation Eq. (20.12.8) to  $R_X(t, s)$  and  $R_Y(t, s)$  as given below:

$$\int_{-\alpha}^{\alpha} R_X(t, s) \phi_X(s) ds = \lambda_X \phi_X(t), \quad a < t \leq a$$

$$\int_{-\alpha}^{\alpha} R_Y(t, s) \phi_Y(s) ds = \lambda_Y \phi_Y(t), \quad -a < t \leq a$$

Since  $X(t)$  and  $V(t)$  are uncorrelated, we can substitute  $R_Y(t, s) = R_X(t, s) + \sigma^2\delta(t - s)$  in the preceding integral equation for  $R_Y(t, s)$  and obtain

$$\int_{-\alpha}^{\alpha} [R_X(t, s) + \sigma^2\delta(t - s)]\phi_Y(s)ds = \lambda_Y\phi_Y(t), \quad -\alpha < t \leq \alpha$$

or

$$\int_{-\alpha}^{\alpha} R_X(t, s)\phi_Y(s)ds = [\lambda_Y - \sigma^2]\phi_Y(t), \quad -\alpha < t \leq \alpha$$

From the uniqueness of the K-L equation,  $\phi_X(t) = \phi_Y(t)$  and  $\lambda_Y = \lambda_X + \sigma^2$ . Hence, we can use the same eigenfunction set  $\{\phi_X(t)\}$  for the series expansion of both  $X(t)$  and  $Y(t)$ , and from Example 20.12.1 the same can be used for expanding  $V(t)$ . Thus, all three signals  $X(t)$ ,  $Y(t)$ , and  $V(t)$  can be represented for  $-\alpha < t \leq \alpha$  in the same function space as

$$X(t) = \sum_{n=1}^{\infty} X_n\phi_n(t), \quad Y(t) = \sum_{n=1}^{\infty} Y_n\phi_n(t), \quad V(t) = \sum_{n=1}^{\infty} V_n\phi_n(t)$$

The previous two examples are simple ones. However, solving the general K-L integral equation can be difficult. In particular cases, closed-form solutions can be obtained as shown in the following examples.

**Example 20.12.4 (Nonstationary Process)** The autocorrelation function  $R_W(t, s)$  of the Wiener process  $W(t)$  is

$$R_W(t, s) = \sigma^2 \min(t, s) \quad (20.12.9)$$

The series expansion for  $W(t)$  for the time of observation  $[0, T]$  can be obtained from the K-L integral equation [Eq. (20.12.8)] as follows:

$$\sigma^2 \int_0^T \min(t, s)\phi(s)ds = \lambda\phi(t)$$

or

$$\sigma^2 \left[ \int_0^t s\phi(s)ds + t \int_t^T \phi(s)ds \right] = \lambda\phi(t), \quad 0 < t, s \leq T \quad (20.12.10)$$

The usual procedure to solve integral equations of this type is to differentiate them as many times as possible so that they are converted into differential equations. The resulting differential equations are then solved from the boundary conditions obtained from their form. Differentiating Eq. (20.12.10) with respect to  $t$ , there results

$$\sigma^2 \left[ t\phi(t) + \int_t^T \phi(s)ds - t\phi(t) \right] = \lambda\phi(t)$$

or

$$\sigma^2 \int_t^T \phi(s)ds = \lambda\phi(t), \quad 0 < t, s \leq T \quad (20.12.11)$$

Differentiating again with respect to  $t$  [Eq. (20.12.11)], we obtain a second-order homogeneous differential equation given by

$$\phi(t) + \frac{\sigma^2}{\lambda} = 0, \quad 0 < t \leq T, \quad \lambda \neq 0 \quad (20.12.12)$$

The characteristic equation is  $s^2 + (\sigma^2/\lambda) = 0$ , and the roots are  $s_1, s_2 = \pm(\sigma/\sqrt{\lambda})$ . The solution to the differential equation is

$$\phi(t) = A \cos\left(\frac{\sigma t}{\sqrt{\lambda}}\right) + B \sin\left(\frac{\sigma t}{\sqrt{\lambda}}\right) \quad (20.12.13)$$

There are three unknowns— $A$ ,  $B$ , and  $\lambda$ —to be resolved. The boundary conditions are obtained as follows. Substituting  $t = 0+$  in Eq. (20.12.10), we can write  $\sigma^2 \int_0^{0+} \min(t, s)\phi(s)ds = \lambda\phi(0+)$ , or  $\lambda\phi(0+) = 0$  and  $\phi(0+) = 0$ . Substituting this initial condition in Eq. (20.12.13), we have  $A = 0$  and the solution becomes

$$\phi(t) = B \sin\left(\frac{\sigma t}{\sqrt{\lambda}}\right) \quad (20.12.14)$$

Substituting  $t = T-$  in Eq. (20.12.11), we obtain  $\sigma^2 \int_{T-}^T \phi(s)ds = \lambda\phi(T-)$ , or  $\phi(T-) = 0$ . Differentiating Eq. (20.12.14),  $\phi(t) = B \cos(\sigma t/\sqrt{\lambda})\sigma/\sqrt{\lambda}$  and we have

$$B \cos\left(\frac{\sigma T}{\sqrt{\lambda}}\right) = 0 \quad \text{from which} \quad \lambda_n = \left[ \frac{\sigma T}{(2n-1)\pi/2} \right]^2, n \geq 1 \quad (20.12.15)$$

The constant  $B$  is obtained from the normalization integral given by

$$\int_0^T B^2 \sin^2(kt)dt = 1 \quad (20.12.16)$$

where, for convenience, we have defined

$$k = \frac{\sigma}{\sqrt{\lambda_n}} = \frac{\sigma(2n-1)\pi}{2\sigma T} = \frac{(2n-1)\pi}{2T}$$

The integral in Eq. (20.12.16) yields

$$\int_0^T B^2 \sin^2(kt)dt = \int_0^T \frac{B^2}{2} [1 - \cos(2kt)]dt = \frac{B^2 T}{2} - \frac{\sin(2kT)}{2k} = 1 \quad (20.12.17)$$

With  $k = (2n-1)\pi/2T$ , the term  $\sin(2kT)/2k$  in Eq. (20.12.17) goes to zero and  $B = \sqrt{2/T}$ .

Hence the eigenfunctions  $\phi_n(t)$  are

$$\phi_n(t) = \sqrt{\frac{2}{T}} \sin\left[\frac{(2n-1)\pi t}{2T}\right], \quad n \geq 1 \quad 0 < t \leq T \quad (20.12.18)$$

The K-L expansion of the Wiener process  $W(t)$  is given by

$$W(t) = \sum_{n=1}^{\infty} W_n \sqrt{\frac{2}{T}} \sin\left[\frac{(2n-1)\pi t}{2T}\right], \quad 0 < t \leq T \quad (20.12.19)$$

where the coefficients  $W_n$  are given by an equation similar to Eq. (20.12.3)

$$W_n = \int_0^T W(t) \sqrt{\frac{2}{T}} \sin\left[\frac{(2n-1)\pi t}{2T}\right] dt, \quad n \geq 1 \quad 0 < t \leq T \quad (20.12.20)$$

with the property  $E[W_n^2] = \lambda_n, n \geq 1$ .

**Example 20.12.5** A sinusoid with random phase is given by  $X(t) = A \cos(\omega t + \Phi)$ , where  $A$  and  $\omega$  are constants and  $\Phi$  is uniformly distributed in  $(-\pi, \pi)$ . The

autocorrelation function as derived in Example 19.2.1 is  $R_X(t, s) = (A^2/2) \cos[\omega(t - s)]$ . We have to find the K-L expansion for  $X(t)$  in the period  $[-T/2, T/2]$  and show that the K-L coefficients  $X_n$  are orthogonal.

Substituting for the autocorrelation function in the K-L equation [Eq. (20.12.8)], we obtain

$$\frac{A^2}{2} \int_{-T/2}^{T/2} \cos[\omega(t - s)]\phi(s)ds = \lambda\phi(t), \quad -\frac{T}{2} \leq t, s \leq \frac{T}{2}$$

Taking derivative with respect to  $t$  once, we get

$$\frac{-A^2}{2}\omega \int_{-T/2}^{T/2} \sin[\omega(t - s)]\phi(s)ds = \lambda\phi'(t), \quad -\frac{T}{2} \leq t, s \leq \frac{T}{2}$$

Taking the derivative with respect to  $t$  again, we have

$$\frac{A^2}{2}\omega^2 \int_{-T/2}^{T/2} \cos[\omega(t - s)]\phi(s)ds = \lambda\phi''(t)$$

or

$$-\omega^2\lambda\phi(t) = \lambda\phi(t) \quad \text{or} \quad \phi(t) + \omega^2\phi(t) = 0, \quad -\frac{T}{2} \leq t, s \leq \frac{T}{2}$$

The solution to the second-order homogeneous differential equation above is given by

$$\phi(t) = C \cos(\omega t) + D \sin(\omega t)$$

From the form of  $\phi(t)$  we conclude that the eigenfunctions are

$$\phi_1(t) = \sqrt{\frac{2}{T}} \cos(\omega t) \quad \phi_2(t) = \sqrt{\frac{2}{T}} \sin(\omega t)$$

There are only two terms in the finite K-L expansion given by

$$X(t) = X_1 \sqrt{\frac{2}{T}} \cos(\omega t) + X_2 \sqrt{\frac{2}{T}} \sin(\omega t)$$

We can find these two K-L coefficients by equating like coefficients in the expansion of  $X(t)$  given by

$$X(t) = A \cos(\omega t) \cos(\Phi) - A \sin(\omega t) \sin(\Phi)$$

Thus

$$X_1 = A \sqrt{\frac{T}{2}} \cos(\Phi) \quad \text{and} \quad X_2 = A \sqrt{\frac{T}{2}} \sin(\Phi)$$

With  $\Phi$  uniformly distributed in  $(-\pi, \pi)$ , the joint moment  $E[X_1 X_2]$  is

$$E[X_1 X_2] = \frac{A^2 T}{2} E[\cos(\Phi) \sin(\Phi)] = \frac{A^2 T}{2} E[\sin(2\Phi)] = 0$$

showing that the K-L coefficients are orthogonal.

We will now find the eigenvalue  $\lambda_1$  by substituting  $\phi_1(t)$  in the K-L equation:

$$\frac{A^2}{2} \int_{-T/2}^{T/2} \cos[\omega(t - s)] \sqrt{\frac{2}{T}} \cos(\omega s) ds = \sqrt{\frac{2}{T}} \cos(\omega t)$$

or

$$\frac{A^2}{2} \int_{-T/2}^{T/2} \frac{1}{2} \{ \cos(\omega t) + \cos[\omega(t-2s)] \} ds = \lambda_1 \cos(\omega t)$$

or

$$\frac{A^2 T}{4} \cos(\omega t) = \lambda \cos(\omega t) \quad \text{and} \quad \lambda_1 = \frac{A^2 T}{4}$$

In a similar manner, substituting  $\phi_2(t)$  in the K-L equation,  $\lambda_2 = A^2 T / 4$ . The energy values in the K-L coefficients are

$$E[X_1^2] = \frac{A^2 T}{2} E[\cos^2(\Phi)] = \frac{A^2 T}{4} = \lambda_1$$

$$E[X_2^2] = \frac{A^2 T}{2} E[\sin^2(\Phi)] = \frac{A^2 T}{4} = \lambda_2$$

**Example 20.12.6** The autocorrelation function of a random process  $X(t)$  is  $R_X(t, s) = e^{-2|t-s|}$ . The signal is observed in the interval  $[-2, 2]$ . We want to find the eigenvalues and eigenfunctions.

The function  $R_X(t, s) = e^{-2|t-s|}$  is shown in Fig. 20.12.1.

This example is a little more involved than the previous ones, and since  $X(t)$  is not specified, we cannot obtain the K-L coefficients directly.

The K-L equation is obtained by substituting  $R_X(t, s)$  in Eq. (20.12.8). The result is

$$\int_{-2}^t e^{-2(t-s)} \phi(s) ds + \int_t^2 e^{2(t-s)} \phi(s) ds = \lambda \phi(t), \quad -2 \leq t \leq 2$$

Differentiating this equation with respect to  $t$ , we obtain

$$-2 \int_{-2}^t e^{-2(t-s)} \phi(s) ds + 2 \int_t^2 e^{2(t-s)} \phi(s) ds = \lambda \phi(t)$$

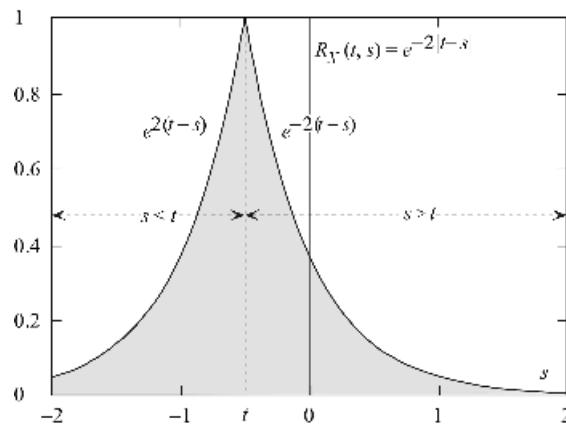


FIGURE 20.12.1

Differentiating again yields

$$-4\phi(t) + 4 \left[ \int_{-2}^t e^{-2(t-s)} \phi(s) ds + \int_t^2 e^{2(t-s)} \phi(s) ds \right] = \lambda \phi(t)$$

The quantity within braces is the K-L equation and equals  $\lambda \phi(t)$ . Hence we obtain

$$\lambda \phi(t) + 4(1 - \lambda) \phi(t) = 0, \quad -2 \leq t \leq 2$$

a second-order homogeneous differential equation. The characteristic equation is  $s^2 + 4(1 - \lambda)/\lambda = 0$ , and the oscillatory roots are given by  $\omega^2 = 4(1 - \lambda)/\lambda$ . We have an implicit relationship between  $\omega$  and  $\lambda$  with  $\lambda = 4/(4 + \omega^2)$ .

The two eigenfunctions are of the form  $\phi_1(t) = A \cos(\omega t)$  and  $\phi_2(t) = B \sin(\omega t)$ , where the constants  $A$  and  $B$  are obtained from normalization.

Substituting  $\phi_1(t)$  in the K-L equation, we obtain

$$\begin{aligned} e^{-2t} \int_{-2}^t e^{2s} \cos(\omega s) ds + e^{2t} \int_t^2 e^{-2s} \cos(\omega s) ds &= \lambda \cos(\omega t) \quad (\text{B}) \\ e^{-2t} \int_{-2}^t e^{2s} \cos(\omega s) ds &= \frac{1}{4 + \omega^2} \{2 \cos(\omega t) + \omega \sin(\omega t) - e^{-4} e^{-2t} [2 \cos(2\omega) + \omega \sin(2\omega)]\} \\ e^{2t} \int_t^2 e^{-2s} \cos(\omega s) ds &= \frac{1}{4 + \omega^2} \{2 \cos(\omega t) - \omega \sin(\omega t) - e^{-4} e^{2t} [-2 \cos(2\omega) + \omega \sin(2\omega)]\} \end{aligned}$$

Substituting these equations in Eq. (B) we obtain

$$\frac{1}{4 + \omega^2} \{4 \cos(\omega t) + e^{-4} (e^{2t} + e^{-2t}) [\omega \sin(2\omega) - 2 \cos(2\omega)]\} = \lambda \cos(\omega t)$$

Equating like coefficients on both sides of this equation, yields

$$\lambda = \frac{4}{4 + \omega^2} \quad \text{and} \quad \omega \sin(2\omega) - 2 \cos(2\omega) = 0$$

The first equation has been obtained earlier. The second equation can be simplified as  $\tan(2\omega) = 2/\omega$ . This is a transcendental equation that has to be solved graphically for  $\omega$ . It is simpler to solve graphically  $g(\omega) = (\omega/2) \tan(2\omega) = 1$  for  $\omega$  as shown in Fig. 20.12.2.

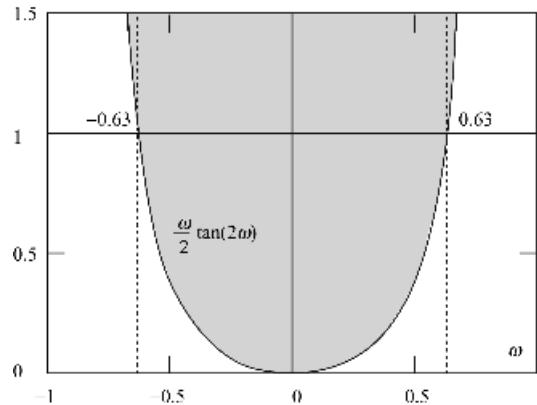


FIGURE 20.12.2

The solution to  $g(\omega) = 1$  is  $\omega = 0.6323$ . Having found  $\omega$ , we can find  $A$  and  $B$  from

$$\int_{-2}^2 A^2 \cos^2(\omega t) dt = 1, \quad \text{or} \quad A^2 \left( 2 + \frac{\sin(4\omega)}{2\omega} \right) = 1, \quad \text{hence } A = 0.6383$$

$$\int_{-2}^2 B^2 \sin^2(\omega t) dt = 1, \quad \text{or} \quad B^2 \left( 2 - \frac{\sin(4\omega)}{2\omega} \right) = 1, \quad \text{hence } B = 0.8044$$

The eigenvalue is  $\lambda = 4/(4 + \omega^2) = 0.9091$ . The eigenfunctions are  $\phi_1(t) = 0.6383 \cos(0.6323t)$  and  $\phi_2(t) = 0.8044 \sin(0.6323t)$ . The K-L expansion is  $X(t) = X_1\phi_1(t) + X_2\phi_2(t)$ ,  $-2 \leq t \leq 2$  and  $E[X_1^2] = E[X_2^2] = \lambda = 0.9091$ .

### Discrete Karhunen–Loeve Transform

In the last section we discussed the continuous Karhunen–Loeve expansion. The same type of an expansion is applicable to a discrete aperiodic random sequence  $\{X(n)\}$ . We shall assume that  $\{X(n)\}$  is a second-order random sequence with autocorrelation  $R_X(n, m)$ . Similar to the continuous-time case, the Fourier expansion for  $X(n)$  can be given by

$$X(n) = \sum_{i=0}^{N-1} X_i \phi_i(n), \quad n = 0, \dots, N-1 \quad (20.12.21)$$

where  $X_i$  is the discrete generalized Fourier transform and  $\{\phi_i(n)\}$  is an orthonormal set in the interval  $0 \leq n \leq N-1$ . We want to find the discrete equivalent of the integral equation (20.12.8) satisfied by the orthonormal sequence  $\{\phi_i(n)\}$ .

The generalized Fourier coefficients  $X_i$  are given by

$$X_i = \sum_{n=0}^{N-1} X(n) \phi_i^*(n), \quad 0 \leq i \leq N-1 \quad (20.12.22)$$

If the sequence  $\{\phi_i(n)\}$  is any orthonormal set, then the coefficients  $\{X_i\}$  in general will not be orthogonal. We are interested in finding the necessary and sufficient conditions for that orthonormal set  $\{\phi_i(n)\}$  that will yield an orthogonal set of Fourier coefficient  $\{X_i\}$ . Multiplying Eq. (20.12.21) by  $X_k^*$  and taking expectations, we can write

$$E[X(n)X_k^*] = \sum_{i=0}^{N-1} E[X_i X_k^*] \phi_i(n), \quad 0 \leq n \leq N-1 \quad (20.12.23)$$

Substituting the desired orthogonality criterion  $E[X_i X_k^*] = 0, i \neq k$  of the Fourier coefficients in Eq. (20.12.23) yields

$$E[X(n)X_k^*] = E|X_k|^2 \phi_k(n), \quad 0 \leq n, k \leq N-1 \quad (20.12.24)$$

We can obtain a second equation by multiplying the equation for  $X_k^*$

$$X_k^* = \sum_{m=0}^{N-1} X^*(m) \phi_k(m), \quad 0 \leq k \leq N-1$$

by  $X(n)$  and taking expectations. Thus, for  $0 \leq n, k \leq N-1$ , we obtain

$$E[X(n)X_k^*] = \sum_{m=0}^{N-1} E[X(n)X^*(m)] \phi_k(m) = \sum_{m=0}^{N-1} R_X(n, m) \phi_k(m) \quad (20.12.25)$$

Comparing Eqs. (20.12.24) and (20.12.25), we can write

$$E[X(n)X_k^*] = \sum_{m=0}^{N-1} R_X(n, m)\phi_k(m) = E|X_k|^2\phi_k(n), \quad 0 \leq n \leq N-1 \quad (20.12.26)$$

As in the continuous case, we can define the signal energy  $E|X_k|^2 = \lambda_k$  and rewrite Eq. (20.12.26) as

$$\sum_{m=0}^{N-1} R_X(n, m)\phi_k(m) = \lambda_k\phi_k(n), \quad 0 \leq n \leq N-1 \quad (20.12.27a)$$

or

$$\sum_{m=0}^{N-1} R_X(n, m)\phi(m) = \lambda\phi(n), \quad 0 \leq n \leq N-1 \quad (20.12.27b)$$

Equation (20.12.27) is a necessary and sufficient condition for the existence of an orthonormal set that will give an orthogonal set of coefficients  $\{X_k\}$  in the Fourier series expansion for random sequence  $X(n)$ . The convergence of Eq. (20.12.21) is in the mean-square sense, or

$$E\left|X(n) - \sum_{k=0}^{N-1} X_k\phi_k(n)\right|^2 = 0, \quad n = 0, \dots, N-1 \quad (20.12.28)$$

### Matrix Formulation of K-L Equation

The K-L equation [Eq. (20.12.27a)] can be compactly expressed in matrix form. The complex random sequence  $\{X(n)\}$  can be expressed as an  $N$ -vector  $\mathbf{X}$  as

$$\mathbf{X} = [X(0) \quad X(1) \quad \dots \quad X(N-1)]^T \quad (20.12.29)$$

with the autocorrelation matrix  $\mathbf{R}_X = E[\mathbf{X}\mathbf{X}^H]$  given by

$$\mathbf{R}_X = \begin{bmatrix} R_X(0, 0) & R_X(0, 1) & \dots & R_X(0, N-1) \\ R_X(1, 0) & R_X(1, 1) & \dots & R_X(1, N-1) \\ \vdots & \vdots & & \vdots \\ R_X(N-1, 0) & R_X(N-1, 1) & \dots & R_X(N-1, N-1) \end{bmatrix} \quad (20.12.30)$$

where the superscript  $H$  stands for conjugate transpose.

We have to find the condition under which  $\Phi$  an  $N \times N$  transformation matrix given by

$$\Phi = [\Phi_0 \quad \Phi_1 \dots \Phi_{N-1}] = \begin{bmatrix} \phi_{00} & \phi_{01} & \dots & \phi_{0N-1} \\ \phi_{10} & \phi_{11} & \dots & \phi_{1N-1} \\ \vdots & \vdots & & \vdots \\ \phi_{N-10} & \phi_{N-11} & \dots & \Phi_{N-1N-1} \end{bmatrix} \quad (20.12.31)$$

transforms the random sequence vector  $\mathbf{X}$  in Eq. (12.20.29) into an orthogonal random sequence vector  $\mathcal{X}$  given by

$$\mathcal{X} = \Phi^H \mathbf{X} = [X_0 \quad X_1 \quad \cdots \quad X_{N-1}]^T \quad (20.12.32)$$

with  $E[\mathcal{X}\mathcal{X}^H] = \Lambda$ , a diagonal matrix with components  $\{\lambda_k\}$ . In this case the transformation matrix  $\Phi$  has to satisfy the matrix K-L equation given by

$$\mathbf{R}_X \Phi = \Lambda \Phi \quad (20.12.33)$$

This equation is similar to the matrix eigenequation (16.2.5). Hence the transforming matrix  $\Phi$  is the modal matrix of  $\mathbf{R}_X$ , and since  $\mathbf{R}_X$  is symmetric and positive definite,  $\Phi$  is unitary, that is,  $\Phi^{-1} = \Phi^H$ . The eigenvalues of  $\mathbf{R}_X$  constitute the diagonal matrix  $\Lambda$  with elements  $\lambda_k = E[X_k^2]$ ,  $n = 0, 1, \dots, N - 1$ .

# Random Processes and Linear Systems

## 21.1 REVIEW OF LINEAR SYSTEMS

Consider the deterministic linear time-invariant system shown in Fig. 21.1.1 with input  $x(t)$  and output  $y(t)$ . The system function is characterized in the time domain as the output to a Dirac delta function  $\delta(t)$  with  $x(0) = 0$  and is called the *impulse response*  $h(t)$ .

A *causal* system is one that does not anticipate the input meaning that the impulse response  $h(t) = 0$ ,  $t < 0$ . The output  $y(t)$  can be given as a convolution of the input  $x(t)$  and the impulse response  $h(t)$ . Depending on the causality of  $x(t)$  and  $h(t)$ , the solution  $y(t)$  in continuous time is given by a convolution integral and may contain an initial condition term  $x(0)$  as shown below:

$$y(t) = \begin{cases} x(0)h(t) + \int_0^t x(\tau)h(t - \tau)d\tau = x(0)h(t) + \int_0^t x(t - \tau)h(\tau)d\tau, & x(t), h(t) \text{ causal} \\ \int_{-\infty}^t x(\tau)h(t - \tau)d\tau = \int_0^\infty x(t - \tau)h(\tau)d\tau, & x(t) \text{ noncausal, } h(t) \text{ causal} \\ \int_{-\infty}^\infty x(\tau)h(t - \tau)d\tau = \int_{-\infty}^\infty x(t - \tau)h(\tau)d\tau, & x(t), h(t) \text{ noncausal} \end{cases} \quad (21.1.1)$$

and similar equations in discrete time for the solution  $y(n)$  with initial condition  $x(0)$

$$y(n) = \begin{cases} x(0)h(n) + \sum_{m=1}^n x(m)h(n - m) \\ = \sum_{m=0}^n x(n - m)h(m), & x(n), h(n) \text{ causal} \\ \sum_{m=-\infty}^n x(m)h(n - m) \\ = \sum_{m=0}^\infty x(n - m)h(m), & x(n) \text{ noncausal, } h(n) \text{ causal} \\ \sum_{m=-\infty}^\infty x(m)h(n - m) \\ = \sum_{m=-\infty}^\infty x(n - m)h(m), & x(n), h(n) \text{ noncausal} \end{cases} \quad (21.1.2)$$

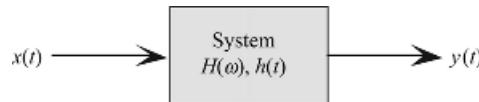


FIGURE 21.1.1

The Fourier transform (FT) of the impulse response  $h(t)$  is called the *transfer function*  $H(\omega)$  and is shown below:

$$H(\omega) = \int_0^{\infty} h(t) e^{-j\omega t} dt \quad (21.1.3)$$

The input FT  $X(\omega)$  and the output FT  $Y(\omega)$  are related by

$$Y(\omega) = H(\omega)X(\omega) \quad (21.1.4)$$

The time-invariant system can also be characterized in continuous time by state equations of the form

$$\begin{aligned} \frac{d}{dt} \mathbf{x}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \mathbf{x}(0) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) \end{aligned} \quad (21.1.5)$$

and in discrete time by

$$\begin{aligned} \mathbf{x}(n+1) &= \mathbf{A}\mathbf{x}(n) + \mathbf{B}\mathbf{u}(n), \mathbf{x}(0) \\ \mathbf{y}(n) &= \mathbf{C}\mathbf{x}(n) \end{aligned} \quad (21.1.6)$$

In Eqs. (21.1.5) and (21.1.6)  $\mathbf{x}$  is an  $n$ -vector,  $\mathbf{A}$  is an  $n \times n$  matrix,  $\mathbf{B}$  is an  $n \times m$  matrix,  $\mathbf{u}$  is an  $m$ -vector,  $\mathbf{y}$  is a  $p$ -vector,  $\mathbf{C}$  is a  $p \times n$  matrix, and  $\mathbf{D}$  is a  $p \times m$  matrix. If these equations are time-varying, then the matrices  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  will become  $\mathbf{A}(t)$ ,  $\mathbf{B}(t)$ , and  $\mathbf{C}(t)$ . The corresponding matrix solutions for  $\mathbf{y}$  for a causal system with initial condition  $\mathbf{x}(0)$  are

*Continuous Time:*

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{C}\Phi(t)\mathbf{x}(0) + \int_0^t \Phi(t-\tau)\mathbf{B}\mathbf{u}(\tau)d\tau, \text{time-invariant} \\ \mathbf{y}(t) &= \mathbf{C}(t)\Phi(t,0)\mathbf{x}(0) + \int_0^t \Phi(t,\tau)\mathbf{B}(t)\mathbf{u}(\tau)d\tau, \text{time-varying} \end{aligned} \quad (21.1.7)$$

*Discrete Time:*

$$\begin{aligned} \mathbf{y}(n) &= \mathbf{C}\Phi(n)\mathbf{x}(0) + \sum_{m=0}^n \Phi(n-m)\mathbf{B}\mathbf{u}(m), \text{time-invariant} \\ \mathbf{y}(n) &= \mathbf{C}(n)\Phi(n,0)\mathbf{x}(0) + \sum_{m=0}^n \Phi(n,m)\mathbf{B}(m)\mathbf{u}(m), \text{time-varying} \end{aligned} \quad (21.1.8)$$

In Eqs. (21.1.7) and (21.1.8) the term  $\Phi(\cdot)$  is called the *state transition matrix*.

The time-invariant system in Eq. (21.1.5) can be solved in the frequency domain by Laplace transform (LT) techniques. Taking unilateral LT of the first equation in

Eq. (21.1.5) and using the initial condition  $\mathbf{x}(0)$ , we obtain

$$s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{AX}(s) + \mathbf{BU}(s) \text{ or } (s\mathbf{I} - \mathbf{A})\mathbf{X}(s) = \mathbf{x}(0) + \mathbf{BU}(s)$$

Hence

$$\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(0) + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{BU}(s) \quad (21.1.9)$$

The inverse LT of the matrix  $(s\mathbf{I} - \mathbf{A})^{-1}$  term by term yields the state transition matrix  $\Phi(t)$ . The solution for  $\mathbf{Y}(s)$  is given by substituting Eq. (21.1.9) in the LT of the second equation in Eq. (21.1.5):

$$\mathbf{Y}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(0) + \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{BU}(s) \quad (21.1.10)$$

In general, deterministic systems deal with causal input  $x(t)$  and causal impulse response  $h(t)$ , and the resulting output  $y(t)$  will also be causal. However, for random inputs to a system,  $x(t)$  need not be causal whereas  $h(t)$  will be causal. The resulting output  $y(t)$  also will not be causal, as shown in the following example.

**Example 21.1.1** In the system represented by Fig. 21.1.1,  $x(t)$ , is a noncausal signal given by  $x(t) = e^{-2|t|}$  and the impulse response is causal, given by  $h(t) = e^{-t}u(t)$ . We will find the output using Laplace transforms (LTs). The LTs of  $x(t)$  and  $h(t)$  are given by

$$X(s) = \frac{4}{4 - s^2} \quad \text{and} \quad H(s) = \frac{1}{s + 1}$$

The LT  $Y(s)$  of the output  $y(t)$  is given by

$$Y(s) = H(s)X(s) = \frac{4}{4 - s^2} \cdot \frac{1}{s + 1}$$

The partial fraction expansion of  $Y(s)$  yields

$$Y(s) = \frac{4}{4 - s^2} \cdot \frac{1}{s + 1} = \frac{4}{3(s + 1)} - \frac{1}{s + 2} + \frac{1}{3(2 - s)}$$

and the inverse LT gives the output as

$$y(t) = \frac{4}{3}e^{-t}u(t) - e^{-2t}u(t) + \frac{1}{3}e^{2t}u(-t)$$

$x(t)$ ,  $h(t)$ , and  $y(t)$  are shown in Fig. 21.1.2.

The results can also be obtained in the time domain directly from the convolution integral

$$y(t) = \int_0^\infty x(t - \tau)h(\tau)d\tau$$

and  $h(\tau)$  and  $x(t - \tau)$  are shown in Fig. 21.1.3. From the figure we can evaluate the convolution.

Since  $x(t - \tau)$  is given by

$$x(t - \tau) = e^{-2|t-\tau|} = \begin{cases} e^{-2(t-\tau)}, & \tau \leq t \\ e^{2(t-\tau)}, & \tau > t \end{cases}$$

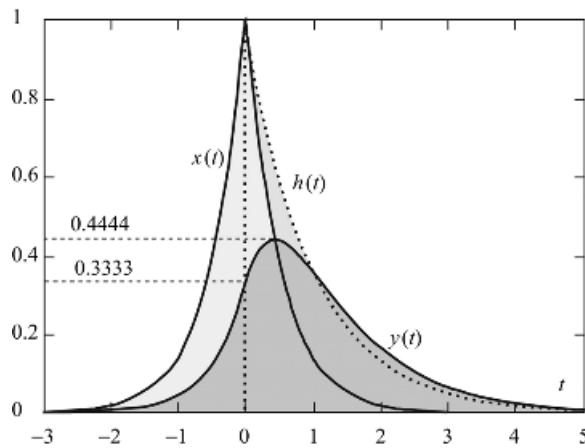


FIGURE 21.1.2

we can calculate  $y(t)$  in the two regions  $t \leq 0$  and  $t > 0$ . For  $t \leq 0$ , we have

$$y(t) = \int_0^\infty e^{2(t-\tau)} e^{-\tau} d\tau = e^{2t} \int_0^\infty e^{-3\tau} d\tau = \frac{1}{3} e^{2t}, \quad t \leq 0, \tau > t$$

The case for  $t > 0$  is shown in Fig. 21.1.3. We calculate the convolution integral in the two regions  $\tau \leq t$  and  $\tau > t$  as shown below:

$$\begin{aligned} y(t) &= \int_0^t e^{-2(t-\tau)} e^{-\tau} d\tau + \int_t^\infty e^{2(t-\tau)} e^{-\tau} d\tau = e^{-2t} \int_0^t e^\tau d\tau + e^{2t} \int_t^\infty e^{-3\tau} d\tau \\ &= e^{-t} - e^{-2t} + \frac{1}{3} e^{-t}, \quad t > 0 \end{aligned}$$

The final result for  $y(t)$  is as obtained by the transform method.

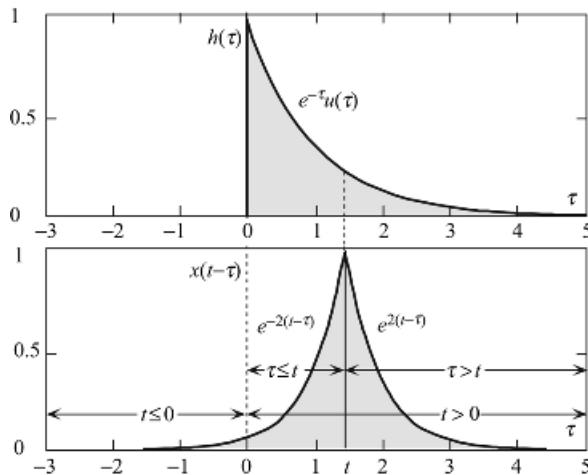


FIGURE 21.1.3

## 21.2 RANDOM PROCESSES THROUGH LINEAR SYSTEMS

The input to the system of Fig. 21.1.1 is a random process  $\{X(t), -\infty < t < \infty\}$ . The system function  $h(t)$  is causal. The sample output  $Y(t, \xi)$  of the system for each sample function  $X(t, \xi)$  is given by

$$Y(t, \xi) = \int_0^\infty X(t - \tau, \xi)h(\tau)d\tau \quad (21.2.1)$$

This is a more convenient form for a noncausal  $X(t)$  and causal  $h(t)$  than the form

$$Y(t, \xi) = \int_{-\infty}^t X(t, \xi)h(t - \tau)d\tau \quad (21.2.2)$$

Under the following mild conditions on  $h(t)$

$$\int_0^\infty |h(\tau)|^2 d\tau < \infty \quad \text{and} \quad \int_0^\infty h(t - \tau)C_X(\tau, s)h(t - s)d\tau < \infty \quad (21.2.3)$$

the output  $\{Y(t), -\infty < t < \infty\}$  is also a random process.

We can now obtain the statistics of the random process  $Y(t)$  under the condition that the input random process  $X(t)$  is stationary. (It has already been mentioned that *stationary* means *wide-sense stationary*.) The mean value of  $Y(t)$  is obtained by taking expectations of both sides of the equation

$$Y(t) = \int_0^\infty X(t - \tau)h(\tau)d\tau$$

or

$$E[Y(t)] = \mu_Y(t) = \int_0^\infty E[X(t - \tau)]h(\tau)d\tau = \mu_X \int_0^\infty h(\tau)d\tau = \mu_Y \quad (21.2.4)$$

and on substituting  $H(0) = \int_0^\infty h(\tau)d\tau$  from Eq. (21.1.3), we obtain

$$\mu_Y = \mu_X H(0) \quad (21.2.5)$$

Thus the mean value of  $Y(t)$  is a constant.

The first step in obtaining the autocorrelation function  $R_Y(t, \tau)$  of the output  $Y(t)$  is to premultiply Eq. (21.2.4) by  $X(t - \tau)$  and take expectations. Thus

$$E[X(t - \tau)Y(t)] = \int_0^\infty E[X(t - \tau)X(t - \xi)]h(\xi)d\xi$$

or

$$R_{XY}(\tau) = \int_0^\infty R_X(\tau - \xi)h(\xi)d\xi$$

or

$$R_{XY}(\tau) = R_X(\tau) * h(\tau) \quad (21.2.6)$$

Thus the cross-correlation function  $R_{XY}(\tau)$  between the input  $X(t)$  and the output  $Y(t)$  is the convolution between the input autocorrelation function  $R_X(\tau)$  and the impulse

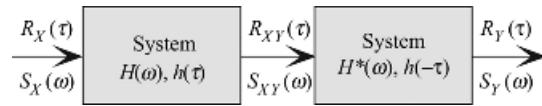


FIGURE 21.2.1

response  $h(\tau)$ . The second step is to postmultiply Eq. (21.2.4) by  $Y(t + \tau)$  and take expectations. Thus

$$E[Y(t)Y(t + \tau)] = \int_0^\infty E[X(t - \xi)Y(t + \tau)]h(\xi)d\xi$$

or

$$R_Y(\tau) = \int_0^\infty R_{XY}(\tau + \xi)h(\xi)d\xi$$

or

$$R_Y(\tau) = R_{XY}(\tau) * h(-\tau) \quad (21.2.7)$$

Combining Eqs. (21.2.6) and (21.2.7), we obtain

$$R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau) \quad (21.2.8)$$

In many cases it is easier to solve Eq. (21.2.8) in the frequency domain. Taking the Fourier transform of Eq. (21.2.8), the power spectral densities of the input and the output are related by

$$\begin{aligned} S_{XY}(\omega) &= S_X(\omega)H(\omega) \\ S_Y(\omega) &= S_{XY}(\omega)H^*(\omega) \\ S_Y(\omega) &= S_X(\omega)|H(\omega)|^2 \end{aligned} \quad (21.2.9)$$

The schematic of the input–output transformation is shown in Fig. 21.2.1.

In conclusion, we can state that if a stationary random process  $X(t)$  is passed through a linear system whose impulse response  $h(t)$  satisfies the regularity conditions of Eq. (21.2.3), then the output  $Y(t)$  is also a stationary process.

In most problems involving random processes through linear systems, the output autocorrelation  $R_Y(\tau)$  can be more easily found by first finding the psd  $S_Y(\omega)$  from Eq. (21.1.9) and taking the inverse Fourier transform.

**Example 21.2.1** A white-noise process  $X(t)$  with autocorrelation function  $R_X(\tau) = \sigma^2\delta(\tau)$  is passed through a linear system whose impulse response  $h(t)$  is given by  $h(t) = e^{-\alpha t}u(t)$ . We want to find the mean value  $\mu_Y$  of the output, the cross-correlation function  $R_{XY}(\tau)$ , the output autocorrelation function  $R_Y(\tau)$ , and the variance of the output  $\sigma_Y^2$ .

The transfer function  $H(\omega)$  and its absolute value squared  $|H(\omega)|^2$  are given by

$$H(\omega) = \frac{1}{\alpha + j\omega}, |H(\omega)|^2 = \frac{1}{\alpha^2 + \omega^2}$$

and the input power spectral density is given by  $S_X(\omega) = \sigma^2$ . The mean value  $\mu_Y$  is

$$\mu_Y = \mu_X H(0) = \frac{\mu_X}{\alpha} = 0 \quad \text{since} \quad \mu_X = 0$$

The cross-spectral density  $S_{XY}(\omega)$  and the output psd  $S_Y(\omega)$  from Eq. (21.2.9) are

$$S_{XY}(\omega) = \sigma^2 H(\omega) = \frac{\sigma^2}{\alpha + j\omega}; \quad S_Y(\omega) = S_X(\omega) |H(\omega)|^2 = \frac{\sigma^2}{\alpha^2 + \omega^2}$$

Taking inverse Fourier transforms of  $S_{XY}(\omega)$  and  $S_Y(\omega)$ , we obtain the cross-correlation function  $R_{XY}(\tau)$  and the autocorrelation function  $R_Y(\tau)$ :

$$R_{XY}(\tau) = \sigma^2 e^{-\alpha\tau} u(\tau); \quad R_Y(\tau) = \frac{\sigma^2}{2\alpha} e^{-\alpha|\tau|}$$

Since the mean  $\mu_Y = 0$ , the variance of the output is  $\sigma_Y^2 = R_Y(0) = E[Y^2] = \sigma^2/2\alpha$ . This filter is called a *shaping filter* since it shapes a white noise to any desired psd by means of the factor  $\alpha$ .

**Example 21.2.2** The autocorrelation function  $R_X(\tau)$  of the input random process  $X(t)$  in the previous example is modified as  $R_X(\tau) = a^2 + \sigma^2 \delta(\tau)$  with the impulse response  $h(t) = e^{-\alpha t} u(t)$ . The mean value of  $X(t)$  from Eq. (19.2.26) is  $\mu_X = a$ . Hence the mean value of  $Y(t)$  is

$$\mu_Y = \mu_X H(0) = \frac{\mu_X}{\alpha} = \frac{a}{\alpha}$$

The input psd is  $S_X(\omega) = a^2 2\pi \delta(\omega) + \sigma^2$ . The cross-spectral density  $S_{XY}(\omega)$  and the output psd  $S_Y(\omega)$  from Eq. (21.2.9) are

$$S_{XY}(\omega) = [\sigma^2 + 2\pi a^2 \delta(\omega)] H(\omega) = \frac{\sigma^2}{\alpha + j\omega} + \frac{2\pi a^2}{\alpha} \delta(\omega)$$

$$S_Y(\omega) = \frac{\sigma^2}{\alpha^2 + \omega^2} + \frac{a^2}{\alpha^2} 2\pi \delta(\omega)$$

The corresponding cross-correlation and autocorrelation functions of  $Y(t)$  are

$$R_{XY}(\tau) = \sigma^2 e^{-\alpha\tau} u(\tau) + \frac{a^2}{\alpha}; \quad R_Y(\tau) = \frac{\sigma^2}{2\alpha} e^{-\alpha|\tau|} + \frac{a^2}{\alpha^2}$$

**Example 21.2.3** In this example the input  $X(t)$  is a random telegraph wave as in Example 19.2.7 with  $h(t) = e^{-\alpha t} u(t)$  and autocorrelation function  $R_X(\tau) = \frac{1}{4} \{1 + e^{-2\lambda|\tau|}\}$ . We have to find the statistics of the output process  $Y(t)$ .

The mean value, the autocovariance, and the variance  $X(t)$  are

$$\mu_X = \sqrt{\frac{1}{4}} = \frac{1}{2} \quad C_X(\tau) = \frac{1}{4} e^{-2\lambda|\tau|} \quad \sigma_X^2 = C_X(0) = \frac{1}{4}$$

The input psd is

$$S_X(\omega) = \frac{1}{4} \left\{ 2\pi \delta(\omega) + \frac{4\lambda}{4\lambda^2 + \omega^2} \right\}$$

The mean value of the output is

$$\mu_Y = \mu_X H(0) = \frac{\mu_X}{\alpha} = \frac{1}{2\alpha}$$

The cross-spectral density  $S_{XY}(\omega)$  and the output psd  $S_Y(\omega)$  from Eq. (21.2.9) are

$$S_{XY}(\omega) = \frac{1}{4} \left\{ 2\pi\delta(\omega) + \frac{4\lambda}{4\lambda^2 + \omega^2} \right\} \frac{1}{\alpha + j\omega} = \frac{2\pi\delta(\omega)}{4\alpha} + \frac{\lambda}{(4\lambda^2 + \omega^2)(\alpha + j\omega)}$$

$$S_Y(\omega) = \frac{1}{4} \left\{ 2\pi\delta(\omega) + \frac{4\lambda}{4\lambda^2 + \omega^2} \right\} \frac{1}{\alpha^2 + \omega^2} = \frac{2\pi\delta(\omega)}{4\alpha^2} + \frac{\lambda}{(4\lambda^2 + \omega^2)(\alpha^2 + \omega^2)}$$

The partial fraction expansions of  $S_{XY}(\omega)$  and  $S_Y(\omega)$  are

$$S_{XY}(\omega) = \frac{2\pi\delta(\omega)}{4\alpha} + \frac{1}{2\lambda + j\omega} + \frac{1}{2\lambda - j\omega} - \frac{\lambda}{\alpha + j\omega}$$

$$S_Y(\omega) = \frac{2\pi\delta(\omega)}{4\alpha^2} + \frac{1}{4(\alpha^2 - 4\lambda^2)} \frac{4\lambda}{(4\lambda^2 + \omega^2)} - \frac{\lambda}{2\alpha(\alpha^2 - 4\lambda^2)} \frac{2\alpha}{(\alpha^2 + \omega^2)}$$

The corresponding cross-correlation and autocorrelation functions of  $Y(t)$  are

$$R_{XY}(\tau) = \begin{cases} \frac{1}{4} \left( \frac{e^{2\lambda\tau}}{\alpha + 2\lambda} + \frac{1}{\alpha} \right), & \tau \leq 0 \\ \frac{1}{4} \left( \frac{e^{-2\lambda\tau}}{\alpha - 2\lambda} + \frac{1}{\alpha} \right) - \frac{\lambda e^{-\alpha\tau}}{\alpha^2 - 4\lambda^2}, & \tau > 0 \end{cases}$$

$$R_Y(\tau) = \frac{1}{4} \left( \frac{e^{-2\lambda|\tau|}}{\alpha^2 - 4\lambda^2} + \frac{1}{\alpha^2} \right) - \frac{\lambda e^{-\alpha|\tau|}}{2\alpha(\alpha^2 - 4\lambda^2)}$$

The cross-spectral density  $S_{XY}(\omega)$ , the cross-correlation function  $R_{XY}(\tau)$ , the power spectral density  $S_Y(\omega)$ , and the autocorrelation function  $R_Y(\tau)$  are shown for  $\lambda=1$  and  $\alpha = 0.5$  in Figs. 21.2.2–21.2.5. The variance of the output process  $Y(t)$  is given by

$$\sigma_Y^2 = R_Y(0) - \mu_Y^2 = \frac{1}{4(\alpha^2 - 4\lambda^2)} - \frac{\lambda}{2\alpha(\alpha^2 - 4\lambda^2)} = 0.2$$

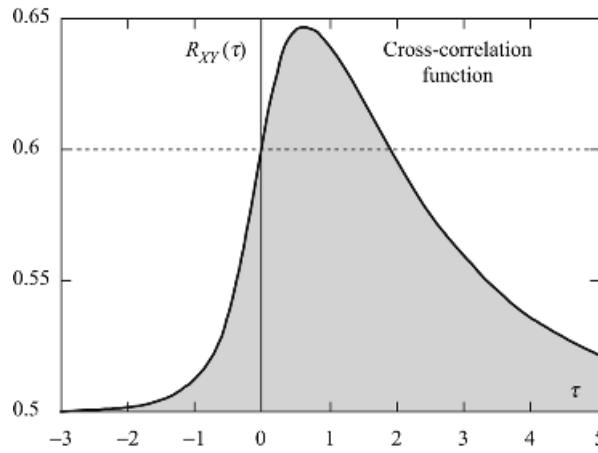


FIGURE 21.2.2

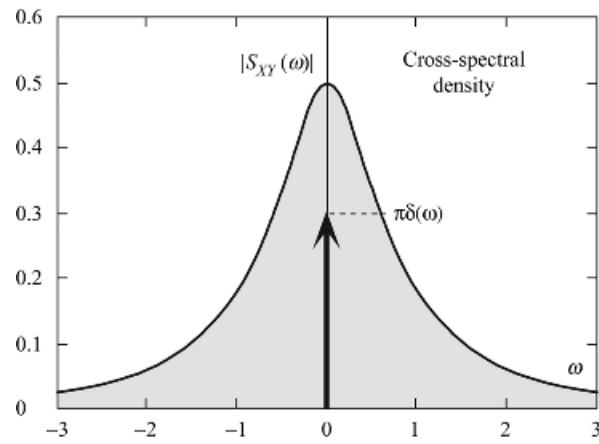


FIGURE 21.2.3

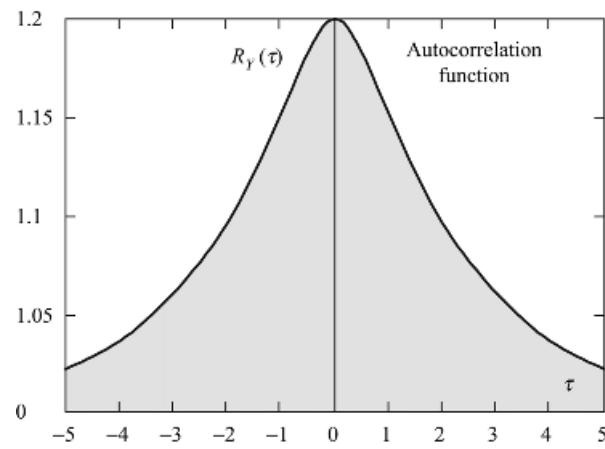


FIGURE 21.2.4

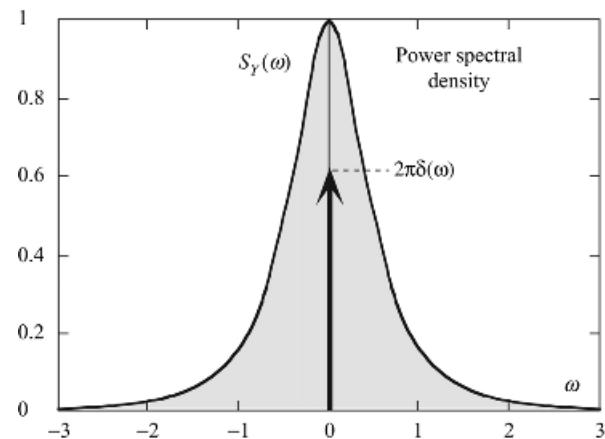


FIGURE 21.2.5

**Example 21.2.4** The input to a linear system is a zero mean stationary random process  $X(t)$ , and the output is  $Y(t)$ . Their psds are given by

$$S_X(\omega) = \frac{1}{4 + \omega^2}; \quad S_Y(\omega) = \frac{4}{(4 + \omega^2)(64 + \omega^2)}.$$

The autocorrelation of the input and the output are given by their respective inverse Fourier transforms (IFT):

$$\begin{aligned} R_X(\tau) &= \text{IFT} \left\{ \frac{1}{4(4 + \omega^2)} \right\} = \frac{e^{-2|\tau|}}{4} \\ R_Y(\tau) &= \text{IFT} \left\{ \frac{4}{(4 + \omega^2)(64 + \omega^2)} \right\} = \text{IFT} \left\{ \frac{1}{60} \frac{4}{(4 + \omega^2)} - \frac{1}{240} \frac{16}{(64 + \omega^2)} \right\} \\ &= \frac{1}{240} (4e^{-2|\tau|} - e^{-8|\tau|}) \end{aligned}$$

Since the mean  $\mu_X = 0$ , the mean  $\mu_Y = 0$ . The variances are

$$\sigma_X^2 = \frac{1}{4}, \quad \sigma_Y^2 = \frac{1}{60} - \frac{1}{240} = \frac{1}{80}$$

Transfer function  $H(\omega)$  and the impulse response  $h(t)$  can be found as follows. From Eq. (21.2.9) we can write

$$|H(\omega)|^2 = \frac{S_Y(\omega)}{S_X(\omega)} = \frac{4}{64 + \omega^2} = H(\omega)H(-\omega) = \frac{2}{(8 + j\omega)} \frac{2}{(8 - j\omega)}$$

Hence the transfer function and the impulse responses are given by

$$H(\omega) = \frac{2}{(8 + j\omega)}; \quad h(t) = 2e^{-8t}u(t)$$

The cross-spectral density is given by

$$S_{XY}(\omega) = \frac{1}{(4 + \omega^2)} \frac{2}{(8 + j\omega)}$$

The partial fraction expansion of  $S_{XY}(\omega)$  yields

$$S_{XY}(\omega) = \frac{1}{12(2 + j\omega)} + \frac{1}{20(2 - j\omega)} - \frac{1}{30(8 + j\omega)}$$

and the cross-correlation function  $R_{XY}(\tau)$  is given by the IFT of  $S_{XY}(\omega)$ :

$$R_{XY}(\tau) = \frac{1}{20} e^{2\tau} u(-\tau) + \frac{1}{12} e^{-2\tau} u(\tau) - \frac{1}{30} e^{-8\tau} u(\tau)$$

The Cross-correlation function  $R_{XY}(\tau)$  is shown in Fig. 21.2.6.

**Example 21.2.5** A random process  $X(t)$  with autocorrelation function  $R_X(\tau) = 1 + 4e^{-2|\tau|}$  excites a linear system with impulse response  $h(t)$ . The autocorrelation function of the output  $Y(t)$  is given by  $R_Y(\tau) = 1 + \frac{8}{3}e^{-|\tau|} - \frac{4}{3}e^{-2|\tau|}$ . We will find the various statistics of this system. Since  $R_X(\infty) = 1$ , the mean value  $\mu_X = 1$ . The variance  $\sigma_X^2 = 4$ .

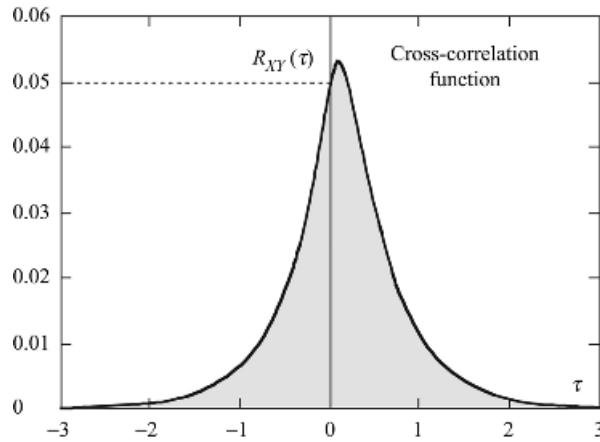


FIGURE 21.2.6

The power spectral densities of  $X(t)$  and  $Y(t)$  are the FTs of the corresponding autocorrelations:

$$\begin{aligned} S_X(\omega) &= 2\pi\delta(\omega) + \frac{16}{4 + \omega^2} \\ S_Y(\omega) &= 2\pi\delta(\omega) + \frac{16}{3(1 + \omega^2)} - \frac{16}{3(4 + \omega^2)} \\ &= 2\pi\delta(\omega) + \frac{16}{(1 + \omega^2)(4 + \omega^2)} \end{aligned}$$

The presence of the delta function complicates the analysis somewhat. If  $H(\omega)$  is the required transfer function, then from Eq. (21.2.9) we have

$$\begin{aligned} \left(2\pi\delta(\omega) + \frac{16}{4 + \omega^2}\right)|H(\omega)|^2 &= 2\pi\delta(\omega) + \frac{16}{(1 + \omega^2)(4 + \omega^2)} \\ 2\pi\delta(\omega)|H(0)|^2 + \frac{16}{4 + \omega^2}|H(\omega)|^2 &= 2\pi\delta(\omega) + \frac{16}{(1 + \omega^2)(4 + \omega^2)} \end{aligned}$$

Since the singularity term on the lefthand side can be matched only by the singularity term on the righthand side, we have  $|H(0)|^2 = 1$  and

$$\frac{16}{4 + \omega^2}|H(\omega)|^2 = \frac{16}{(1 + \omega^2)(4 + \omega^2)}$$

from which

$$|H(\omega)|^2 = \frac{1}{(1 + \omega^2)} = \frac{1}{1 + j\omega} \cdot \frac{1}{1 - j\omega}$$

Hence, the transfer function  $H(\omega) = 1/(1 + j\omega)$  and the impulse response  $h(t) = e^{-t}u(t)$ .

The mean value of the output  $\mu_Y = \mu_X H(0) = 1$ . The variance of the output is

$$\sigma_Y^2 = R_Y(0) - \mu_Y^2 = \frac{8}{3} - \frac{4}{3} = \frac{4}{3}$$

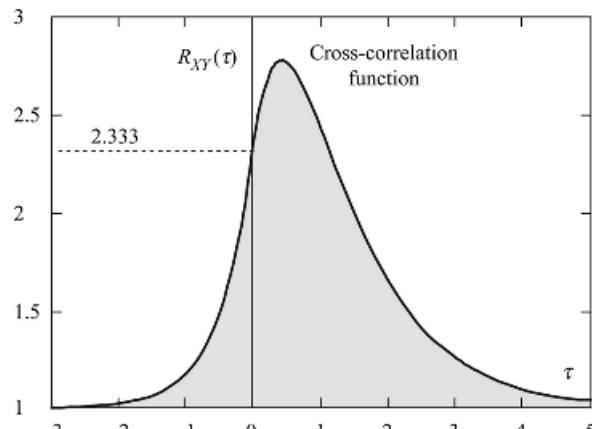


FIGURE 21.2.7

The cross-spectral density  $S_{XY}(\omega)$  is given by

$$S_{XY}(\omega) = \left[ 2\pi\delta(\omega) + \frac{16}{4 + \omega^2} \right] \left( \frac{1}{1 + j\omega} \right) = 2\pi\delta(\omega) + \frac{16}{4 + \omega^2} \frac{1}{1 + j\omega}$$

and the cross-correlation function  $R_{XY}(\tau)$  by

$$R_{XY}(\tau) = 1 + \frac{4}{3}e^{2\tau}u(-\tau) - 4e^{-2\tau}u(\tau) + \frac{16}{3}e^{-\tau}u(\tau)$$

which is shown in Fig. 21.2.7.

**Example 21.2.6** A zero mean stationary Gaussian stochastic process  $X(t)$  with autocorrelation function  $R_X(\tau) = 100e^{-100|\tau|}$  is applied to a system shown in Fig. 21.2.8.

From the diagram the output  $Y(t)$  is another random process given by  $Y(t) = X(t) - X(t - 0.1\pi)$ . We will determine the various statistics of the random processes. Since  $\mu_X = 0$ , it follows that  $R_X(\tau) = C_X(\tau)$  and hence the variance of  $X(t)$ ,  $\sigma_X^2 = C_X(0) = 100$ . The joint density of  $\{X_1 = X(t - 0.1\pi), X_2 = X(t)\}$  can be found by first obtaining the correlation coefficient  $\rho_{X_1 X_2}$ :

$$E[X_1 X_2] = E[X(t - 0.1\pi)X(t)] = R_X(0.1\pi) = C_X(0.1\pi) = 100e^{-10\pi}$$

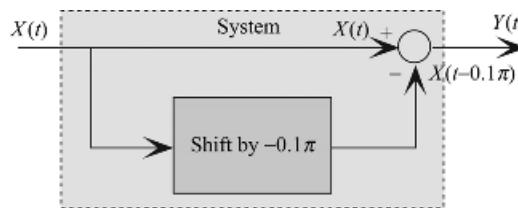


FIGURE 21.2.8

Hence

$$\rho_{X_1 X_2} = \frac{C_X(0.1\pi)}{C_X(0)} = e^{-10\pi}$$

The joint Gaussian density between  $X(t_1)$  and  $X(t_2)$  can be obtained from Eq. (9.3.1) as follows:

$$\begin{aligned} f_{X_1 X_2}(x_1, x_2) &= \frac{1}{2\pi\sigma_X^2\sqrt{1-\rho_{X_1 X_2}^2}} \exp \left[ \begin{aligned} &-\frac{1}{2(1-\rho_{X_1 X_2}^2)} \\ &\times \left( \frac{x_1^2 - 2\rho_{X_1 X_2}x_1 x_2 + x_2^2}{\sigma_X^2} \right) \end{aligned} \right] \\ &= \frac{1}{200\pi\sqrt{1-e^{-20\pi}}} \exp \left[ \begin{aligned} &-\frac{1}{2(1-e^{-20\pi})} \\ &\left( \frac{x_1^2 - 2e^{-10\pi}x_1 x_2 + x_2^2}{100} \right) \end{aligned} \right] \end{aligned}$$

The power spectral density  $S_X(\omega)$  of the input random process  $X(t)$  is

$$S_X(\omega) = \frac{20,000}{10,000 + \omega^2}$$

The transfer function of the system is

$$H(\omega) = \frac{Y(\omega)}{X(\omega)} = \frac{(1 - e^{-j0.1\pi\omega})X(\omega)}{X(\omega)} = 1 - e^{-j0.1\pi\omega}$$

Hence the mean value  $\mu_Y = \mu_X H(0) = 0$ . The cross-spectral density  $S_{XY}(\omega)$  and the power spectral density  $S_Y(\omega)$  are obtained from Eqs. (21.2.9) as follows:

$$\begin{aligned} S_{XY}(\omega) &= S_X(\omega)H(\omega) = \frac{20,000}{10,000 + \omega^2}(1 - e^{-j0.1\pi\omega}) \\ S_Y(\omega) &= S_X(\omega)H(\omega)H(-\omega) = \frac{20,000}{10,000 + \omega^2}(1 - e^{-j0.1\pi\omega})(1 - e^{j0.1\pi\omega}) \\ &= \frac{20,000}{10,000 + \omega^2}(2 - e^{-j0.1\pi\omega} - e^{j0.1\pi\omega}) \\ &= \frac{40,000[1 - \cos(0.1\pi\omega)]}{10,000 + \omega^2} \end{aligned}$$

The graphs of  $|S_{XY}(\omega)|$  and  $S_Y(\omega)$  are shown in Figs. 21.2.9 and 21.2.10.

Taking the IFTs of  $S_{XY}(\omega)$  and  $S_Y(\omega)$  from the FT tables yields the correlation functions  $R_{XY}(\tau)$  and  $R_Y(\tau)$ , given by

$$\begin{aligned} R_{XY}(\tau) &= \text{IFT}[S_{XY}(\omega)] = 100(e^{-100|\tau|} - e^{-100|\tau-0.1\pi|}) \\ R_Y(\tau) &= \text{FT}[S_Y(\omega)] = 100(2e^{-100|\tau|} - e^{-100|\tau-0.1\pi|} - e^{-100|\tau+0.1\pi|}) \end{aligned}$$

The graphs of  $R_{XY}(\tau)$ , and  $R_Y(\tau)$  are shown in Figs. 21.2.11 and 21.2.12.

The variance of the process  $Y(t)$  is given by  $\sigma_Y^2 = C_Y(0) = R_Y(0) = 200$ . With  $\sigma_Y^2 = 200$ , the pdf of  $Y(t)$  is given by

$$f_Y(y) = \frac{1}{20\sqrt{\pi}} e^{-(1/2)\{y^2/200\}}$$

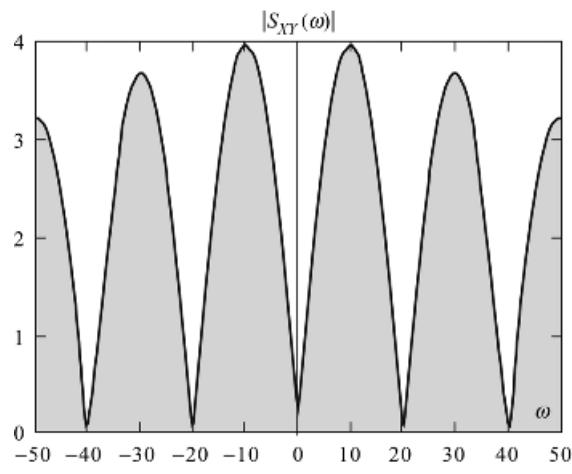


FIGURE 21.2.9

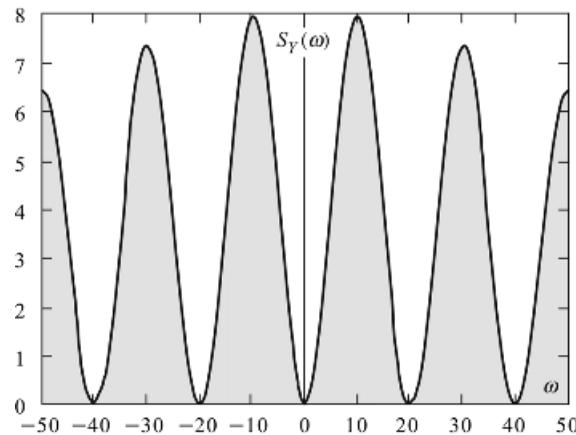


FIGURE 21.2.10

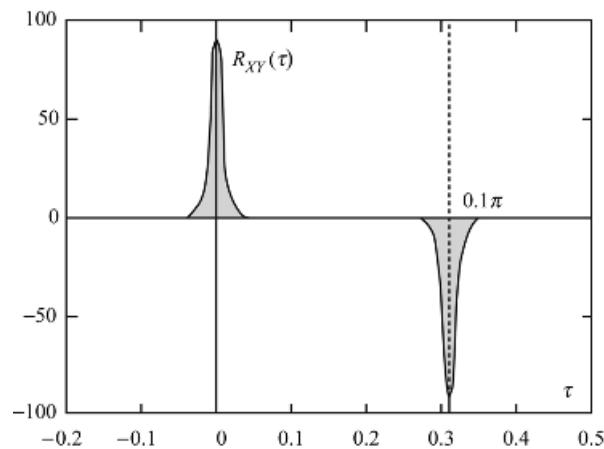


FIGURE 21.2.11

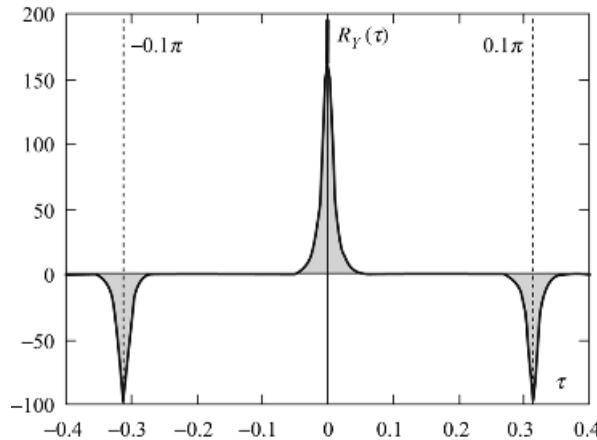


FIGURE 21.2.12

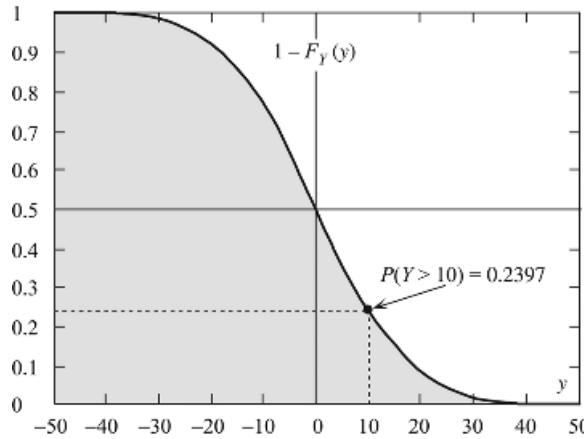


FIGURE 21.2.13

from which the probability of  $Y(t) > y$  can be calculated from

$$P\{Y(t) > y\} = 1 - F_Y(y) = 1 - \frac{1}{20\sqrt{\pi}} \int_{-\infty}^y e^{-(1/2)\{\zeta^2/200\}} d\zeta$$

and  $P(Y(t) > 10)$  can be evaluated. The function  $1 - F_Y(y)$  is shown in Fig. 21.2.13. However, this probability can be evaluated from the Gaussian tables by first substituting  $z = y/\sigma = y/10\sqrt{2}$  in the integral above and obtaining

$$P\{Y(t) > 10\} = 1 - \frac{1}{20\sqrt{\pi}} \int_{-\infty}^{10} e^{-(1/2)\{y^2/200\}} dy = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{1/\sqrt{2}} e^{-(z^2/2)} dz$$

and from the tables  $P\{Y \leq 10\} = 1/\sqrt{2\pi} \int_{-\infty}^{1/\sqrt{2}} e^{-(z^2/2)} dz = 0.7603$ . Hence

$$P\{Y(t) > 10\} = 1 - 0.7603 = 0.2397$$

**Example 21.2.7 (Threshold Filter)** This example is a computer simulation of using the threshold filter to retrieve from the output signal corrupted with white noise, the input

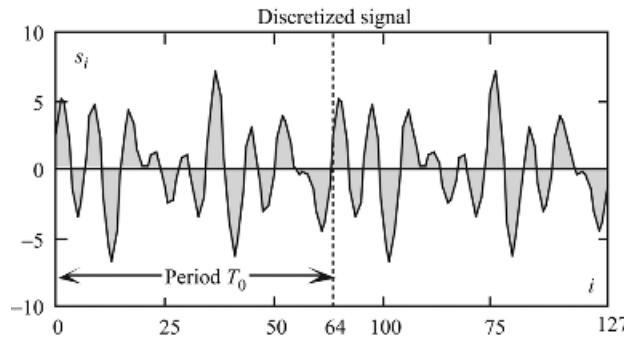


FIGURE 21.2.14

sinusoidal signal in an environment of high signal-to-noise ratio (SNR). The input sinusoidal signal is given by

$$s(t) = 2 \sin(4\omega_0 t) + \frac{5}{2} \cos(7\omega_0 t) + 3 \sin(7\omega_0 t)$$

where  $\omega_0 = (2\pi/T_0) = (2\pi/64)$  and the period  $T_0 = 64$ . The input  $s(t)$  is discretized with  $\Delta t = 1$ , and the discretized signal  $s(i\Delta t) = s_i$ , given by

$$s_i = 2 \sin\left(8 \frac{2\pi i}{128}\right) + \frac{5}{2} \cos\left(14 \frac{2\pi i}{128}\right) + 3 \sin\left(18 \frac{2\pi i}{128}\right), \quad i = 0, \dots, 127$$

is shown in Fig. 21.2.14.

The number of points  $N$  is chosen to be  $N = 2^7 = 128$  for ease in applying the fast Fourier algorithm (FFT) [36] to transform  $s_i$  into the frequency domain.

The zero mean additive discrete white noise  $w_i$  with variance 4 is shown in Fig. 21.2.15. The input power  $P_{s_i}$  is

$$P_{s_i} = \frac{1}{N} \sum_{i=0}^{N-1} s_i^2 = \frac{1}{128} \sum_{i=0}^{127} \left[ 2 \sin\left(8 \frac{2\pi i}{128}\right) + \frac{5}{2} \cos\left(14 \frac{2\pi i}{128}\right) + 3 \sin\left(18 \frac{2\pi i}{128}\right) \right]^2 = 9.625$$

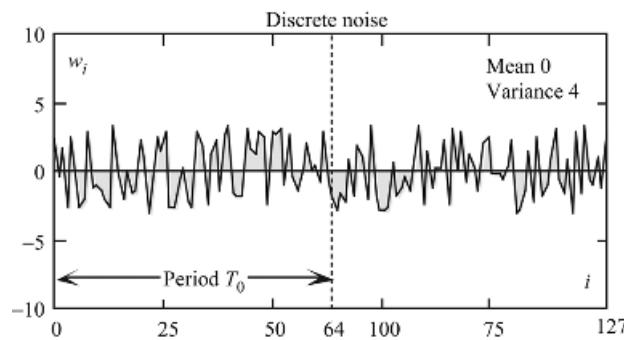


FIGURE 21.2.15

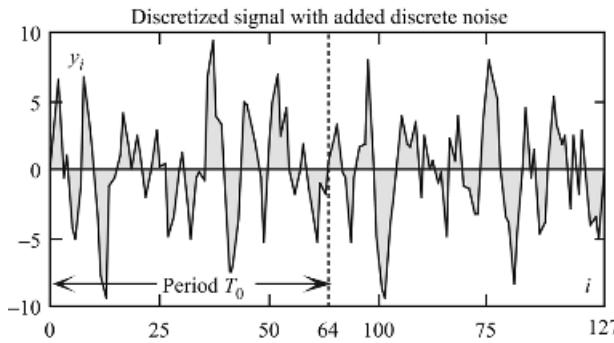


FIGURE 21.2.16

and the noise power  $P_{n_i}$  is

$$P_{n_i} = \frac{1}{N} \sum_{i=0}^{N-1} w_i^2 = \frac{1}{128} \sum_{i=0}^{127} w_i^2 = 4$$

and  $\text{SNR} = 9.625/4 = 2.406$ .

The output signal  $y_i$ , given by

$$y_i = s_i + w_i, \quad i = 0, \dots, N-1$$

is shown in Fig. 21.2.16.

The output power  $P_{y_i}$  is given by

$$P_{y_i} = \frac{1}{N} \sum_{i=0}^{N-1} y_i^2 = \frac{1}{128} \sum_{i=0}^{127} y_i^2 = 14.811$$

The input and the noise signals are transformed into the frequency domain by using the discrete Fourier transform (DFT) [36], defined by

$$S_k = \sum_{i=0}^{N-1} s_i e^{-jk(2\pi/N)i} = \sum_{i=0}^{127} s_i e^{-jk(2\pi/128)i}$$

$$W_k = \sum_{i=0}^{N-1} w_i e^{-jk(2\pi/N)i} = \sum_{i=0}^{127} w_i e^{-jk(2\pi/128)i}$$

For the purely sinusoidal  $s_i$  the DFT  $S_k$  is given by Kronecker delta functions as

$$S_k = \frac{N}{2} \left( -2j\delta_{k-8} + \frac{5}{2}\delta_{k-14} - 3j\delta_{k-18} \right)$$

where  $\delta_m$  is the Kronecker delta equaling 1 when  $m = 0$  and 0 when  $m \neq 0$ . The DFT of the output  $y_i$  is given by

$$Y_k = S_k + W_k, \quad k = 0, \dots, N-1$$

The input and the output spectra  $|S_k|$  and  $|Y_k|$  are shown in Fig. 12.2.17. At frequencies of 8, 14, and 18, the peak values of the spectra are given in Table 21.2.1.

In Fig. 21.2.17, if we choose those data points  $Y_k$  that are above a threshold value  $\alpha$ , then we can isolate all the peaks. A simple threshold filter can be designed by multiplying

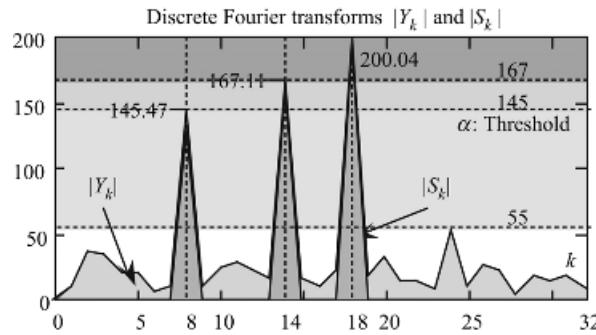


FIGURE 21.2.17

TABLE 21.2.1

$f$	Input Spectrum	Output Spectrum
8	$ S_8  = 128$	$ Y_8  = 145.47$
14	$ S_{14}  = 160$	$ Y_{14}  = 167.11$
18	$ S_{18}  = 192$	$ Y_{18}  = 200.04$

$\{Y_k, k = 0, \dots, N - 1\}$  by a unit step function  $u(|Y_k| - \alpha)$ , defined by

$$u(|Y_k| - \alpha) = \begin{cases} 1, & |Y_k| \geq \alpha \\ 0, & |Y_k| < \alpha \end{cases}$$

The output of this filter will be only those values of  $Y_k$  that are above the threshold. If  $55 \leq \alpha \leq 145$ , then the output will be all three peaks. If  $146 \leq \alpha \leq 167$ , then the output will be only two peaks, and if  $168 \leq \alpha \leq 200$ , the output will be only one peak. Finally, if  $\alpha > 200$ , the output will be zero. The output  $Z_k$  of this filter will be given by

$$Z_k = Y_k u(|Y_k| - \alpha), \quad k = 0, \dots, N - 1$$

TABLE 21.2.2

$k$	$ S_k $	$ Y_k $	$u( Y_k  - \alpha)$	$ Z_k $
5	0	19.709	0	0
6	0	5.858	0	0
7	0	11.158	0	0
8	128	145.473	1	145.473
9	0	11.091	0	0
10	0	23.615	0	0
11	0	28.109	0	0
12	0	23.032	0	0
13	0	15.388	0	0
14	160	167.111	1	167.111
15	0	15.412	0	0
16	0	10.366	0	0
17	0	22.076	0	0
18	192	200.036	1	200.036
19	0	17.753	0	0
20	0	32.344	0	0

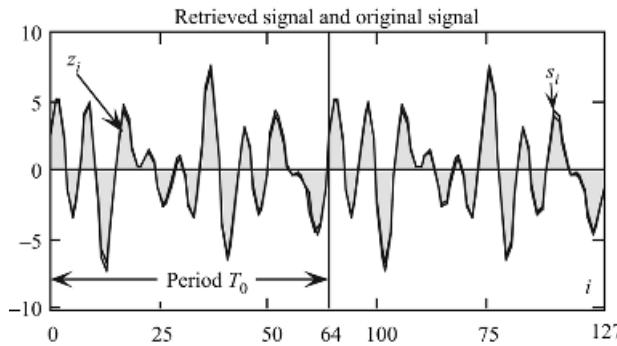


FIGURE 21.2.18

To clarify the operation of the threshold filter, the quantities,  $|S_k|$ ,  $|Y_k|$ ,  $u(|Y_k| - \alpha)$ , and  $|Z_k|$  are tabulated in Table 21.2.2.

The inverse DFT of  $Z_k$  yields the signal  $z_i$  that will approximate the input signal  $s_i$ :

$$\text{IDFT}\{Z_k\} = z_i \approx s_i$$

With  $\alpha = 65$ , this filter as applied and the IDFT of  $Z_k$  yields the retrieved signal  $z_i$  shown in Fig. 21.2.18 along with the superposed original signal  $s_i$ . We see from the figure that the retrieved signal is almost the same as the original signal.

The power  $P_{z_i}$  of the retrieved signal is

$$P_{z_i} = \frac{1}{N} \sum_{i=0}^{N-1} z_i^2 = \frac{1}{128} \sum_{i=0}^{127} z_i^2 = 10.877$$

## 21.3 LINEAR FILTERS

The previous section dealt with only lowpass filters. Among ideal filters we can define lowpass, highpass, and bandpass filters. The transfer functions and impulse responses of these filters are defined as follows:

*Ideal Lowpass:*

$$H_{\text{low}}(\omega) = \begin{cases} 1, & |\omega| \leq \omega_B \\ 0, & \text{otherwise} \end{cases}; \quad h_{\text{low}}(t) = \frac{\sin(\omega_B t)}{\pi t} \quad (21.3.1)$$

*Ideal Highpass:*

$$H_{\text{high}}(\omega) = \begin{cases} 1, & |\omega| > \omega_B \\ 0, & \text{otherwise} \end{cases}; \quad h_{\text{high}}(t) = \delta(t) - \frac{\sin(\omega_B t)}{\pi t} \quad (21.3.2)$$

*Ideal Bandpass:*

$$H_{\text{band}}(\omega) = \begin{cases} 1, & |\omega - \omega_c| \leq \omega_B \\ 0, & \text{otherwise} \end{cases}; \quad h_{\text{band}}(t) = \frac{\sin(\omega_B t)}{\pi t} \cos(\omega_c t), \quad \omega_c > 2\omega_B \quad (21.3.3)$$

In these equations  $\omega_B$  is the *cutoff frequency* and  $\omega_c$  is the *carrier frequency*. The transfer functions and the impulse responses are shown in Fig. 21.3.1.

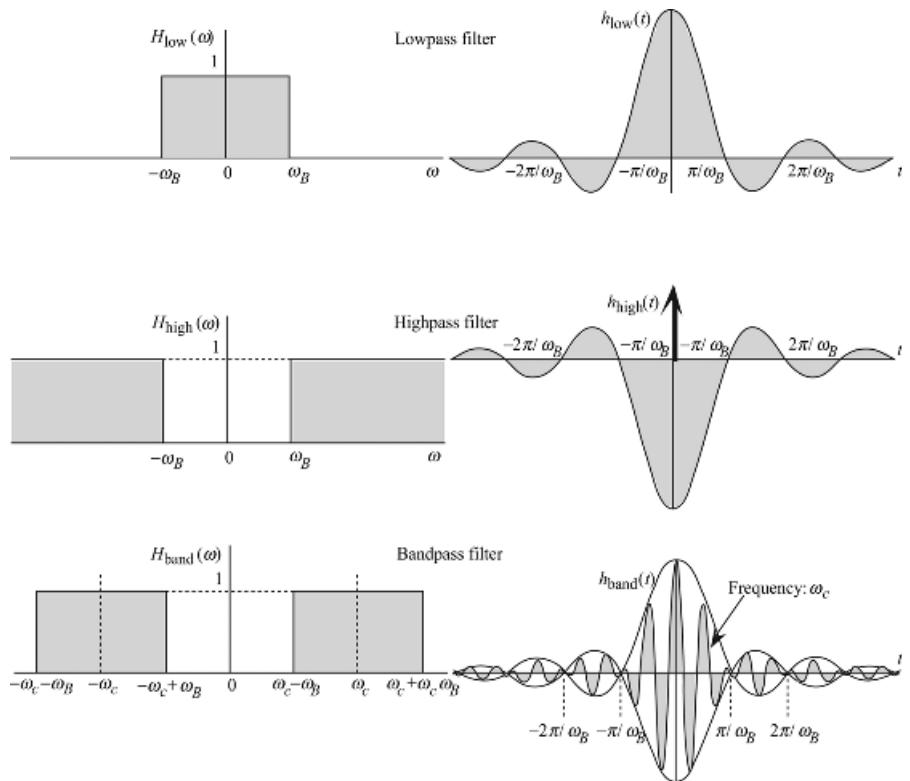


FIGURE 21.3.1

As seen in the figures, the impulse responses are noncausal and hence the filters are not physically realizable. All physically realizable transfer functions must satisfy the Paley–Wiener criterion [45], given by

$$\int_{-\infty}^{\infty} \frac{\ln |H(\omega)|}{1 + \omega^2} d\omega < \infty \quad (21.3.4)$$

Nevertheless, these transfer functions can be approximated by physically realizable transfer functions.

**Example 21.3.1** A random telegraph wave  $X(t)$  with mean  $\mu_X = 0$  and autocorrelation function  $R_X(\tau) = Ke^{-\alpha|\tau|}$  is multiplied with a carrier signal  $\cos(\omega_c t + \Phi)$ , where  $\Phi$  is a random phase independent of  $X(t)$  and  $K = 2$ ,  $\alpha = 3$ , and  $\omega_c = 5\pi$ . We will calculate the statistics of the output  $Y(t) = X(t)\cos(\omega_c t + \Phi)$  and devise a scheme to retrieve the input signal. Using the independence of  $X(t)$  and  $\Phi$ , we have  $\mu_Y = \mu_X E[\cos(\omega_c t + \Phi)] = 0$  since  $\mu_X$  and  $E[\cos(\omega_c t + \Phi)] = 0$ . Again, from the independence of  $X(t)$  and  $\Phi$ , the autocorrelation function  $R_Y(\tau)$  can be calculated as follows:

$$\begin{aligned} R_Y(\tau) &= E[Y(t)Y(t + \tau)] = E[X(t)X(t + \tau)]E[\cos(\omega_c t + \Phi)\cos(\omega_c(t + \tau) + \Phi)] \\ &= E[X(t)X(t + \tau)]\frac{1}{2}E[\cos(\omega_c\tau) + \cos(\omega_c(2t + \tau) + 2\Phi)] \\ &= \frac{1}{2}R_X(\tau)\cos(\omega_c\tau) = \frac{K}{2}e^{-\alpha|\tau|}\cos(\omega_c\tau) \end{aligned}$$

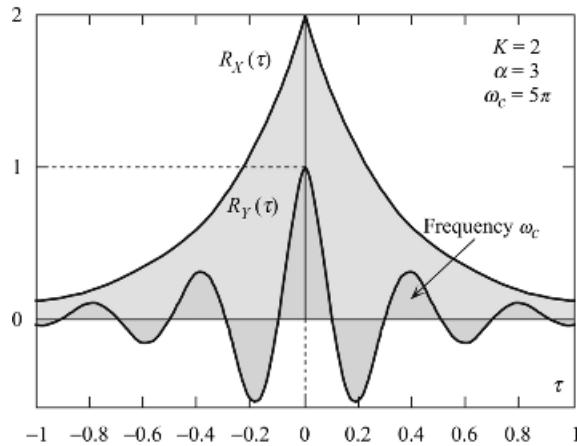


FIGURE 21.3.2

Hence the random process  $Y(t)$  is also stationary. The two autocorrelation functions  $R_X(\tau)$  and  $R_Y(\tau)$  are shown in Fig. 21.3.2.

The psd  $S_X(\omega)$  of the input process  $X(t)$  is given by

$$S_X(\omega) = \frac{2K\alpha}{\alpha^2 + \omega^2}$$

From the form of  $R_Y(\tau)$ , we conclude that  $Y(t)$  is a bandpass signal, and the psd  $S_Y(\omega)$  of the output process  $Y(t)$  is given by

$$\begin{aligned} S_Y(\omega) &= \frac{1}{2\pi} \frac{S_X(\omega)}{2} * \pi[\delta(\omega + \omega_c) + \delta(\omega - \omega_c)] \\ &= \frac{1}{4} [S_X(\omega + \omega_c) + S_X(\omega - \omega_c)] \\ &= \frac{2K\alpha}{4} \left[ \frac{1}{\alpha^2 + (\omega + \omega_c)^2} + \frac{1}{\alpha^2 + (\omega - \omega_c)^2} \right] \end{aligned}$$

The first step in retrieving the original signal  $X(t)$  is to multiply  $Y(t)$  by a carrier  $\cos(\omega_c t + \Theta)$  where  $\Theta$  is another random phase angle independent of  $Y(t)$ . Thus,  $Z(t) = Y(t)\cos(\omega_c t + \Theta)$ , and the autocorrelation  $R_Z(\tau)$  is given by

$$\begin{aligned} R_Z(\tau) &= E[Z(t)Z(t + \tau)] = E[Y(t)Y(t + \tau)]E[\cos(\omega_c t + \Theta)\cos(\omega_c(t + \tau) + \Theta)] \\ &= E[Y(t)Y(t + \tau)] \frac{1}{2} E[\cos(\omega_c \tau) + \cos(\omega_c(2t + \tau) + 2\Theta)] \\ &= \frac{1}{2} R_Y(\tau) \cos(\omega_c \tau) = \frac{K}{4} e^{-\alpha|\tau|} \cos^2(\omega_c \tau) = \frac{Ke^{-\alpha|\tau|}}{8} [1 + \cos(2\omega_c \tau)] \end{aligned}$$

The two autocorrelation functions  $R_Y(\tau)$  and  $R_Z(\tau)$  are shown in Fig. 21.3.3.

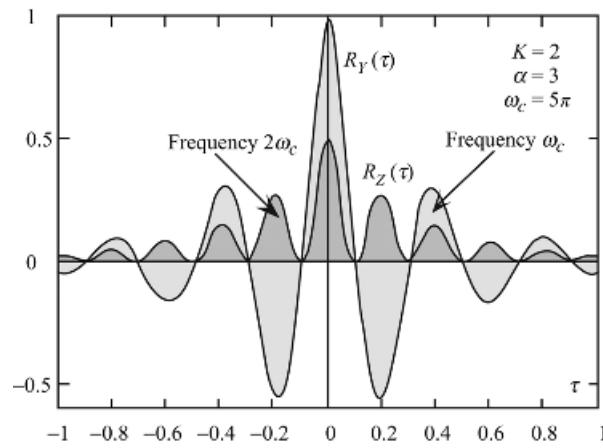


FIGURE 21.3.3

Clearly,  $Z(t)$  is also a stationary process. Its psd is given by

$$\begin{aligned}
 S_Z(\omega) &= \frac{1}{2\pi} \frac{S_Y(\omega)}{2} * \pi[\delta(\omega + \omega_c) + \delta(\omega - \omega_c)] \\
 &= \frac{1}{4} [S_Y(\omega + \omega_c) + S_Y(\omega - \omega_c)] \\
 &= \frac{K\alpha}{8} \left[ \frac{1}{\alpha^2 + (\omega + 2\omega_c)^2} + \frac{1}{\alpha^2 + (\omega - 2\omega_c)^2} + \frac{2}{\alpha^2 + \omega^2} \right]
 \end{aligned}$$

The power spectral densities  $S_X(\omega)$ ,  $S_Y(\omega)$ , and  $S_Z(\omega)$  are shown in Fig. 21.3.4. The maximum value of  $S_X(\omega)$  occurs at zero frequency and is equal to

$$S_{X,\max} = S_X(\omega)|_{\omega=0} = \frac{4}{3}$$

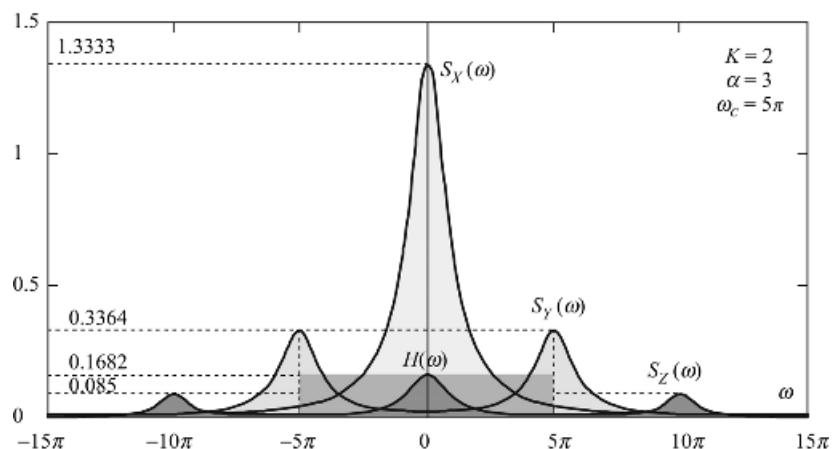


FIGURE 21.3.4

The maximum value of  $S_Y(\omega)$  occurs at frequencies  $\omega_c = \pm 5\pi$  and is equal to,

$$S_{Y,\max} = S_Y(\omega)|_{\omega=\pm 5\pi} = \frac{2}{3} \cdot \frac{9 + 50\pi^2}{9 + 100\pi^2} = 0.3364$$

The value of  $S_Z(\omega)$  occurring at zero frequency is equal to

$$S_Z(\omega)|_{\omega=0} = \frac{1}{3} \cdot \frac{9 + 50\pi^2}{9 + 100\pi^2} = 0.1682$$

and the values of  $S_Z(\omega)$  occurring at frequencies of  $\pm 10\pi$  are equal to

$$S_Z(\omega)|_{\omega=\pm 10\pi} = \frac{1}{3} \cdot \frac{81 + 3150\pi^2 + 10,000\pi^4}{(9 + 100\pi^2)(9 + 400\pi^2)} = 0.085$$

As shown in Fig. 21.3.4, passing the random process  $Z(t)$  through a lowpass filter of bandwidth  $2\omega_c$  and amplitude equal to

$$\frac{1}{3} \cdot \frac{9 + 50\pi^2}{9 + 100\pi^2} = 0.1682$$

multiplied by a suitable gain factor will reconstruct the original signal  $X(t)$ .

### Matched Filters

In the previous section we discussed basic filtering methods where a causal transfer function  $H(\omega)$  was used to shape an input random process  $X(t)$  to any desired output random process  $Y(t)$ . We will take this process one step further where we want to design an optimum filter for any given signal corrupted by noise with known statistics. There are three main factors in designing optimum systems:

1. Specifications of the input or a minimum knowledge of the characteristics of the input. We will assume that the input to the filter consists of a known deterministic signal  $s(t)$  with an additive stationary Gaussian noise  $N(t)$ . The autocorrelation of the noise is  $R_N(\tau)$  with psd  $\Phi_N(\omega)$ :

$$X(t) = s(t) + N(t), \quad E[N(t)N(t + \tau)] = R_N(\tau) \quad (21.3.5)$$

2. The nature of the impulse response  $h(t)$  such as linear time-invariant, linear time-varying, nonlinear, or causal, must be specified. We will consider linear time-invariant impulse responses, not necessarily causal. In this case the transfer function will be given by  $H(\omega)$ .
3. The criterion of optimality has to be specified. Some of the criteria of optimality similar to the criteria of estimation discussed in Section 18.1 are
  - (a) Maximum SNR
  - (b) Minimum mean-square error
  - (c) Bayesian optimality criteria such as maximum a posteriori and maximum likelihood

The criterion of optimality chosen will be maximizing the signal-to-noise ratio (SNR). In the schematic shown in Fig. 21.3.5, the signal  $s(t)$  is corrupted with an additive stationary Gaussian noise  $N(t)$  with autocorrelation function  $R_N(\tau)$ . The input to the system is  $X(t)$ , given by Eq. (21.3.5). We have to find the optimum impulse response  $h(t)$ , not necessarily causal, such that the output SNR  $\gamma_o(t)$  at the time of observation  $T$  is maximized.

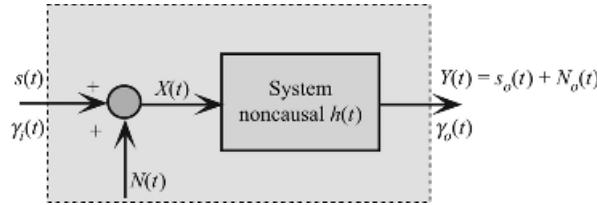


FIGURE 21.3.5

Since the system is linear, the output signal  $Y(t)$  will be given by

$$Y(t) = s_o(t) + N_o(t) \quad (21.3.6)$$

where the output signal  $s_o(t)$  and the output noise  $N_o(t)$  will be convolutions of the noncausal impulse response  $h(t)$  with the input signal  $s(t)$  and the input noise  $N(t)$ . Or

$$s_o(t) = \int_{-\infty}^{\infty} s(\tau)h(t-\tau)d\tau : N_o(t) = \int_{-\infty}^{\infty} N(\tau)h(t-\tau)d\tau \quad (21.3.7)$$

The instantaneous input and output SNRs  $\gamma_i(t)$  and  $\gamma_o(t)$  can be defined as follows:

$$\gamma_i(t) = \frac{|s^2(t)|}{E[N^2(t)]} = \frac{|s^2(t)|}{R_N(0)}; \quad \gamma_o(t) = \frac{|s_o^2(t)|}{E[N_o^2(t)]} = \frac{|s_o^2(t)|}{R_{N_o}(0)} \quad (21.3.8)$$

Since  $S_o(\omega) = S(\omega)H(\omega)$ , we can express  $s_o(t)$  as

$$s_o(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega)H(\omega)e^{j\omega t}d\omega \quad (21.3.9)$$

From Eq. (21.3.9) we can relate the output noise psd  $\Phi_{N_o}(\omega)$  to the input noise psd  $\Phi_N(\omega)$  as

$$\Phi_{N_o}(\omega) = |H(\omega)|^2\Phi_N(\omega) \quad (21.3.10)$$

We can express  $R_{N_o}(0)$  in terms of Eq. (21.3.10) as

$$R_N(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2\Phi_N(\omega)e^{j\omega\tau}d\omega \Big|_{\tau=0} = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2\Phi_N(\omega)d\omega \quad (21.3.11)$$

Substituting Eqs. (21.3.10) and (21.3.11) in the output SNR  $\gamma_o(t)$  in Eq. (21.3.8), we have

$$\gamma_o(t) = \frac{\left| \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega)H(\omega)e^{j\omega t}d\omega \right|^2}{\frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2\Phi_N(\omega)d\omega} = \frac{1}{2\pi} \frac{\left| \int_{-\infty}^{\infty} S(\omega)H(\omega)e^{j\omega t}d\omega \right|^2}{\int_{-\infty}^{\infty} |H(\omega)|^2\Phi_N(\omega)d\omega} \quad (21.3.12)$$

The problem now is to find the transfer function  $H(\omega)$  in Eq. (21.3.12) that maximizes the output SNR  $\gamma_o(t)$  without regard to causality. We will address the question of causality later. The Schwarz inequality [Eq. (14.5.2)] modified in the frequency form given by Eq. (21.3.13) will be used to solve for  $H(\omega)$ :

$$\left| \int_{-\infty}^{\infty} A(\omega)B^*(\omega)d\omega \right|^2 \leq \int_{-\infty}^{\infty} |A(\omega)|^2d\omega \cdot \int_{-\infty}^{\infty} |B(\omega)|^2d\omega \quad (21.3.13)$$

with equality holding when  $A(\omega) = KB(\omega)$ . We define the following quantities:

$$A(\omega) = H(\omega)\sqrt{\Phi_N(\omega)} \quad \text{and} \quad B^*(\omega) = \frac{S(\omega)e^{j\omega t}}{\sqrt{\Phi_N(\omega)}} \quad (21.3.14)$$

Substituting Eq. (21.3.14) in Eq. (21.3.13), Schwartz' inequality becomes

$$\left| \int_{-\infty}^{\infty} H(\omega)S(\omega)e^{j\omega t}d\omega \right|^2 \leq \int_{-\infty}^{\infty} H^2(\omega)\Phi_N(\omega)\Phi_N(\omega)d\omega \cdot \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{\Phi_N(\omega)}d\omega$$

or

$$\frac{\left| \int_{-\infty}^{\infty} H(\omega)S(\omega)e^{j\omega t}d\omega \right|^2}{\int_{-\infty}^{\infty} H^2(\omega)\Phi_N(\omega)d\omega} \leq \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{\Phi_N(\omega)}d\omega \quad (21.3.15)$$

Substituting Eq. (21.3.15) in Eq. (21.3.12), we obtain

$$\gamma_o(t) = \frac{1}{2\pi} \frac{\left| \int_{-\infty}^{\infty} S(\omega)H(\omega)e^{j\omega t}d\omega \right|^2}{\int_{-\infty}^{\infty} |H(\omega)|^2\Phi_N(\omega)d\omega} \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{\Phi_N(\omega)}d\omega \quad (21.3.16)$$

The equality holds when  $A(\omega) = KB(\omega)$ , or

$$H(\omega)\sqrt{\Phi_N(\omega)} = K \frac{S^*(\omega)e^{-j\omega t}}{\sqrt{\Phi_N^*(\omega)}} \quad \text{and} \quad H_{\text{opt}}(\omega) = K \frac{S^*(\omega)e^{-j\omega t}}{|\Phi_N(\omega)|} \quad (21.3.17)$$

If  $T$  is the time of observation, then the optimum transfer function will be given by

$$H_{\text{opt}}(\omega) = K \frac{S^*(\omega)e^{-j\omega T}}{|\Phi_N(\omega)|} \quad (21.3.18)$$

where  $K$  is an arbitrary real gain constant. Since  $H_{\text{opt}}(\omega)$  is matched to the input signal, it is called a *matched filter*. Under these optimum conditions the maximum output SNR  $\gamma_{o,\text{max}}$  for the time of observation  $T$  becomes

$$\gamma_{o,\text{max}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{\Phi_N(\omega)}d\omega \quad (21.3.19)$$

Simplifications in Eq. (21.3.18) can result if the noise  $N(t)$  is white with psd  $\Phi_N(\omega) = N_0/2$  and the gain constant is adjusted to be equal to  $K = N_0/2$ . In this case the optimum transfer function is given by

$$H_{\text{opt}}(\omega) = S^*(\omega)e^{-j\omega T} \quad (21.3.20)$$

The optimum impulse response is the IFT of  $H_{\text{opt}}(\omega)$  and is given by

$$h_{\text{opt}}(t) = s(T - t) \quad (21.3.21)$$

This equation shows clearly that the optimum output impulse response  $h_{\text{opt}}(t)$  is matched to the input signal  $s(t)$ .

The autocorrelation of the signal  $s(t)$  is defined by

$$R_s(t) = s(t)^*s(-t) = \int_{-\infty}^{\infty} s(\tau - t)s(\tau)d\tau \quad (12.3.22)$$

Taking FT on both sides of Eq. (12.3.22), we obtain

$$\text{FT}[R_s(t)] = \Phi_s(\omega) = S(\omega)S^*(\omega) = |S(\omega)|^2 \quad (12.3.23)$$

$|S(\omega)|^2$  is called the *energy spectral density*. The autocorrelation function  $R_s(t)$  can be given as the IFT of  $|S(\omega)|^2$ :

$$R_s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |S(\omega)|^2 e^{j\omega t} d\omega \quad (12.3.24)$$

The optimum output signal spectrum  $S_o(\omega)$  under the condition  $K = N_0/2$  is given by the product

$$S_o(\omega) = S(\omega)H_{\text{opt}}(\omega) = S(\omega)S^*(\omega)e^{-j\omega T} = |S(\omega)|^2 e^{-j\omega T} \quad (21.3.25)$$

Taking the IFT of  $S_o(\omega)$ , the optimum output signal  $s_o(t)$  is given by

$$s_o(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |S(\omega)|^2 e^{-j\omega T} e^{j\omega t} d\omega = R_s(t - T), \quad t \geq 0 \quad (21.3.26)$$

Thus, the optimum output signal  $s_o(t)$  under the conditions of additive white noise is the autocorrelation of the input signal  $R_s(t)$  shifted by the time of observation  $T$ .

The corresponding maximum SNR is given by

$$\gamma_{o,\text{max}} = \frac{1}{2\pi} \frac{2}{N_0} \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega \quad (21.3.27)$$

Using Parseval's theorem  $(1/2\pi) \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega = \int_{-\infty}^{\infty} s^2(t) dt$  in Eq. (21.3.27), we obtain

$$\gamma_{o,\text{max}} = \frac{2}{N_0} \int_{-\infty}^{\infty} s^2(t) dt = \frac{2E_s}{N_0} \quad (21.3.28)$$

where  $E_s$  is the input energy.

### Summary of Matched-Filters Properties

Some of the properties of matched filters under the assumptions that (1) the additive noise  $N(t)$  is white with psd  $S_N(\omega) = N_0/2$  and (2) the gain constant  $K = N_0/2$  can be summarized as follows:

1. The Fourier spectrum  $S_o(\omega)$  of the output signal  $s_o(t)$  is equal to the energy spectral density  $|S(\omega)|^2$  multiplied by a delay factor  $e^{-j\omega T}$  of the input signal  $s(t)$ . This can be shown from Eq. (21.3.20) as follows:

$$S_o(\omega) = S(\omega)H_{\text{opt}}(\omega) = S(\omega)S^*(\omega)e^{-j\omega T} = |S(\omega)|^2 e^{-j\omega T} \quad (21.3.29)$$

2. The output response  $s_o(t)$  is equal to the autocorrelation  $R_s(t)$  of the input signal shifted in time by  $T$  as follow:

$$s_o(t) = R_s(t - T), \quad t \geq 0 \quad (21.3.26)$$

3. From Eq. (21.3.28), the optimum SNR  $\gamma_{o,\text{max}}$  is

$$\gamma_{o,\text{max}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\Phi_s(\omega)}{\Phi_N(\omega)} d\omega = \frac{1}{2\pi} \frac{2}{N_0} \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_s(\omega) d\omega = \frac{2E_s}{N_0} \quad (21.3.30)$$

4. *Causality*: If the original signal  $s(t)$  is time limited to  $(0, T)$ , then the matched filter given by  $s(T - t)$  is causal.

In the following three examples (Examples 21.3.2–21.3.4) the additive noise  $N(t)$  is white with psd  $S_N(\omega) = N_0/2$  and the gain constant  $K = N_0/2$ .

**Example 21.3.2 (Rectangular Pulse)** The signal  $s(t)$  is a rectangular pulse defined by

$$s(t) = \begin{cases} A, & 0 < t < T \\ 0, & \text{otherwise} \end{cases}$$

From the tables, the FT of the signal  $s(t)$  is given by

$$S(\omega) = AT \frac{\sin(\omega T/2)}{(\omega T/2)} e^{-j\omega T/2}$$

Hence, from Eq. (21.3.20) the optimum transfer function is

$$H_{\text{opt}}(\omega) = AT \frac{\sin(\omega T/2)}{(\omega T/2)} e^{j\omega T/2} e^{-j\omega T} = AT \frac{\sin(\omega T/2)}{(\omega T/2)} e^{-j\omega T/2}$$

and the impulse response  $h_{\text{opt}}(t) = s(T - t) = s(t)$  and the matched filter are the same as the input signal.

The autocorrelation function of the input from Eq. (12.3.22) is

$$R_s(t) = \begin{cases} A^2(T + t), & -T < t \leq 0 \\ A^2(T - t), & 0 < t \leq T \end{cases}$$

The output  $s_o(t)$  is obtained from Eq. (12.3.26) as follows:

$$R_s(t - T) = \begin{cases} A^2t, & 0 < t \leq T \\ A^2(T - t), & T < t \leq 2T \end{cases}$$

From Eq. (21.3.26), the maximum output SNR is

$$\gamma_{o,\text{max}} = \frac{2}{N_0} \int_{-\infty}^{\infty} s^2(t) dt = \frac{2}{N_0} \int_0^T A^2 dt = \frac{2A^2 T}{N_0}$$

**Example 21.3.3 (Sinusoidal Pulse)** The input signal  $s(t)$  is a time-limited sinusoidal pulse with energy  $E_s = 1$  as given below:

$$s(t) = \begin{cases} \sqrt{\frac{2}{T}} \sin(\omega_0 t), & 0 \leq t \leq T, \quad \omega_0 = \frac{2\pi n}{T} \\ 0, & \text{otherwise} \end{cases}$$

The signal  $s(t)$  is shown in Fig.(21.3.6a). It is simpler to approach the problem from the time domain. From Eq. (21.3.21) the optimum impulse response is given by

$$h_{\text{opt}}(t) = s(T - t) = \begin{cases} \sqrt{\frac{2}{T}} \sin[\omega_0(T - t)], & 0 \leq t \leq T, \quad \omega_0 = \frac{2\pi n}{T} \\ 0, & \text{otherwise} \end{cases}$$

The optimum output signal  $s_o(t)$  is given by the convolution integral

$$s_o(t) = s(t)^* s(T - t) = \int_{-\infty}^{\infty} s(\tau) s(T - t + \tau) d\tau$$

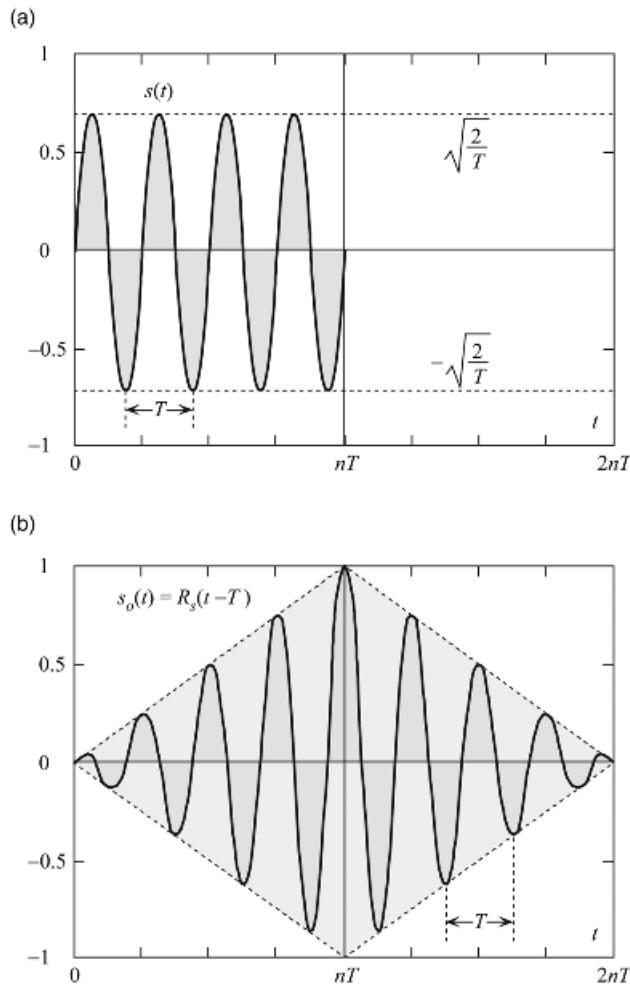


FIGURE 21.3.6

The convolution integral can be written as follows:

$$s_o(t) = \begin{cases} \frac{2}{T} \int_0^t \sin(\omega_0 \tau) \sin[\omega_0(T-t+\tau)d\tau], & 0 \leq t \leq T \\ \frac{2}{T} \int_{t-T}^T \sin(\omega_0 \tau) \sin[\omega_0(T-t+\tau)d\tau], & T \leq t \leq 2T \\ 0, & \text{otherwise} \end{cases}$$

The integrals are evaluated, and the results are

$$s_o(t) = \begin{cases} \frac{1}{\omega_0 T} [2\omega_0 t \cos(\omega_0 t) - \sin(\omega_0 t)], & 0 \leq t \leq T \\ \frac{1}{\omega_0 T} [\omega_0(2T-t) \cos(\omega_0 t) + \sin(\omega_0 t)], & T \leq t \leq 2T \\ 0, & \text{otherwise} \end{cases}$$

The output signal  $s_o(t)$  is shown in Fig. (21.3.6b).

It is also instructive to evaluate the output  $s_o(t)$  as the shifted version of the autocorrelation of the input signal  $s(t)$  as given in Eq. (21.3.26). The autocorrelation  $R_s(t)$  of the input  $s(t)$  is given by the following integrals:

$$R_s(t) = \begin{cases} \frac{2}{T} \int_0^{t+T} \sin[(\omega_0(\tau - t)] \sin(\omega_0\tau) d\tau, & -T \leq t \leq 0 \\ \frac{2}{T} \int_t^T \sin[\omega_0(\tau - t)] \sin(\omega_0\tau) d\tau, & 0 \leq t \leq T \\ 0, & \text{otherwise} \end{cases}$$

In this example the integrand in the integral shown above is the same as those in  $s_o(t)$  except that the limits are different. Evaluation of these integrals results in

$$R_s(t) = \begin{cases} \frac{1}{\omega_0 T} [\omega_0(T+t) \cos(\omega_0 t) - \sin(\omega_0 t)], & -T \leq t \leq 0 \\ \frac{1}{\omega_0 T} [\omega_0(T-t) \cos(\omega_0 t) + \sin(\omega_0 t)], & 0 \leq t \leq T \\ 0, & \text{otherwise} \end{cases}$$

and indeed  $R_s(t-T) = s_o(t)$  and its graph also superpose the graph of  $s_o(t)$  exactly in Fig. 21.3.6, thus validating property 2 of Eq. (21.3.26).

**Example 21.3.4** In this example, the input  $s(t)$  is a more complicated time-limited signal, given by

$$s(t) = \begin{cases} e^{-t} \sin\left(\frac{\pi t}{2}\right), & 0 \leq t \leq 4 \\ 0, & \text{otherwise} \end{cases}$$

The psd of noise  $N(t)$  is given by  $N_0/2 = 2$ . The time of observation  $T = 4$

The FT of  $s(t)$  is given by

$$S(\omega) = \frac{2\pi}{4(1+j\omega)^2 + \pi^2} [1 - e^{-4(1+j\omega)}]$$

From Eq. (21.3.20) the optimum filter  $H_{\text{opt}}(\omega)$  is given by

$$H_{\text{opt}}(\omega) = S^*(\omega) e^{-j\omega T} = \frac{2\pi}{4(1-j\omega)^2 + \pi^2} [1 - e^{-4(1-j\omega)}] e^{-4j\omega}$$

From Eq. (21.3.25) the output spectrum  $S_o(\omega)$  is given by

$$S_o(\omega) = |S(\omega)|^2 e^{-j\omega T} = \frac{4\pi^2 [1 - 2e^{-4} \cos(4\omega) + e^{-8}]}{16(1+\omega^2)^2 + 8\pi^2(1-\omega^2) + \pi^4} e^{-4j\omega}$$

The amplitude spectra  $|S(\omega)|$  and  $|H_{\text{opt}}(\omega)|$  are the same and are shown in Fig. 21.3.7 along with the amplitude spectrum of  $|S_o(\omega)|$ .

The impulse response  $h_{\text{opt}}(t)$  corresponding to the transfer function  $H_{\text{opt}}(\omega)$  can be obtained from Eq. (21.3.21) as

$$\begin{aligned} h_{\text{opt}}(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S^*(\omega) e^{-j\omega(4-t)} d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S^*(\omega) e^{j\omega(4-t)} d\omega = s(4-t) = e^{-(4-t)} \sin\left[\frac{\pi(4-t)}{2}\right], \quad 0 \leq t \leq 4 \end{aligned}$$

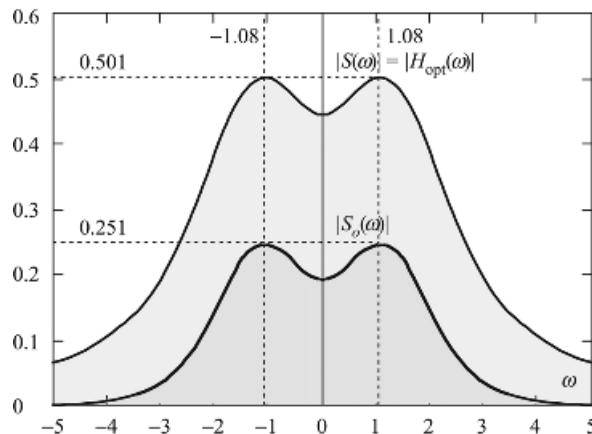


FIGURE 21.3.7

It is easier to obtain the output signal  $s_o(t)$  from the shifted version of the autocorrelation  $R_s(t)$  of the input  $s(t)$  rather than from the IFT of  $S_o(\omega)$ .  $R_s(t)$  can be obtained from the following integrals:

$$R_s(t) = \begin{cases} \int_0^{T+t} e^{-2\pi\tau+t} \sin\left(\frac{\pi\tau}{2}\right) \sin\left(\frac{\pi(\tau-t)}{2}\right) d\tau, & -T \leq t \leq 0 \\ \int_t^T e^{-2\pi\tau+t} \sin\left(\frac{\pi\tau}{2}\right) \sin\left(\frac{\pi(\tau-t)}{2}\right) d\tau, & 0 \leq t \leq T \\ 0, & \text{otherwise} \end{cases}$$

After some involved algebra, these integrals can be evaluated as

$$R_s(t) = \frac{-\pi}{4(4 + \pi^2)} \begin{cases} 2 \sin\left(\frac{\pi t}{2}\right) [e^{-(8+t)} + e^t] + \pi \cos\left(\frac{\pi t}{2}\right) [e^{-(8+t)} - e^t], & -4 \leq t \leq 0 \\ 2 \sin\left(\frac{\pi t}{2}\right) [e^{-(8-t)} + e^{-t}] - \pi \cos\left(\frac{\pi t}{2}\right) [e^{-(8-t)} - e^{-t}], & -0 \leq t \leq 4 \\ 0, & \text{otherwise} \end{cases}$$

and the output signal  $s_o(t)$  can be evaluated from the convolution integral or by shifting the autocorrelation  $R_s(t)$  by  $T = 4$ :

$$s_o(t) = s(t) * s(4 - t) = R_s(t - 4)$$

The signal  $s(t)$ , the optimum impulse response  $h_{\text{opt}}(t)$ , the autocorrelation  $R_s(t)$ , and the output signal  $s_o(t)$  are shown in Fig. 21.3.8.

The maximum SNR from Eq. (21.3.19) is

$$\gamma_{o,\max} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{S_N(\omega)} d\omega = \frac{1}{2\pi N_0} \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega = \frac{2}{N_0} \int_0^T s^2(t) dt = \frac{2E_s}{N_0}$$

The input energy  $E_s = \int_0^T s^2(t) dt = 0.17784$ , and with  $N_0/2 = 2$ , the maximum SNR is  $\gamma_{o,\max} = 0.0889$ .

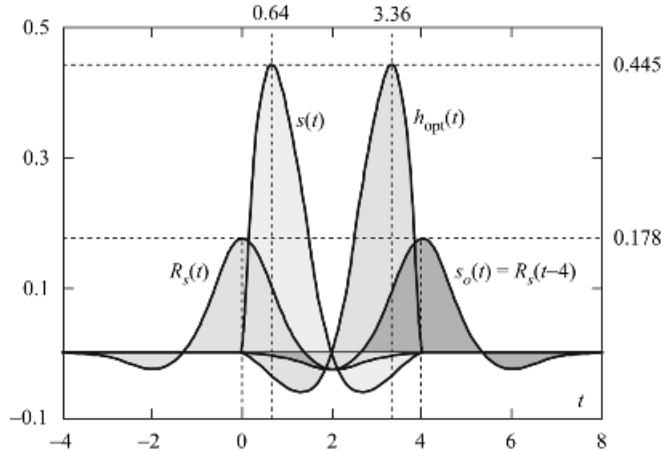


FIGURE 21.3.8

**Example 21.3.5 (Nonwhite Gaussian Noise)** In this example the signal  $s(t)$  and the psd  $\Phi_N(\omega)$  of the nonwhite stationary noise  $N(t)$  are given by

$$s(t) = \begin{cases} \frac{3}{4}(e^{-\frac{t}{2}} - e^{-\frac{3t}{2}}), & t \geq 0 \\ 0, & t < 0 \end{cases} \quad \Phi_N(\omega) = \frac{4}{1 + 4\omega^2}$$

The time of observation  $T$  is equal to 5, and the dc value of the optimum transfer function  $H_{\text{opt}}(0) = 1$ . The Fourier spectrum of  $s(t)$  can be obtained from the tables as

$$S(\omega) = \frac{3}{(1 + 2j\omega)(3 + 2j\omega)} = \frac{3}{3 - 4\omega^2 + 8j\omega}$$

The transfer function  $H_{\text{opt}}(\omega)$  of the matched filter is found from Eq. (21.3.18) as

$$H_{\text{opt}}(\omega) = K \frac{S^*(\omega)e^{-j\omega T}}{|\Phi_N(\omega)|} = K \frac{3e^{-j\omega 5}}{3 - 4\omega^2 - 8j\omega} \frac{4 + \omega^2}{4}$$

Since  $H(0) = 1$ , we obtain  $K = 1$ , and  $H_{\text{opt}}(\omega)$  is

$$H_{\text{opt}}(\omega) = \frac{3(4 + \omega^2)e^{-j\omega 5}}{4(3 - 4\omega^2 - 8j\omega)}$$

The Fourier spectrum  $S_o(\omega)$  of the optimum output signal  $s_o(t)$  is given by

$$S_o(\omega) = S(\omega)H_{\text{opt}}(\omega) = \frac{|S(\omega)|^2 e^{-j\omega 5}}{\Phi_N(\omega)} = \frac{9(4 + \omega^2)e^{-5j\omega}}{4(16\omega^4 + 40\omega^2 + 9)}$$

The amplitude spectra  $|S(\omega)|$ ,  $|H_{\text{opt}}(\omega)|$ , and  $|S_o(\omega)|$  are shown in Fig. 21.3.9.

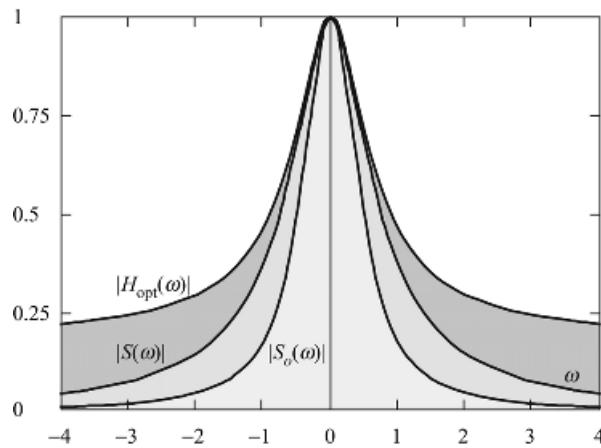


FIGURE 21.3.9

The optimum impulse response  $h_{\text{opt}}(t)$  and the output time function  $s_o(t)$  are obtained from the IFTs of their respective Fourier spectra  $H_{\text{opt}}(\omega)$  and  $S_o(\omega)$  and are given by

$$h_{\text{opt}}(t) = \left[ \frac{45}{64} e^{(t-5)/2} - \frac{21}{64} e^{3(t-5)/2} \right] u(5-t) - \frac{3}{16} \delta(5-t)$$

$$s_o(t) = \left[ \frac{135}{512} e^{(t-5)/2} - \frac{21}{512} e^{3(t-5)/2} \right] u(5-t) + \left[ \frac{135}{512} e^{(5-t)/2} - \frac{21}{512} e^{3(5-t)/2} \right] u(t-5)$$

The input signal  $s(t)$ , the impulse response  $h_{\text{opt}}(t)$ , and the output time function  $s_o(t)$  are shown in Fig. 21.3.10.

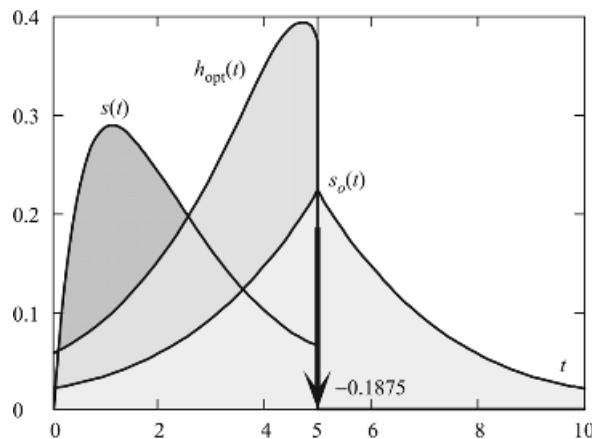


FIGURE 21.3.10

The optimum SNR is obtained from Eq. (21.3.19) and is given by

$$\begin{aligned}\gamma_{o, \max} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|S(\omega)|^2}{\Phi_N(\omega)} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{9(4 + \omega^2)}{4(16\omega^4 + 40\omega^2 + 9)} d\omega \\ &= \frac{1}{2\pi} \left[ \frac{135}{128} \tan^{-1}(\infty) - \frac{21}{128} \tan^{-1}(\infty) \right] = \frac{57}{256}\end{aligned}$$

## 21.4. BANDPASS STATIONARY RANDOM PROCESSES

### Review of Bandpass Signals

Bandpass signals form the backbone of signal transmission. The channel through which the signal is transmitted is bandwidth-limited to a band of frequencies centered about the carrier. If the bandwidth  $2\omega_B$  is much smaller than the carrier frequency  $\omega_c$ , then such signals are called *narrowband signals*. They can be represented in three ways as

1. *Natural Envelope Representation*. This is the usual representation of a narrowband process and is given by

$$s(t) = a(t) \cos[\omega_c t + \theta(t)] \quad (21.4.1)$$

where  $a(t)$  (called the *natural envelope*) is a lowpass signal bandlimited to  $2\omega_B$ ,  $\omega_c$  is the carrier frequency, and  $\theta(t)$  is the phase angle. If the Fourier spectrum of  $a(t)$  is  $A(\omega)$ , then the spectrum of the signal  $S(\omega)$  is given by

$$\begin{aligned}S(\omega) &= \frac{1}{2\pi} A(\omega) * \pi[\delta(\omega + \omega_c) + \delta(\omega - \omega_c)] \\ &= \frac{1}{2} [A(\omega + \omega_c) + A(\omega - \omega_c)] \quad (21.4.2)\end{aligned}$$

The amplitude spectra of  $A(\omega)$  and  $S(\omega)$  are shown in Fig. 21.4.1.

In Fig. 21.4.1  $2\omega_B < \omega_c$ , and hence the signal  $s(t)$  is bandpass. The band of frequencies centered about  $\omega_c$  is called the *upper sideband*, and the band centered about  $-\omega_c$  is called

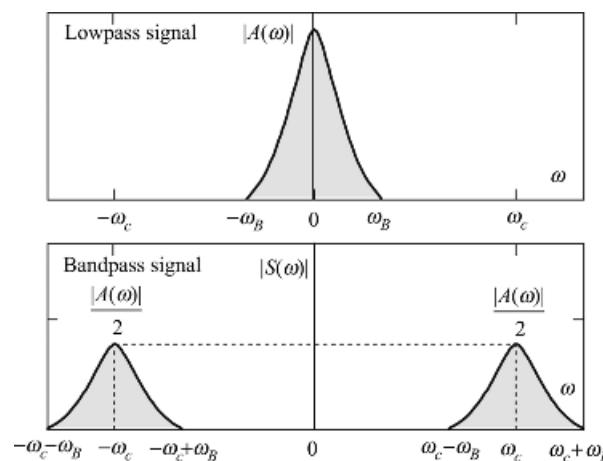


FIGURE 21.4.1

the *lower sideband*. Each of the sidebands carries the same information as the original lowpass signal  $s(t)$ .

**2. Quadrature Representation.** The signal  $s(t)$  in Eq. (21.4.1) can be expanded and written as

$$\begin{aligned} s(t) &= a(t) \cos[\omega_c t + \theta(t)] \\ &= a(t)[\cos(\omega_c t) \cos \theta(t) - \sin(\omega_c t) \sin \theta(t)] \end{aligned} \quad (21.4.3)$$

In Eq. (21.4.3) we can define the *quadrature components*  $s_c(t)$  and  $s_s(t)$  as

$$\begin{aligned} s_c(t) &= a(t) \cos \theta(t), \quad \text{in phase} \\ s_s(t) &= a(t) \sin \theta(t), \quad \text{quadrature} \end{aligned} \quad (21.4.4)$$

and write Eq. (21.4.3) as

$$s(t) = s_c(t) \cos \theta(t) - s_s(t) \sin \theta(t) \quad (21.4.5)$$

The envelope  $a(t)$  and the phase  $\theta(t)$  are obtained from the quadrature components as

$$a(t) = \sqrt{s_c^2(t) + s_s^2(t)}; \quad \theta(t) = \tan^{-1}\left(\frac{s_s(t)}{s_c(t)}\right) \quad (21.4.6)$$

Since the signal is narrowband  $a(t)$  varies slowly compared to  $t_c = 2\pi/\omega_c$ .

**3. Analytic Signal Representation.** Since both sidebands in a bandpass signal carry the same information, we should be able to transform the spectrum into positive frequencies. This is accomplished by using Hilbert transforms (HTs), [36]. The Hilbert transform  $\hat{s}(t)$  of any signal  $s(t)$  is a time-domain transformation, defined by

$$\hat{s}(t) = s(t)^* \frac{1}{\pi t} = \int_{-\infty}^{\infty} \frac{s(\tau)}{\pi(t-\tau)} d\tau \quad (21.4.7)$$

under the assumptions, that (a)  $s(t)$  is square integrable and (b) the FT  $S(\omega)$  is minimum phase.

The improper integral in Eq. (21.4.7) is evaluated in the sense of Cauchy principal value, defined by

$$\int_{-\infty}^{\infty} \frac{s(\tau)}{\pi(t-\tau)} d\tau = \lim_{\varepsilon \rightarrow 0} \left\{ \int_{-\infty}^{t-\varepsilon} \frac{s(\tau)}{\pi(t-\tau)} d\tau + \int_{t+\varepsilon}^{\infty} \frac{s(\tau)}{\pi(t-\tau)} d\tau \right\}$$

The inverse Hilbert transform (IHT) is defined by

$$s(t) = -\hat{s}(t) * \frac{1}{\pi t} = \int_{-\infty}^{\infty} \frac{\hat{s}(\tau)}{\pi(t-\tau)} d\tau \quad (21.4.8)$$

Since  $(1/\pi t) \Leftrightarrow -j\text{sgn}(\omega)$ , the FT of  $\hat{s}(t)$  is obtained by using the convolution property as

$$\text{FT}\{\hat{s}(t)\} = \hat{S}(\omega) = -S(\omega)j \text{ sgn}(\omega) \quad (21.4.9)$$

We can also define the HT of the Fourier transform of  $S(\omega)$  as,

$$\text{FT}\{\hat{S}(\omega)\} = \hat{S}(\omega) = S(\omega) * \frac{1}{\pi\omega} = \int_{-\infty}^{\infty} \frac{S(p)}{\pi(\omega-p)} dp \quad (21.4.10)$$

In general,  $\hat{S}(\omega) \neq \hat{\hat{S}}(\omega)$ , but if  $s(t)$  is a real causal time function, then it can be shown that

$$\hat{S}(\omega) = -\hat{\hat{S}}(\omega)\text{sgn}(\omega) \quad (21.4.11)$$

The analytic signal  $z_s(t)$  corresponding to  $s(t)$  is defined as

$$z_s(t) = s(t) + j\hat{s}(t) \quad (21.4.12)$$

The Fourier spectrum of the analytic signal  $z_s(t)$  can be obtained by using Eq. (21.4.9):

$$Z_s(\omega) = S(\omega + j(-j)S(\omega)\text{sgn}(\omega)) = 2S(\omega)u(\omega) \quad (21.4.13)$$

Thus the spectrum  $Z_s(\omega)$  is causal in the frequency domain.

In most problems it may be easier to find the HT from the frequency representation of Eq. (21.4.9) rather than the convolution integral of Eq. (21.4.7).

**Example 21.4.1** We will find the Hilbert transforms for the following functions:

(a)  $x_c(t) = a(t)\cos(\omega_c t)$ : The FT of  $x_c(t)$  is

$$X_c(\omega) = \frac{1}{2\pi}A(\omega) * \pi[\delta(\omega + \omega_c) + \delta(\omega - \omega_c)] = \frac{1}{2}[A(\omega + \omega_c) + A(\omega - \omega_c)]$$

and from Eq. (21.4.9)

$$\hat{X}_c(\omega) = \frac{1}{2}[A(\omega + \omega_c) + A(\omega - \omega_c)][-j\text{sgn}(\omega)] = \frac{j}{2}[A(\omega + \omega_c) - A(\omega - \omega_c)]$$

Hence

$$\hat{x}_c(t) = a(t)\sin(\omega_c t)$$

(b)  $x_s(t) = a(t)\sin(\omega_c t)$ : The FT of  $x_s(t)$  is

$$X_s(\omega) = \frac{1}{2\pi}A(\omega) * j\pi[\delta(\omega + \omega_c) - \delta(\omega - \omega_c)] = \frac{j}{2}[A(\omega + \omega_c) - A(\omega - \omega_c)]$$

and from Eq. (21.4.9)

$$\hat{X}_s(\omega) = \frac{j}{2}[A(\omega + \omega_c) - A(\omega - \omega_c)][-j\text{sgn}(\omega)] = -\frac{1}{2}[A(\omega + \omega_c) + A(\omega - \omega_c)]$$

Hence

$$\hat{x}_s(t) = -a(t)\cos(\omega_c t)$$

(c)  $y(t) = \sin(\omega_B t)\sin(\omega_c t)$ ,  $\omega_B \ll \omega_c$ : We will solve this problem from basics and show that the HT  $\hat{y}(t)$  follows directly from the previous example (b). The

Fourier spectrum  $Y(\omega)$  is obtained from the tables as

$$\begin{aligned}
 Y(\omega) &= \frac{1}{2\pi} \{ j\pi(\omega + \omega_B) - \delta[\omega - \omega_B] \} * \{ j\pi[\delta(\omega + \omega_c) - \delta(\omega - \omega_c)] \} \\
 &= -\frac{\pi}{2} \{ \delta[\omega + \omega_B + \omega_c] - \delta[\omega + \omega_B - \omega_c] \\
 &\quad - \delta[\omega - \omega_B + \omega_c] + \delta[\omega - \omega_B - \omega_c] \} \\
 &= -\frac{\pi}{2} \{ \delta[\omega + (\omega_c + \omega_B)] + \delta[\omega - (\omega_c + \omega_B)] \} \\
 &\quad + \frac{\pi}{2} \{ \delta[\omega + (\omega_c - \omega_B)] + \delta[\omega - (\omega_c - \omega_B)] \}
 \end{aligned}$$

From Eq. (21.4.9) the FT  $\hat{Y}(\omega)$  is given by

$$\begin{aligned}
 \hat{Y}(\omega) &= Y(\omega)(-j \operatorname{sgn}(\omega)) \\
 &= -\frac{j\pi}{2} \{ \delta[\omega + (\omega_c + \omega_B)] - \delta[\omega - (\omega_c + \omega_B)] \} \\
 &= \frac{j\pi}{2} \{ \delta[\omega + (\omega_c - \omega_B)] - \delta[\omega - (\omega_c - \omega_B)] \}
 \end{aligned}$$

The functions  $Y(\omega)$ ,  $-j \operatorname{sgn}(\omega)$ , and  $\hat{Y}(\omega)$  are shown in Fig. 21.4.2

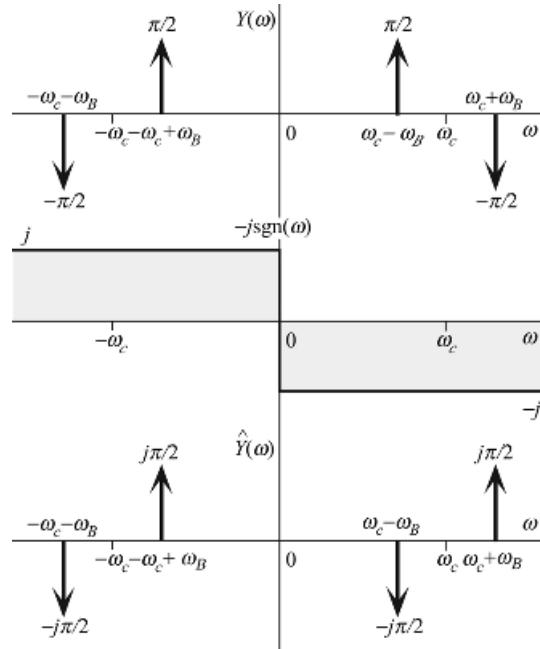


FIGURE 21.4.2

**TABLE 21.4.1 Table of Hilbert transforms**

No.	Time Function $s(t)$	Hilbert Transform $\hat{s}(t)$	Comments
1.	$\cos(\omega_c t)$	$\sin(\omega_c t)$	
2.	$\sin(\omega_c t)$	$-\cos(\omega_c t)$	
3.	$a(t)\cos(\omega_c t)$	$a(t)\sin(\omega_c t)$	$a(t)$ bandlimited to $2\omega_B, <\omega_c$
4.	$a(t)\sin(\omega_c t)$	$-a(t)\cos(\omega_c t)$	$a(t)$ bandlimited to $2\omega_B, <\omega_c$
5.	$\frac{\sin(\omega_c t)}{\omega_c t}$	$\frac{1 - \cos(\omega_c t)}{\omega_c t}$	
6.	$\frac{1}{1 + t^2}$	$\frac{t}{1 + t^2}$	
7.	$u\left(t + \frac{1}{2}\right) - u\left(t - \frac{1}{2}\right)$	$\frac{1}{\pi} \ln\left(\frac{t + 1/2}{t - 1/2}\right)$	
8.	$\delta(t)$	$\frac{1}{\pi t}$	
9.	$\frac{1}{t}$	$-\pi\delta(t)$	

Hence the HT of  $y(t)$  is

$$\begin{aligned}\hat{y}(t) &= \frac{1}{2} \{ \sin[(\omega_c - \omega_B)t] - \sin[(\omega_c + \omega_B)t] \} \\ &= -\sin(\omega_B t) \cos(\omega_c t)\end{aligned}$$

A result that can be obtained by substituting  $a(t) = \sin(\omega_B t)$  in example (b) above.

A summary of some Hilbert transforms is given in Table 21.4.1.

Substituting  $s(t) = a(t) \cos[\omega_c t + \theta(t)]$  from Eq. (21.4.1) and  $\hat{s}(t) = -a(t) \sin[\omega_c t + \theta(t)]$  from Example 21.4.1 in Eq. (21.4.12) for the analytic signal, we can write

$$\begin{aligned}z_s(t) &= a(t) \{ \cos[\omega_c t + \theta(t)] - j \sin[\omega_c t + \theta(t)] \} \\ &= a(t) e^{j[\omega_c t + \theta(t)]} = [a(t) e^{j\theta(t)}] e^{j\omega_c t} = \tilde{s}(t) e^{j\omega_c t}\end{aligned}\quad (21.4.14)$$

where  $\tilde{s}(t) = a(t) e^{j\theta(t)}$  is defined as the *complex envelope* of the analytic signal  $z_s(t)$ . The bandpass signal  $s(t)$  can be written in terms of the analytic signal as

$$s(t) = \operatorname{Re}[z_s(t)] = \operatorname{Re}[\tilde{s}(t) e^{j\omega_c t}] = a(t) \cos[\omega_c t + \theta(t)] \quad (21.4.15)$$

Using the relationship  $\operatorname{Re}[z] = \frac{1}{2}[z + z^*]$  we can rewrite Eq. (21.4.15) as

$$s(t) = \frac{1}{2} [s(t) e^{j\omega_c t} + s^*(t) e^{-j\omega_c t}] \quad (21.4.16)$$

and its Fourier spectrum as

$$S(\omega) = \frac{1}{2} [S(\omega - \omega_c) + S^*(\omega + \omega_c)] \quad (21.4.17)$$

The absolute value  $|\tilde{s}(t)|$  of the complex envelope  $\tilde{s}(t)$  is the natural envelope  $a(t)$ , and  $\tilde{s}(t)$  can be written in terms of the quadrature components of Eq. (21.4.4) as

$$\tilde{s}(t) = s_c(t) + j s_s(t) \quad (21.4.18)$$

It should be noted that  $a(t)$ ,  $\tilde{s}(t)$ ,  $s_c(t)$ , and  $s_s(t)$  are all lowpass signals, whereas  $s(t)$  and  $z_s(t)$  are bandpass signals.

**Example 21.4.2** A narrowband bandpass signal  $X(t)$  with bandwidth  $2\omega_B$  is passed through a bandpass filter with bandwidth  $2\omega_B$  and impulse response  $h(t)$ . We will show that the output of this system  $Y(t)$  can be represented by

$$Y(t) = \frac{1}{2} \operatorname{Re}[z_h(t)^* z_X(t)]$$

where  $z_h(t)$  and  $z_X(t)$  are the analytic signals corresponding to  $h(t)$  and  $X(t)$ .

We will now evaluate

$$\begin{aligned} \operatorname{Re} \left[ \int_{-\infty}^{\infty} z_h(t) z_X^*(t) dt \right] &= \operatorname{Re} \int_{-\infty}^{\infty} [h(t) + j\hat{h}(t)] [X(t) - j\hat{X}(t)] dt \\ &= \int_{-\infty}^{\infty} [h(t)X(t) + \hat{h}(t)\hat{X}(t)] dt \end{aligned}$$

By Parseval's theorem, we obtain

$$\begin{aligned} \int_{-\infty}^{\infty} \hat{h}(t)\hat{X}(t) dt &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{H}(\omega) \hat{X}^*(\omega) d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} -j \operatorname{sgn}(\omega) H(\omega) \cdot j \operatorname{sgn}(\omega) X^*(\omega) d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \operatorname{sgn}^2(\omega) H(\omega) \cdot X^*(\omega) d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) \cdot X^*(\omega) d\omega = \int_{-\infty}^{\infty} h(t)X(t) dt \end{aligned}$$

Hence

$$\operatorname{Re} \left[ \int_{-\infty}^{\infty} z_h(t) z_X^*(t) dt \right] = 2 \int_{-\infty}^{\infty} h(t)X(t) dt = 2 \int_{-\infty}^{\infty} \operatorname{Re}[z_h(t)] \operatorname{Re}[z_X(t)] dt \quad (\text{A})$$

We will evaluate the output  $Y(t)$  given by the convolution integral

$$Y(t) = \int_{-\infty}^{\infty} h(\tau)X(t-\tau) d\tau$$

and with  $z_h(t) = h(t) + j\hat{h}(t)$  and  $z_X(t) = X(t) + j\hat{X}(t)$  we have

$$Y(t) = \int_{-\infty}^{\infty} \operatorname{Re}[z_h(\tau)] \operatorname{Re}[z_X(t-\tau)] d\tau$$

Substituting the result (A) derived earlier in this equation, we obtain

$$Y(t) = \frac{1}{2} \operatorname{Re} \int_{-\infty}^{\infty} z_h(\tau) z_X(t-\tau) d\tau = \frac{1}{2} \operatorname{Re}[z_h(t) * z_X(t)]$$

**Example 21.4.3** A pulsed radiofrequency (RF) wave  $x(t) = Ap_{T/2}(t) \cos(\omega_c t)$ , where  $p_{T/2}$  is a pulse of width  $T$  defined by  $p_{T/2} = u(t + T/2) - u(t - T/2)$ , is applied to an ideal bandpass filter shown as in Fig. 21.4.3.

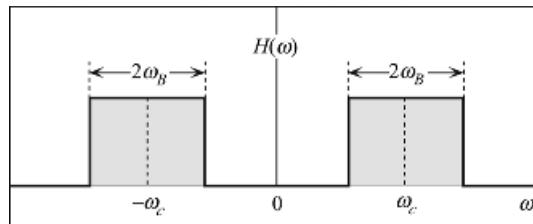


FIGURE 21.4.3

We want to find the output  $y(t)$  using Hilbert transform techniques.

The HT of the input can be obtained from the HT table (Table 21.4.1) as  $\tilde{x}(t) = Ap_{T/2}(t) \sin(\omega_c t)$ , the analytic signal and the complex envelope corresponding to  $x(t)$  are

$$z_x(t) = Ap_{T/2}(t)e^{j\omega_c t}; \quad \tilde{x}(t) = Ap_{T/2}(t)e^{jt \cdot 0}$$

and the FT of  $z_x(t)$  is

$$Z_x(\omega) = AT \frac{\sin[(\omega - \omega_c)T/2]}{(\omega - \omega_c)T/2}$$

The equivalent lowpass filter for  $H(\omega)$  can be obtained by retaining the positive half of  $H(\omega)$  and backshifting it to the origin. Thus  $\tilde{H}(\omega)$  and the corresponding complex impulse response  $\tilde{h}(t)$  are given by

$$\tilde{H}(\omega) = u(\omega + \omega_B) - u(\omega - \omega_B)e^{j\omega 0}; \quad \tilde{h}(t) = \frac{\omega_B \sin(\omega_B t)}{\pi \omega_B t}$$

The complex envelope  $\tilde{y}(t)$  of the output  $y(t)$  is the convolution given by

$$\tilde{y}(t) = \tilde{x}(t) * \tilde{h}(t) = Ap_{T/2}(t) * \frac{\omega_B \sin(\omega_B t)}{\pi \omega_B t} = A \int_{-T/2}^{T/2} \frac{\omega_B \sin(\omega_B t)}{\pi \omega_B t} P_{T/2}(t - \tau) d\tau$$

Using the sine integral  $\text{Si}(x) = \int_0^x (\sin(t)/t) dt$ , the preceding equation can be written as

$$y(t) = \frac{A}{\pi} \left\{ \text{Si} \left[ \omega_B \left( t + \frac{T}{2} \right) \right] - \text{Si} \left[ \omega_B \left( t - \frac{T}{2} \right) \right] \right\}$$

The RF pulse  $x(t)$  is shown in Fig. 21.4.4a, and  $y(t)$  is shown in Fig. 21.4.4b.

### Bandpass Stationary Random Processes

A stationary random process  $N(t)$  is called a *narrowband bandpass process* if the power spectral density (psd)  $S_N(\omega)$  satisfies two conditions:

1.  $S_N(\omega)$  is zero outside a band of frequencies centered around the carrier frequency  $\pm \omega_c$ .
2. The width  $2\omega_B$  of the band of frequencies is very much smaller than the carrier frequency  $\omega_c$ .

These conditions are similar to those of deterministic narrowband signals except that the Fourier spectrum has been replaced by the power spectral density. Similar to the deterministic process, these narrowband stationary processes can also be represented in three ways.

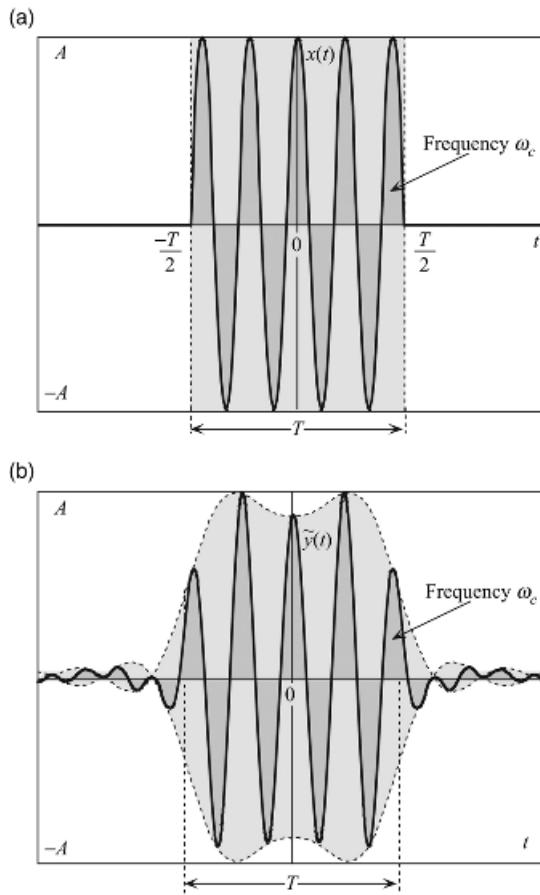


FIGURE 21.4.4

The natural envelope representation is

$$N(t) = W(t) \cos[\omega_c t + \Theta(t)] \quad (21.4.19)$$

where  $W(t)$  is a lowpass stationary random process with psd  $S_W(\omega)$  and bandwidth  $2\omega_B$  and  $\Theta(t)$  is a random phase.

The quadrature representation can be obtained by expanding Eq. (21.4.19) as

$$N(t) = W(t)[\cos(\omega_c t) \cos[\Theta(t)] - \sin(\omega_c t) \sin[\Theta(t)]]$$

and rewriting it as

$$N(t) = N_c(t) \cos(\omega_c t) - N_s(t) \sin(\omega_c t) \quad (21.4.20a)$$

where

$$N_c(t) = W(t) \cos[\Theta(t)]; \quad N_s(t) = W(t) \sin[\Theta(t)] \quad (21.4.20b)$$

are the lowpass stationary quadrature components.

The analytic signal representation is

$$N(t) = \operatorname{Re}[z_N(t)] = \operatorname{Re}[N(t)e^{j\omega_c t}] \quad (21.4.21)$$

where  $z_N(t)$ , the analytic signal corresponding to  $N(t)$ , and  $\tilde{N}(t)$ , the complex envelope in terms of the quadrature components  $N_c(t)$  and  $N_s(t)$ , are given by

$$\begin{aligned} z_N(t) &= N(t) + j\hat{N}(t) \\ \tilde{N}(t) &= N_c(t) + jN_s(t) \end{aligned} \quad (21.4.22)$$

where  $\hat{N}(t)$  is Hilbert transform of  $N(t)$ .

### Quadrature Representation

We will now investigate the quadrature representation of Eq. (21.4.20a). Since  $N(t)$  is stationary, the correlation functions of any two stationary random processes  $N(t)$  and  $M(t)$  are governed by the symmetry conditions

$$\begin{aligned} E[N(t)N(t + \tau)] &= R_N(\tau) = E[N(t + \tau)N(t)] = R_N(-\tau) \\ E[N(t)M(t + \tau)] &= R_{NM}(\tau) = E[M(t + \tau)N(t)] = R_{MN}(-\tau) \end{aligned} \quad (21.4.23)$$

We will determine the autocorrelation functions  $R_N(\tau)$ ,  $R_{N_c}(\tau)$ , and  $R_{N_s}(\tau)$  and the cross-correlation functions  $R_{N_c N_s}(\tau)$  and  $R_{N_s N_c}(\tau)$ :

$$\begin{aligned} R_N(\tau) &= E\{[N_c(t) \cos(\omega_c t) - N_s(t) \sin(\omega_c t)] \\ &\quad \times [N_c(t + \tau) \cos(\omega_c(t + \tau)) - N_s(t + \tau) \sin(\omega_c(t + \tau))]\} \\ &= R_{N_c}(\tau) \cos(\omega_c t) \cos(\omega_c(t + \tau)) + R_{N_s}(\tau) \sin(\omega_c t) \sin(\omega_c(t + \tau)) \\ &\quad - R_{N_c} N_s(\tau) \cos(\omega_c t) \sin(\omega_c(t + \tau)) - R_{N_s} N_c(\tau) \sin(\omega_c t) \cos(\omega_c(t + \tau)) \end{aligned} \quad (21.4.24)$$

Using trigonometric identities, we obtain

$$\begin{aligned} R_N(\tau) &= \frac{1}{2} [R_{N_c}(\tau) + R_{N_s}(\tau)] \cos(\omega_c \tau) \\ &\quad + \frac{1}{2} [R_{N_c}(\tau) - R_{N_s}(\tau)] \cos[\omega_c(2t + \tau)] \\ &\quad - \frac{1}{2} [R_{N_c N_s}(\tau) - R_{N_s N_c}(\tau)] \sin(\omega_c t) \\ &\quad - \frac{1}{2} [R_{N_c N_s}(\tau) + R_{N_s N_c}(\tau)] \sin[\omega_c(2t + \tau)] \end{aligned} \quad (21.4.25)$$

Since there can be no dependence on  $t$  for a stationary process, the following conditions must be satisfied by  $N(t)$  in Eq. (21.4.25):

$$\begin{aligned} R_{N_c}(\tau) &= R_{N_s}(\tau) \\ R_{N_c N_s}(\tau) &= -R_{N_s N_c}(\tau) \end{aligned} \quad (21.4.26)$$

Substituting Eqs. (21.4.6) in Eq. (21.4.25), we obtain the autocorrelation  $R_N(\tau)$  also in quadrature form as

$$R_N(\tau) = R_{N_c}(\tau) \cos(\omega_c \tau) - R_{N_c N_s}(\tau) \sin(\omega_c \tau) \quad (21.4.27)$$

In a similar manner we can also derive the autocorrelation  $\hat{N}(t)$  of the HT of  $N(t)$ , but it is easier to use the HT table (Table 21.4.1) and write

$$R_{\hat{N}}(\tau) = R_{N_c}(\tau) \sin(\omega_c \tau) + R_{N_c N_s}(\tau) \cos(\omega_c \tau) \quad (21.4.28)$$

Equations (21.4.27) and (21.4.28) can be combined as a matrix and written as

$$\begin{bmatrix} R_N(\tau) \\ R_{\hat{N}}(\tau) \end{bmatrix} = \begin{bmatrix} \cos(\omega_c \tau) & -\sin(\omega_c \tau) \\ \sin(\omega_c \tau) & \cos(\omega_c \tau) \end{bmatrix} \begin{bmatrix} R_{N_c}(\tau) \\ R_{N_c N_s}(\tau) \end{bmatrix} \quad (21.4.29)$$

and the correlation functions of quadrature components  $N_c(t)$  and  $N_s(t)$  can be given in terms of the correlation functions of  $N(t)$  and  $\hat{N}(t)$  as

$$\begin{bmatrix} R_{N_c}(\tau) \\ R_{N_c N_s}(\tau) \end{bmatrix} = \begin{bmatrix} \cos(\omega_c \tau) & \sin(\omega_c \tau) \\ -\sin(\omega_c \tau) & \cos(\omega_c \tau) \end{bmatrix} \begin{bmatrix} R_N(\tau) \\ R_{\hat{N}}(\tau) \end{bmatrix} \quad (21.4.30)$$

Equations (21.4.29) and (21.4.30) are orthogonal transformations.

### Analytic Signal Representation

We will investigate the analytic signal representation of the narrowband bandpass stationary random process  $N(t)$ . The analytic signal corresponding to  $N(t)$  is

$$z_N(t) = N(t) + j\hat{N}(t) = \tilde{N}(t)e^{j\omega_c t} \quad (21.4.31)$$

and the complex envelope  $\tilde{N}(t)$  will be given by

$$\tilde{N}(t) = N_c(t) + jN_s(t) = z_N(t)e^{-j\omega_c t} = [N(t) + j\hat{N}(t)]e^{-j\omega_c t} \quad (21.4.32)$$

where  $N_c(t)$  and  $N_s(t)$  are the lowpass quadrature components given by Eq. (21.4.20b). The quadrature components  $N_c(t)$ ,  $N_s(t)$  can be expressed in terms of the components of the analytic signal  $N(t)$ ,  $\hat{N}(t)$  by expanding Eq. (21.4.32) as

$$\begin{aligned} N_c(t) + jN_s(t) &= [N(t) + j\hat{N}(t)][\cos(\omega_c t) - j \sin(\omega_c t)] \\ &= N(t) \cos(\omega_c t) + \hat{N}(t) \sin(\omega_c t) + j[\hat{N}(t) \cos(\omega_c t) - N(t) \sin(\omega_c t)] \end{aligned}$$

and equating real and imaginary parts to obtain, in matrix form

$$\begin{bmatrix} N_c(t) \\ N_s(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega_c t) & \sin(\omega_c t) \\ -\sin(\omega_c t) & \cos(\omega_c t) \end{bmatrix} \begin{bmatrix} N(t) \\ \hat{N}(t) \end{bmatrix} \quad (21.4.33)$$

Inverting the matrix equation [Eq. (21.4.33)] the components of the analytic signal,  $N(t)$ ,  $\hat{N}(t)$  are expressed in terms of the quadrature components

$$\begin{bmatrix} N(t) \\ \hat{N}(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega_c t) & -\sin(\omega_c t) \\ \sin(\omega_c t) & \cos(\omega_c t) \end{bmatrix} \begin{bmatrix} N_c(t) \\ N_s(t) \end{bmatrix} \quad (21.4.34)$$

Equations (21.4.33) and (21.4.34) are again orthogonal transformations.

### Properties of Quadrature Components

Without loss in generality we shall assume that  $N(t)$  is a zero mean stationary narrowband bandpass Gaussian process.

1. Since  $N(t)$  is Gaussian,  $\hat{N}(t)$ , which is linear transformation, is also a zero mean stationary narrowband bandpass Gaussian process.
2. From Eq. (21.4.33),  $N_c(t)$  and  $N_s(t)$  are linear transformations of  $N(t)$  and  $\hat{N}(t)$  and hence they are also stationary Gaussian processes.

3. Because of the stationarity of  $N(t)$ , Eq. (21.4.26) must be satisfied.

$$\begin{aligned} R_{N_c}(\tau) &= R_{N_s}(\tau) \\ R_{N_c N_s}(\tau) &= -R_{N_s N_c}(\tau) \end{aligned} \quad (21.4.26)$$

4. Since  $R_{N_c}(\tau) = R_{N_s}(\tau)$ ,  $N_c(t)$  and  $N_s(t)$  have the same psd  $S_{N_c}(\omega) = S_{N_s}(\omega)$ .  
5. Substituting  $\tau = 0$  in Eq. (21.4.27), we obtain  $R_N(0) = R_{N_c}(0)$  and from Eq. (21.4.26)

$$R_N(0) = R_{N_c}(0) = R_{N_s}(0) = \sigma_N^2 \quad (21.4.35)$$

Thus, the original random process  $N(t)$  and its quadrature components  $N_c(t)$  and  $N_s(t)$  have the same variance  $\sigma_N^2$ .

6. The autocorrelation function  $R_N(\tau)$  will be computed:

$$\begin{aligned} R_N(\tau) &= E\{[N_c(t) + jN_s(t)][N_c(t + \tau) - jN_s(t + \tau)]\} \\ &= R_{N_c}(\tau) + R_{N_s}(\tau) - j[R_{N_c N_s}(\tau) - R_{N_s N_c}(\tau)] \\ &= 2[R_{N_c}(\tau) - jR_{N_c N_s}(\tau)] \quad [\text{from Eq. (21.4.26)}] \end{aligned} \quad (21.4.36)$$

Multiplying Eq. (21.4.36) by  $e^{-j\omega_c \tau}$ , we obtain

$$\begin{aligned} R_N(\tau)e^{-j\omega_c \tau} &= 2[R_{N_c}(\tau) - jR_{N_c N_s}(\tau)][\cos(\omega_c \tau) - j \sin(\omega_c \tau)] \\ &= 2[R_{N_c}(\tau) - \cos(\omega_c t) - R_{N_c N_s} \sin(\omega_c t)] \\ &\quad - 2j[R_{N_c}(\tau) \sin(\omega_c \tau) + R_{N_c N_s}(\tau) \cos(\omega_c t) +] \\ &= 2[R_N(\tau) - jR_{\hat{N}}(\tau)] \quad [\text{from Eq. (21.4.29)}] \end{aligned} \quad (21.4.37)$$

Taking the real parts of both sides of Eq. (21.4.37) results in

$$R_N(\tau) = \frac{1}{2} \operatorname{Re}[R_N(\tau)e^{-j\omega_c \tau}] \quad (21.4.38)$$

7. The symmetry condition of any cross-correlation function is  $R_{N_c N_s}(\tau) = R_{N_s N_c}(-\tau)$ , and together with the stationarity condition of Eq. (21.4.26),  $R_{N_c N_s}(\tau) = -R_{N_s N_c}(\tau)$  implies that

$$R_{N_c N_s}(\tau) = R_{N_s N_c}(-\tau) = -R_{N_s N_c}(\tau) \quad (21.4.39)$$

Hence the cross-correlation function  $R_{N_c N_s}(\tau)$  is odd.

8. The autocorrelation function  $R_N(\tau)$  has Hermitian symmetry. From Eq. (21.4.36) and the fact that  $R_{N_c N_s}(\tau)$  is odd, from Eq. (21.4.39), we obtain

$$\begin{aligned} R_N^*(-\tau) &= 2[R_{N_c}(-\tau) + jR_{N_c N_s}(-\tau)] \\ &= 2[R_{N_c}(\tau) - jR_{N_c N_s}(\tau)], \quad R_{N_c N_s}(\tau) \text{ odd function} \\ &= R_N(\tau) \end{aligned} \quad (21.4.40)$$

showing Hermitian symmetry.

9. The cross-spectral density  $S_{N_c N_s}(\omega)$  is purely imaginary, as shown below:

$$\begin{aligned} S_{N_c N_s}(\omega) &= \int_{-\infty}^{\infty} R_{N_c N_s}(\tau) e^{-j\omega_c \tau} d\tau = - \int_{-\infty}^{\infty} R_{N_c N_s}(-\tau) e^{-j\omega_c \tau} d\tau \\ &= - \int_{-\infty}^{\infty} R_{N_c N_s}(\tau) e^{j\omega_c \tau} d\tau = -S_{N_c N_s}^*(\omega) \end{aligned} \quad (21.4.41)$$

The only way that Eq. (21.4.41) can be satisfied is when  $\text{Re}[S_{N_c N_s}(\omega)] = 0$ .

10. Since  $R_{N_c N_s}(\tau)$  is odd from, Eq. (21.4.39), we have  $R_{N_c N_s}(0) = 0$ . Hence  $E[N_c(t)N_s(t)] = 0$ , showing that the quadrature components  $N_c(t)$  and  $N_s(t)$  are orthogonal. Since they are zero mean, they are uncorrelated, and because they are Gaussian, they are also independent.
11. Structure of  $S_N(\omega)$  and  $S_N(\omega)$ . Since  $R_N(\tau)$  has Hermitian symmetry, from Eq. (21.4.40), its FT  $S_N(\omega)$  is a real-valued function of the frequency  $\omega$ . Using the relationship  $\text{Re}[z] = (z + z^*)/2$  in Eq. (21.4.38), we can write

$$R_N(\tau) = \frac{1}{4} [R_N(\tau) e^{-j\omega_c \tau} + R_N^*(\tau) e^{j\omega_c \tau}] \quad (21.4.42)$$

Taking FT on both sides of Eq. (21.4.42), we have

$$S_N(\tau) = \frac{1}{4} [S_N(\omega - \omega_c) + S_N(\omega + \omega_c)] \quad (21.4.43)$$

12. Structure of  $S_{N_c}(\omega) = S_{N_s}(\omega)$ . We will assume that the bandwidth  $2\omega_B$  of  $S_{N_c}(\omega)$  is governed by  $-\omega_B \leq \omega_c \leq \omega_B$ . From Eq. (21.4.30) we have

$$R_{N_c}(\tau) = R_{N_s}(\tau) = R_N(\tau) \cos(\omega_c \tau) + R_{\hat{N}}(\tau) \sin(\omega_c \tau) \quad (21.4.44)$$

Taking FT on both sides of Eq. (21.4.14), we obtain

$$\begin{aligned} S_{N_c}(\omega) = S_{N_s}(\omega) &= \frac{1}{2\pi} S_N(\omega)^* \pi [\delta(\omega + \omega_c) + \delta(\omega - \omega_c)] \\ &\quad - \frac{j}{2\pi} S_N(\omega) \text{sgn}(\omega)^* j\pi [\delta(\omega + \omega_c) - (\omega - \omega_c)] \\ &= \frac{1}{2} S_N(\omega + \omega_c) [1 + \text{sgn}(\omega + \omega_c)] \\ &\quad + \frac{1}{2} S_N(\omega - \omega_c) [1 - \text{sgn}(\omega - \omega_c)], \quad -\omega_B \leq \omega_c \leq \omega_B \end{aligned} \quad (21.4.45)$$

In Eq. (21.4.45) the terms containing the signum functions are given by

$$[1 + \text{sgn}(\omega + \omega_c)] = \begin{cases} 2, \omega > -\omega_c; \\ 0, \omega \leq -\omega_c \end{cases} \quad [1 - \text{sgn}(\omega - \omega_c)] = \begin{cases} 0, \omega > \omega_c \\ 2, \omega \leq \omega_c \end{cases} \quad (21.4.46)$$

Substituting Eq. (21.4.46) in Eq. (21.4.45), we obtain

$$\begin{aligned} S_{N_c}(\omega) = S_{N_s}(\omega) &= S_N(\omega + \omega_c) u(\omega + \omega_c) \\ &\quad + S_N(\omega - \omega_c) u[-(\omega - \omega_c)], \quad -\omega_B \leq \omega_c \leq \omega_B \end{aligned} \quad (21.4.47)$$

The development of the psd  $S_{N_c}(\omega)$  from the psd  $S_N(\omega)$  is shown in Fig. 21.4.5.

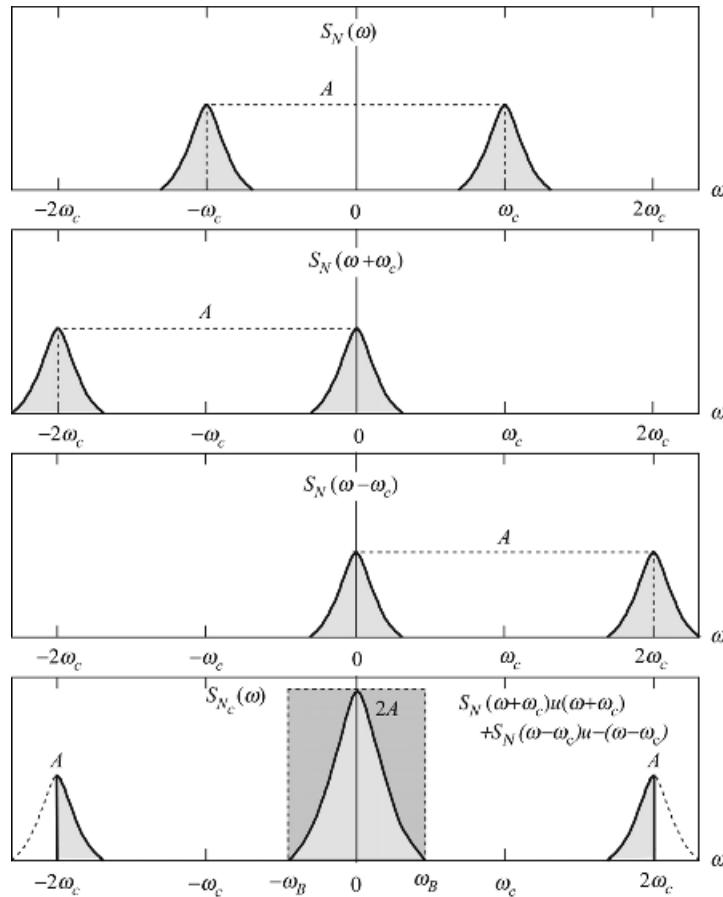


FIGURE 21.4.5

### Natural Envelope Representation

The natural envelope representation of a narrowband bandpass stationary random process is given by

$$N(t) = W(t) \cos[\omega_c t + \Theta(t)] \quad (21.4.19)$$

where  $W(t)$  in terms of the quadrature components  $N_c(t)$  and  $N_s(t)$  is given by

$$W(t) = \sqrt{N_c^2(t) + N_s^2(t)} \quad (21.4.48)$$

and the phase angle  $\Theta(t)$ , by

$$\Theta(t) = \tan^{-1} \left( \frac{N_s(t)}{N_c(t)} \right) \quad (21.4.49)$$

If  $N(t)$  is Gaussian-distributed, then from property 2 above, the quadrature components  $N_c(t)$  and  $N_s(t)$  are also Gaussian-distributed. We will assume that they are zero mean. Under these conditions we want to find the distributions of the natural envelope  $W(t)$  and the phase angle  $\Theta(t)$ . By property 10,  $N_c(t)$  and  $N_s(t)$  are independent, and by property 5, their variance  $\sigma_N^2$  is the same as the variance of  $N(t)$ . Hence their joint

density can be given as follows:

$$f_{N_c N_s}(n_c, n_s) = \frac{1}{2\pi\sigma_N^2} e^{-(1/2)[(n_c^2 + n_s^2)/\sigma_N^2]} \quad (21.4.50)$$

The transformation between  $N_c(t)$  and  $N_s(t)$  to  $W(t)$  is given by

$$\begin{aligned} N_c(t) &= W(t) \cos[\Theta(t)] \\ N_s(t) &= W(t) \sin[\Theta(t)] \end{aligned} \quad (21.4.51)$$

The pdf  $f_W(w)$  can be found by the techniques described in Section 13.3. The Jacobian of the transformation is

$$\|J(w, \theta)\| = \left\| \begin{bmatrix} \frac{\partial N_c(w, \theta)}{\partial w} & \frac{\partial N_c(w, \theta)}{\partial \theta} \\ \frac{\partial N_s(w, \theta)}{\partial w} & \frac{\partial N_s(w, \theta)}{\partial \theta} \end{bmatrix} \right\| = \left\| \begin{bmatrix} \cos(\theta) & -w \sin(\theta) \\ \sin(\theta) & w \cos(\theta) \end{bmatrix} \right\| \quad (21.4.52)$$

or

$$\|J(w, \theta)\| = w[\cos^2(\theta) + \sin^2(\theta)] = w$$

The solutions for  $w$  and  $\theta$  are

$$w = \sqrt{n_c^2 + n_s^2}; \quad \theta = \tan^{-1}\left(\frac{n_s}{n_c}\right) \quad (21.4.53)$$

Hence, from Eq. (13.3.5), we obtain

$$f_{N_c N_s}(n_c, n_s) = \frac{1}{2\pi\sigma_N^2} e^{-(1/2)[(n_c^2 + n_s^2)/\sigma_N^2]} = \frac{w}{2\pi\sigma_N^2} e^{-(1/2)(w^2/\sigma_N^2)} = f_{W, \Theta}(w, \theta) \quad (21.4.54)$$

which is the desired joint density  $f_{W, \Theta}(w, \theta)$ . Integrating Eq. (21.4.54) with respect to  $\theta$ , we obtain

$$f_W(w) = \int_0^{2\pi} \frac{w}{2\pi\sigma_N^2} e^{-(1/2)(w^2/\sigma_N^2)} d\theta = \frac{w}{\sigma_N^2} e^{-(1/2)(w^2/\sigma_N^2)}, \quad w > 0 \quad (21.4.55)$$

which is a Rayleigh distribution shown in Fig. 7.8.2. Similarly, integrating with respect to  $w$ , we obtain

$$f_\Theta(\theta) = \frac{1}{2\pi} \int_0^{2\pi} \frac{w}{\sigma_N^2} e^{-(1/2)(w^2/\sigma_N^2)} d\theta = \frac{1}{2\pi}, \quad 0 \leq \theta \leq 2\pi \quad (21.4.56)$$

which is a uniform distribution in  $(0, 2\pi)$ . Thus we find that  $W(t)$  and  $\Theta(t)$  are also independent. This derivation holds good only if the mean values of the quadrature components  $N_c(t)$  and  $N_s(t)$  are zero. We will now discuss the case where the means are not zero.

### Narrowband Bandpass Stationary Process with Additive Sinusoid

A stationary bandpass random process  $X(t)$  is given by

$$X(t) = N(t) + a \cos[\omega_c t + \Psi] \quad (21.4.57)$$

where  $N(t)$  is a zero mean Gaussian stationary narrowband bandpass random process defined by Eq. (21.4.19) with quadrature components  $N_c(t)$  and  $N_s(t)$ , defined by Eq. (21.4.51).  $\Psi$  is a random phase uniformly distributed in  $(0, 2\pi)$ , and  $a$  is a constant.

The variances of  $N(t)$ ,  $N_c(t)$ , and  $N_s(t)$  are all equal to  $\sigma_N^2$ . The natural envelope of  $X(t)$  will be given by

$$X(t) = R(t) \cos[\omega_c t + \Phi(t)] \quad (21.4.58)$$

where  $\Phi(t)$  is some function of both  $\Psi$  and  $\Theta(t)$ . Expanding Eq. (21.4.58), we obtain

$$X(t) = R(t)[\cos[\Phi(t)] \cos(\omega_c t) - \sin[\Phi(t)] \sin(\omega_c t)] \quad (21.4.59)$$

Substituting Eq. (21.4.20a) for  $N(t)$  in Eq. (21.4.57), we obtain

$$X(t) = [N_c(t) + a \cos(\Psi)] \cos(\omega_c t) - [N_s(t) + a \sin(\Psi)] \sin(\omega_c t) \quad (21.4.60)$$

and the quadrature components of  $X(t)$  can be defined as

$$\begin{aligned} X_c(t) &= N_c(t) + a \cos(\Psi), \text{ in phase} \\ X_s(t) &= N_s(t) + a \sin(\Psi), \text{ quadrature} \end{aligned} \quad (21.4.61)$$

Unlike  $N_c(t)$  and  $N_s(t)$ , which are zero mean Gaussian lowpass random processes,  $X_c(t)$  and  $X_s(t)$  are Gaussian but not zero mean processes. Their mean values are given by

$$\mu_{X_c} = a \cos(\Psi); \quad \mu_{X_s} = a \sin(\Psi) \quad \text{and} \quad a = \sqrt{\mu_{X_c}^2 + \mu_{X_s}^2} \quad (21.4.62)$$

Without loss of generality, we can assume that  $\Psi = \psi$  is a constant and the quadrature components  $X_c(t)$  and  $X_s(t)$  are independent. Under these assumptions their joint Gaussian density is

$$\begin{aligned} f_{X_c X_s}(x_c, x_s) &= \frac{1}{2\pi\sigma_N^2} e^{-(1/2\sigma_N^2)[x_c - a \cos(\psi)]^2 + [x_s - a \sin(\psi)]^2} \\ &= \frac{1}{2\pi\sigma_N^2} e^{-(1/2\sigma_N^2)\{x_c^2 + x_s^2 + a^2 - 2a[x_c \cos(\psi) + x_s \sin(\psi)]\}} \end{aligned} \quad (21.4.63)$$

Comparing terms in Eqs. (21.4.59) and (21.4.60), we can write

$$\begin{aligned} X_c(t) &= R(t) \cos[\Phi(t)] = N_c(t) + a \cos(\psi) \\ X_s(t) &= R(t) \sin[\Phi(t)] = N_s(t) + a \sin(\psi) \end{aligned} \quad (21.4.64)$$

To obtain the joint density of  $R(t)$  and  $\Phi(t)$ , we use the following transformation:

$$x_c = r \cos(\phi); \quad x_s = r \sin(\phi) \quad (21.4.65)$$

The Jacobian of the transformation of Eq. (21.4.65)  $\|J\| = r$ . In Eq. (21.4.63) we note that the term

$$x_c \cos(\psi) + x_s \sin(\psi) = r \cos(\phi) \cos(\psi) + r \sin(\phi) \sin(\psi) = r \cos(\phi - \psi) \quad (21.4.66)$$

Substituting Eqs. (21.4.65) and (21.4.66) in Eq. (21.4.63), the joint pdf  $f_{R\Phi}(r, \phi)$  can be written as follows:

$$f_{R,\Phi}(r, \phi) = \frac{r}{2\pi\sigma_N^2} e^{-(1/2\sigma_N^2)\{r^2 + a^2 - 2ar \cos(\phi - \psi)\}}, \quad \begin{cases} 0 \leq r < \infty \\ 0 \leq \phi \leq 2\pi \end{cases} \quad (21.4.67)$$

The marginal density of the envelope  $R(t)$  can be found by integrating Eq. (21.4.67) with respect to  $\phi$ . Substituting  $\theta = \phi - \psi$  in Eq. (21.4.67), we obtain

$$f_R(r) = \frac{r}{\sigma_N^2} e^{[(r^2+a^2)/2\sigma_N^2]} \frac{1}{2\pi} \int_0^{2\pi} e^{\frac{ar \cos(\theta)}{\sigma_N^2}} d\theta = \frac{r}{\sigma_N^2} e^{-(r^2+a^2)/2\sigma_N^2} I_0\left(\frac{ar}{\sigma_N^2}\right) \quad (21.4.68)$$

where we have utilized the fact that  $e^{ar \cos(\theta)}$  is periodic with period  $2\pi$ , that is,  $e^{ar \cos(\theta)} = e^{ar \cos(2\pi + \theta)}$ . The integral in Eq. (21.4.68) is the zero-order modified Bessel function of the first kind given in Eq. (7.8.13) and shown in Fig. 7.8.4.

$$I_0\left(\frac{ar}{\sigma_N^2}\right) = \frac{1}{2\pi} \int_0^{2\pi} e^{[ar \cos(\theta)/\sigma_N^2]} d\theta \quad (21.4.69)$$

Equation (21.4.68) can be normalized by substituting

$$z = \frac{r}{\sigma_N}; \quad k = \frac{a}{\sigma_N}$$

and written as

$$f_Z(z) = z e^{[(z^2+k^2)/2]} I_0(kz) \quad (21.4.70)$$

This is called *Rice's density* and is shown in Fig. 7.8.5 and reproduced in Fig. 21.4.6 for  $k = 0, \dots, 5$ .

If  $a = 0$ , the mean values are 0, and we obtain the Rayleigh density shown in Fig. 21.4.6. To obtain the phase density function, we integrate Eq. (21.4.67) over  $r$ , yielding

$$f_\Phi(\phi) = \frac{1}{2\pi\sigma_N^2} \int_0^{2\pi} r e^{-(1/2\sigma_N^2)\{r^2+a^2-2ar \cos(\phi-\psi)\}} dr \quad (21.4.71)$$

Completing the squares in the expression  $r^2 + a^2 - 2ar \cos(\phi - \psi)$ , we can rewrite Eq. (21.4.71) as follows:

$$f_\Phi(\phi) = \frac{1}{2\pi} e^{-(1/2\sigma_N^2)[1-\cos^2(\phi-\psi)]} \int_0^\infty \frac{r}{\sigma_N} \exp\left\{-\frac{1}{2}\left[\frac{r^2 - a \cos(\phi - \psi)}{\sigma_N}\right]^2\right\} \frac{dr}{\sigma_N} \quad (21.4.72)$$

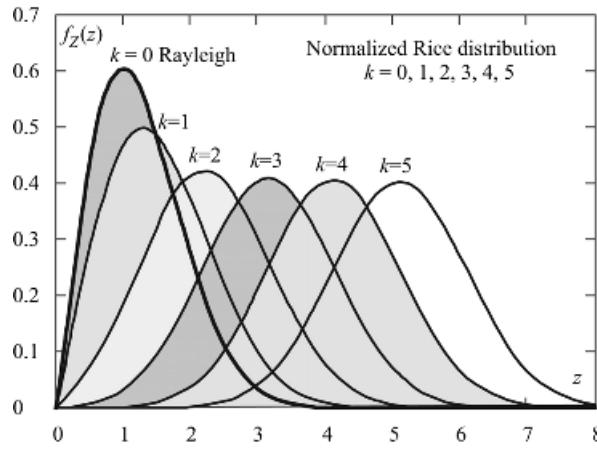


FIGURE 21.4.6

Making the following substitutions for convenience in Eq. (21.4.72)

$$z = \frac{r}{\sigma_N}; k = \frac{a}{\sigma_N}; \theta = \phi - \psi; b = k \cos(\theta)$$

as in the previous case, we can write

$$f_\Phi(\phi) = e^{-(1/2)k^2 \sin^2(\theta)} \frac{1}{2\pi} \int_0^\infty z e^{-(1/2)(z-b)^2} dz \quad (21.4.73)$$

Again substituting  $w = z - b$  in Eq. (21.4.73) and rearranging terms, we obtain

$$\begin{aligned} f_\Phi(\phi) &= e^{-(1/2)k^2 \sin^2(\theta)} \frac{1}{2\pi} \int_{-b}^\infty (w + b) e^{-(1/2)w^2} dw \\ f_\Phi(\phi) &= e^{-(1/2)k^2 \sin^2(\theta)} \frac{1}{2\pi} \int_{-b}^\infty w e^{-(1/2)w^2} dw + \frac{b}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}} \int_{-b}^\infty e^{-(1/2)w^2} dw \end{aligned} \quad (21.4.74)$$

Evaluating the integrals, we obtain

$$f_\Phi(\phi) = e^{-(1/2)k^2 \sin^2(\theta)} \frac{1}{2\pi} e^{-(b^2/2)} + \frac{b}{\sqrt{2\pi}} \Phi(b) \quad (21.4.75)$$

where  $\Phi(b)$  is the Gaussian integral

$$\Phi(b) = \frac{1}{\sqrt{2\pi}} \int_{-b}^\infty e^{-(1/2)w^2} dw = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-(1/2)w^2} dw$$

Substituting back for  $b = k \cos(\phi - \psi)$  in Eq. (21.4.75), we have

$$f_\Phi(\phi) = \frac{1}{2\pi} e^{-(k^2/2)} + \frac{k \cos(\phi - \psi)}{\sqrt{2\pi}} e^{-(1/2)k^2 \sin^2(\phi - \psi)} \Phi(b) \quad (21.4.76)$$

which is the desired result for the phase density. When  $k = 0$ , the phase  $\phi$  is uniformly distributed in  $(0, 2\pi)$ . However, if  $k \neq 0$ , then the quadrature components are not independent, unlike the case when  $k = 0$ . The phase density is shown in Fig. 21.4.7 for  $\psi = \pi$ .

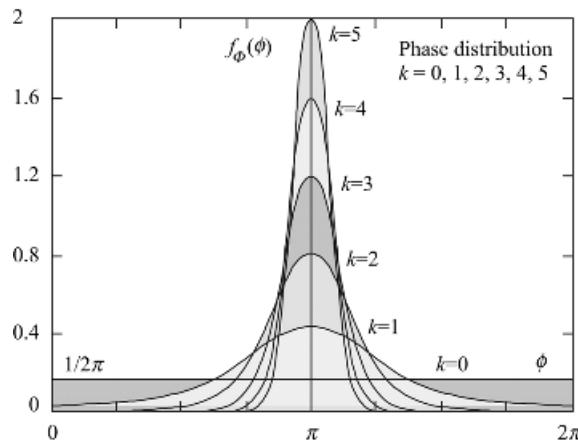


FIGURE 21.4.7

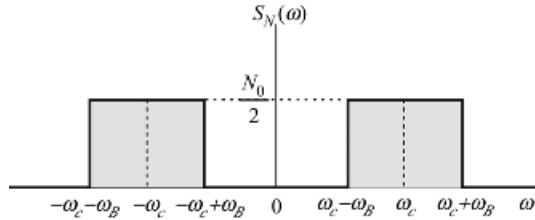


FIGURE 21.4.8

**Example 21.4.4** A received signal is of the form

$$X(t) = A \cos[(\omega_c + \omega_d)t] + N(t) \quad (21.4.77)$$

where  $N(t)$  is a bandpass white-noise signal with psd as shown in Fig. 21.4.8.

We want to express  $N(t)$  in the form

$$N(t) = N'_c(t) \cos[(\omega_c + \omega_d)t] - N'_s(t) \sin[(\omega_c + \omega_d)t] \quad (21.4.78)$$

with the quadrature components of  $N(t)$  given by [Eq. (21.4.20a)]

$$N(t) = N_c(t) \cos(\omega_c t) - N_s(t) \sin(\omega_c t) \quad (21.4.79)$$

We will express  $N'_c(t)$  and  $N'_s(t)$  in terms of  $N_c(t)$  and  $N_s(t)$ . Expanding the righthand side (RHS) of Eq. (21.4.78), we obtain

$$\begin{aligned} N'_c(t) \cos[(\omega_c + \omega_d)t] - N'_s(t) \sin[(\omega_c + \omega_d)t] \\ = [N'_c(t) \cos(\omega_d t) - N'_s(t) \sin(\omega_d t)] \cos(\omega_c t) \\ - [N'_c(t) \sin(\omega_d t) + N'_s(t) \cos(\omega_d t)] \sin(\omega_c t) \end{aligned} \quad (21.4.80)$$

Equating like coefficients in Eqs. (21.4.80) and (21.4.79), we obtain, in matrix form

$$\begin{bmatrix} N_c(t) \\ N_s(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega_d t) & -\sin(\omega_d t) \\ \sin(\omega_d t) & \cos(\omega_d t) \end{bmatrix} \begin{bmatrix} N'_c(t) \\ N'_s(t) \end{bmatrix} \quad (21.4.81)$$

The multiplying matrix in Eq. (21.4.81) is called a *rotation matrix* and is orthogonal. The inverse is the transpose, and hence we can write

$$\begin{bmatrix} N'_c(t) \\ N'_s(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega_d t) & \sin(\omega_d t) \\ -\sin(\omega_d t) & \cos(\omega_d t) \end{bmatrix} \begin{bmatrix} N_c(t) \\ N_s(t) \end{bmatrix} \quad (21.4.82)$$

or

$$\begin{aligned} N'_c(t) &= N_c(t) \cos(\omega_d t) + N_s(t) \sin(\omega_d t) \\ N'_s(t) &= -N_c(t) \sin(\omega_d t) + N_s(t) \cos(\omega_d t) \end{aligned} \quad (21.4.83)$$

We can now find the autocorrelation functions,  $R_{N'_c}(\tau)$  and  $R_{N'_s}(\tau)$ , in terms of  $R_{N_c}(\tau)$  and  $R_{N_s}(\tau)$  as follows:

$$\begin{aligned} R_{N'_c}(\tau) &= E[N'_c(t)N'_c(t + \tau)] \\ &= E\{[N_c(t) \cos(\omega_d t) + N_s(t) \sin(\omega_d t)] \\ &\quad \times [N_c(t + \tau) \cos(\omega_d(t + \tau)) + N_s(t + \tau) \sin(\omega_d(t + \tau))]\} \\ &= R_{N_c}(\tau) \cos(\omega_d t) \cos(\omega_d(t + \tau)) + R_{N_s}(\tau) \sin(\omega_d t) \sin(\omega_d(t + \tau)) \\ &\quad + R_{N_c N_s} \cos(\omega_d t) \sin(\omega_d(t + \tau)) + R_{N_s N_c} \sin(\omega_d t) \cos(\omega_d(t + \tau)) \end{aligned} \quad (21.4.84)$$

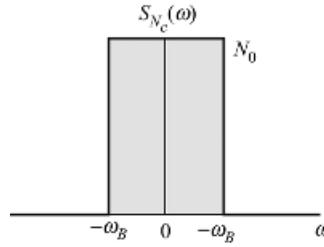


FIGURE 21.4.9

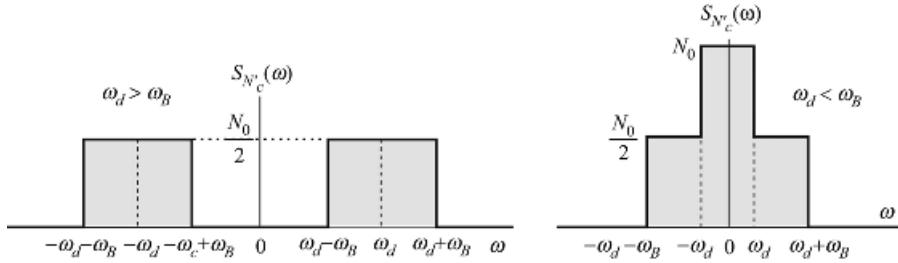


FIGURE 21.4.10

Substituting  $R_{N_c}(\tau) = R_{N_s}(\tau)$ ;  $R_{N_c N_s}(\tau) = -R_{N_s N_c}(\tau)$  from Eq. (21.4.26) in Eq. (21.4.84), we obtain

$$R_{N'_c}(\tau) = R_{N_c}(\tau) \cos(\omega_d \tau) + R_{N_c N_s}(\tau) \sin(\omega_d \tau) \quad (21.4.85)$$

This equation [Eq. (21.4.85)] applies for any stationary random process  $N(t)$ . For a white-noise process, we have

$$R_{N_c N_s}(\tau) = 0, \quad \tau \neq 0 \quad (21.4.86)$$

and substituting Eq. (21.4.86) in Eq. (21.4.85), we obtain

$$R_{N'_s}(\tau) = R_{N_c}(\tau) \cos(\omega_d \tau) \quad (21.4.87)$$

In a similar manner, we can obtain

$$R_{N'_s}(\tau) = R_{N_c}(\tau) \cos(\omega_d \tau) \quad (21.4.88)$$

From Eqs. (21.4.87) and (21.4.88), we find that  $R_{N'_c}(\tau) = R_{N'_s}(\tau)$ . The psd  $S_{N_c}(\omega)$  is obtained from Eq. (21.4.47) as

$$S_{N_c}(\omega) = S_{N_s}(\omega) = N_0[u(\omega + \omega_B) - u(\omega - \omega_B)] \quad (21.4.89)$$

and is shown in Fig. 21.4.9.

The psd  $S_{N'_c}(\omega)$  is obtained by using the modulation property of FT and is given by

$$\begin{aligned} S_{N'_c}(\omega) &= \frac{1}{2\pi} S_{N_c}(\omega) \cdot \pi[\delta(\omega + \omega_d) + \delta(\omega - \omega_d)] \\ &= \frac{1}{2} [S_{N_c}(\omega + \omega_d) + S_{N_c}(\omega - \omega_d)] \end{aligned} \quad (21.4.90)$$

$S_{N'_c}(\omega)$  is shown in Fig. 21.4.10 for the two cases of  $\omega_d > \omega_B$  and  $\omega_d < \omega_B$ .

# Weiner and Kalman Filters

## 22.1 REVIEW OF ORTHOGONALITY PRINCIPLE

### *Random Variables*

The estimate  $\hat{Y}$  of a random variable  $Y$  by a function of another random variable  $g(X)$  using the mean-square error criterion  $J = E[Y - g(X)]^2$  is found in Chapter 18 as the conditional expectation given by Eq. (18.2.7):

$$g(X) = \hat{Y} = E[Y | X] \quad (18.2.7)$$

and this estimator is also the orthogonal projection of  $Y$  in the space of  $X$  in the sense

$$E[\varepsilon \hat{Y}] = E\{[Y - g(X)]g(X)\} = 0 \quad (22.1.1)$$

In the case of linear estimation given by

$$g(\mathbf{X}) = \hat{Y} = \mathbf{a}^T \mathbf{X} \quad (22.1.2)$$

the coefficient vector  $\mathbf{a}$  is determined from Eq. (18.2.23) as

$$\mathbf{a} = \mathbf{R}_X^{-1} \mathbf{R}_{XY} \quad (22.1.3)$$

where  $\mathbf{R}_X$  is the correlation matrix  $E[\mathbf{XX}^T]$  and  $\mathbf{R}_{XY}$  is the cross-correlation vector  $E[\mathbf{XY}]$ .

We will now consider the one-dimensional case where  $\hat{Y} = aX$ . The orthogonality condition of Eq. (22.1.1) yields

$$E\{[Y - aX]aX\} = E[XY] - aE[X^2] \quad \text{or} \quad a = \frac{E[XY]}{E[X^2]} \quad (22.1.4)$$

and using the orthogonality principle, the minimum mean-square error is

$$J_{\min} = E[\varepsilon^2] = E[\varepsilon(Y - aX)] = E[\varepsilon Y] = E[Y^2] - \frac{E[XY]^2}{E[X^2]} \quad (22.1.5)$$

and if  $X$  and  $Y$  have zero mean, then we can define  $\rho_{XY}$  as the correlation coefficient between  $X$  and  $Y$  and write Eq. (22.1.5) as

$$J_{\min} = E[Y^2][1 - \rho_{XY}^2] \quad (22.1.6)$$

### Random Processes

We can extend the preceding analysis to random processes. The input to the system shown in Fig. 22.1.1 is a zero mean random process  $X(t)$  consisting of a zero mean signal process  $S(t)$  and a zero mean additive noise  $N(t)$  that is independent of  $S(t)$ , both defined in the *estimation time interval*  $T$ :

$$X(t) = S(t) + N(t), \quad t \in T; E[S(t)N(t)] = E[S(t)]E[N(t)] = 0, \quad t \in T \quad (22.1.7)$$

The output of the system is  $Y(t)$ . The system is observed during the *observation time interval*  $U$ .  $T$  and  $U$  may be different or the same. We want to design a linear system with impulse response  $h(t,u)$ ,  $u \in U$  such that the output  $Y(t)$  approximates the desired signal  $S(t)$  according to the minimum mean-square error criterion given by

$$J = \min_{h(t,u)} E\{[S(t) - Y(t)]^2\}, \quad t \in T, \quad u \in U \quad (22.1.8)$$

where

$$Y(t) = \int_U h(t,u)X(u)du, \quad t \in T \quad (22.1.9)$$

For this case the orthogonality condition  $E\{[S(t) - Y(t)]Y(t)\} = 0$  becomes

$$E\left\{ \left[ S(t) - \int_U h(t,u)X(u)du \right] \int_U h(t,v)X(v)dv \right\} = 0, \quad t \in T \quad (22.1.10)$$

This equation can be rearranged as follows:

$$\int_U h(t,v) \left\{ E[S(t)X(v)] - \int_U h(t,u)E[X(u)X(v)]du \right\} dv, \quad t \in T \quad (22.1.11)$$

If Eq. (22.1.11) has to be satisfied, then the quantity within braces must be equal to 0:

$$\int_U h_{\text{opt}}(t,u)E[X(u)X(v)]du = E[S(t)X(v)], \quad t \in T, \quad v \in T \quad (22.1.12)$$

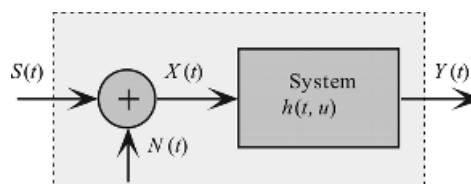


FIGURE 22.1.1

We can now express Eq. (22.1.12) in terms of correlation functions as

$$\int_U h_{\text{opt}}(t,u)R_X(u,v)du = R_{XS}(t,v), \quad t \in T, \quad v \in T \quad (22.1.13)$$

where  $h_{\text{opt}}(t)$  is the optimum impulse response. The corresponding minimum mean-square error  $J_{\text{min}}$  is given by

$$\begin{aligned} J_{\text{min}} &= E\{[S(t) - Y_o(t)]S(t)\} \\ &= E[S^2(t)] - \int_U h_{\text{opt}}(t,u)E[X(u)S(t)]du \\ &= R_S(t,t) - \int_U h_{\text{opt}}(t,u)R_{XS}(t,u)du, \quad t \in T \end{aligned} \quad (22.1.14)$$

If  $S(t)$  is replaced by  $S(t + \alpha)$  in Eqs. (22.1.12), then we have

$$\int_U h_{\text{opt}}(t,u)E[X(v)X(u)]du = E[X(v)S(t + \alpha)], \quad t \in T, \quad v \in T \quad (22.1.15)$$

and Eq. (21.1.13) becomes

$$\int_U h_{\text{opt}}(t,u)R_X(u,v)du = R_{XS}(t + \alpha, v), \quad t \in T, \quad v \in T \quad (22.1.16)$$

In Eq. (22.1.16), if  $\alpha = 0$ , we have a *filtering* problem; if  $\alpha > 0$ , we have a *prediction* or *extrapolation* problem; and if  $\alpha < 0$ , we have a *smoothing* or *interpolation* problem.

The corresponding equation for the minimum mean square error is

$$\begin{aligned} J_{\text{min}} &= E[S^2(t + \alpha)] - \int_U h_{\text{opt}}(t,u)E[X(u)S(t + \alpha)]du \\ &= R_S(t,t) - \int_U h_{\text{opt}}(t,u)R_{XS}(t + \alpha, u)du, \quad t \in T \end{aligned} \quad (22.1.17)$$

## 22.2 WIENER FILTERING

### Noncausal Wiener Filtering (Unrealizable Filter)

We will now assume that the estimation time interval  $T$  and the observation time interval  $U$  are the real line, or,  $T = U = (-\infty, \infty)$ . The processes  $S(t), N(t)$  are stationary, and the system is time-invariant but not causal. Under these assumptions Eq. (22.1.16) can be rewritten as

$$\int_{-\infty}^{\infty} h_{\text{opt}}(t-u)R_X(u-v)du = R_{XS}(t + \alpha - v), \quad t \in T, \quad v \in T \quad (22.2.1)$$

We make a change of variables  $t - u = \xi$  and  $t - v = \tau$  in Eq. (22.2.1) and we have

$$\int_{-\infty}^{\infty} h_{\text{opt}}(\xi)R_X(\tau - \xi)d\xi = R_{XS}(\tau + \alpha)$$

Or

$$h_{\text{opt}}(\tau) * R_X(\tau) = R_{XS}(\tau + \alpha), \quad t \in T \quad (22.2.2)$$

The optimum transfer function  $H_{\text{opt}}(\omega)$  is obtained by taking FT of Eq. (22.2.2), yielding

$$H_{\text{opt}}(\omega) = \frac{S_{XS}(\omega)e^{j\omega\alpha}}{S_X(\omega)} \quad (22.2.3)$$

Since the signal and noise are assumed independent, the relevant correlation functions and the spectral densities are

$$\begin{aligned} R_X(\tau) &= R_S(\tau) + R_N(\tau) \\ R_{XS}(\tau) &= E\{[S(t) + N(t)]S(t + \tau)\} = R_S(\tau) \\ S_X(\omega) &= S_S(\omega) + S_N(\omega); \quad S_{XS}(\omega) = S_S(\omega) \end{aligned} \quad (22.2.4)$$

Substituting Eq. (22.2.4) into Eq. (22.2.3), we obtain

$$H_{\text{opt}}(\omega) = \frac{S_S(\omega)e^{j\omega\alpha}}{S_S(\omega) + S_N(\omega)} \quad (22.2.5)$$

If  $N(t) = 0$  in Eq. (22.1.7), then  $X(t) = S(t)$ ,  $R_X(\tau) = R_S(\tau)$ , and Eq. (22.2.5) becomes

$$H_{\text{opt}}(\omega) = e^{j\omega\alpha} \quad (22.2.6)$$

a pure prediction if  $\alpha > 0$  and a pure delay if  $\alpha < 0$ .

The minimum mean-square error is obtained from Eq. (22.1.17) as

$$\begin{aligned} J_{\text{min}} &= R_S(0) - \int_{-\infty}^{\infty} h_{\text{opt}}(t-u)R_{XS}(t-u+\alpha)du \\ &= R_S(0) - \int_{-\infty}^{\infty} h_{\text{opt}}(\tau)R_{XS}(\tau+\alpha)d\tau \end{aligned} \quad (22.2.7)$$

and using Parseval's theorem, we obtain the following in the frequency domain:

$$J_{\text{min}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} [S_S(\omega) - S_{XS}(\omega)H_{\text{opt}}^*(\omega)e^{-j\omega\alpha}]d\omega \quad (22.2.8)$$

Substituting Eq. (22.2.3) in Eq. (22.2.8), we obtain

$$\begin{aligned} J_{\text{min}} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[ S_S(\omega) - S_{XS}(\omega) \frac{S_{XS}^*(\omega)}{S_X^*(\omega)} \right] d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[ \frac{S_S(\omega)S_X^*(\omega) - |S_{XS}(\omega)|^2}{S_X^*(\omega)} \right] d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[ \frac{S_S(\omega)S_X(\omega) - |S_{XS}(\omega)|^2}{S_X(\omega)} \right] d\omega, \quad \text{since } S_X^*(\omega) = S_X(\omega) \end{aligned} \quad (22.2.9)$$

and if  $S(t)$  and  $N(t)$  are independent, Eq. (22.2.9) reduces to

$$J_{\text{min}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[ \frac{S_S(\omega)S_N(\omega)}{S_S(\omega) + S_N(\omega)} \right] d\omega \quad (22.2.10)$$

If the noise  $N(t) = 0$ , then we obtain  $J_{\text{min}} = 0$  from Eq. (22.2.10). It should be noted that the optimum transfer function  $H_{\text{opt}}(\omega)$  is noncausal and hence unrealizable.

In the following examples we will assume that the signal  $S(t)$  and the noise  $N(t)$  are independent and both are zero mean stationary random processes.

**Example 22.2.1** The autocorrelation functions of the input signal  $S(t)$  and the noise  $N(t)$  to a system are

$$R_S(\tau) = e^{-|\tau|/2} \quad \text{and} \quad R_N(\tau) = 4\delta(\tau)$$

We will estimate the signal  $S(t + \alpha)$  by finding the unrealizable filter  $H_{\text{opt}}(\omega)$ . The power spectral densities (psd) of signal and noise are

$$S_S(\omega) = \frac{4}{1 + 4\omega^2} \quad \text{and} \quad S_N(\omega) = 4$$

Since the signal and noise are independent, we can use Eq. (22.2.5) to obtain the optimum transfer function  $H_{\text{opt}}(\omega)$  as

$$H_{\text{opt}}(\omega) = \frac{\frac{4}{1 + 4\omega^2} e^{j\omega\alpha}}{\frac{4}{1 + 4\omega^2} + 4} = \frac{e^{j\omega\alpha}}{2(1 + 2\omega^2)}$$

The optimum transfer function  $|H_{\text{opt}}(\omega)|$  is shown in Fig. 22.2.1.

The corresponding optimum impulse response  $h_{\text{opt}}(t)$  is obtained by taking the inverse Fourier transform (FT) of the optimum transfer function:

$$h_{\text{opt}}(t) = \frac{1}{4\sqrt{2}} e^{-|t-\alpha|/\sqrt{2}}$$

$h_{\text{opt}}(t)$  is shown in Fig. 22.2.2 for values of the lag  $\alpha = 6, 3, 0, -3$ .

It is seen from the figure that  $h_{\text{opt}}(t)$  is unrealizable because it is noncausal; however, it is approximately causal for lags  $\alpha \geq 6$ . It can be truly causal only for infinite lags.

The minimum mean-square error is obtained from Eq. (22.2.10) as follows:

$$J_{\text{min}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\frac{4}{1 + 4\omega^2} \cdot 4}{\frac{4}{1 + 4\omega^2} + 4} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2}{1 + 2\omega^2} d\omega = \frac{1}{\sqrt{2}}$$

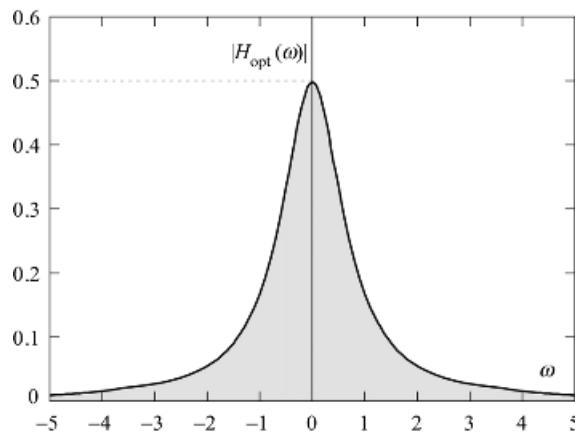


FIGURE 22.2.1

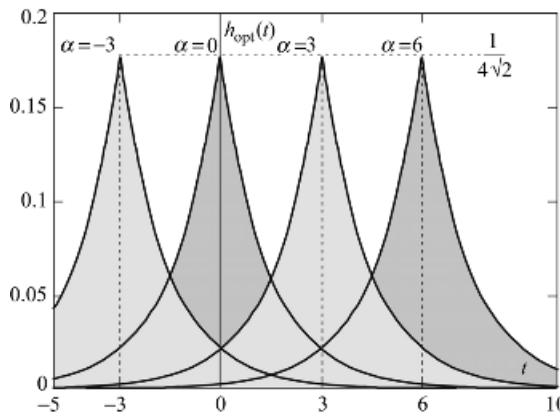


FIGURE 22.2.2

**Example 22.2.2** This example will show that even for an infinite lag,  $h_{\text{opt}}(t)$  cannot be causal.

The autocorrelation functions of the input signal  $S(t)$  and the noise  $N(t)$  to a system are

$$R_S(\tau) = e^{-|\tau|/2} \quad \text{and} \quad R_N(\tau) = e^{-2|\tau|}$$

We have to estimate the signal  $S(t + \alpha)$ . We will determine the unrealizable optimum filter  $H_{\text{opt}}(\omega)$  and the corresponding minimum mean-square error. The psd's of the signal and noise are

$$S_S(\omega) = \frac{4}{1 + 4\omega^2} \quad \text{and} \quad S_N(\omega) = \frac{4}{4 + \omega^2}$$

Since the signal and noise are independent, we can use Eq. (22.2.5) to obtain the transfer function  $H_{\text{opt}}(\omega)$  as follows:

$$H_{\text{opt}}(\omega) = \frac{\frac{4}{1 + 4\omega^2} e^{j\omega\alpha}}{\frac{4}{4 + \omega^2} + \frac{4}{1 + 4\omega^2}} = \frac{4 + \omega^2}{5(1 + \omega^2)} e^{j\omega\alpha}$$

The optimum transfer function is shown in Fig. 22.2.3.

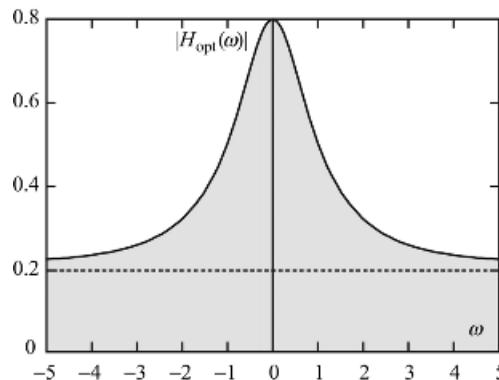


FIGURE 22.2.3

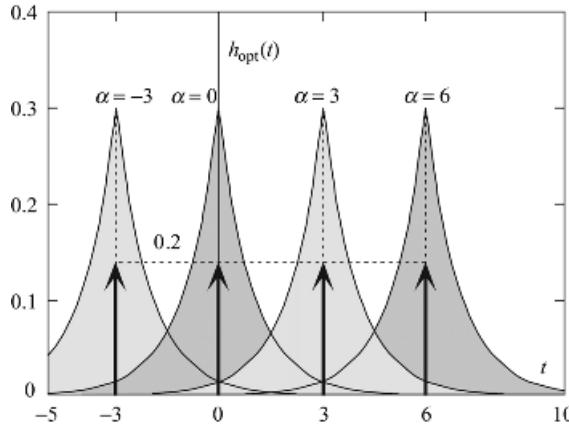


FIGURE 22.2.4

The transfer function  $H_{\text{opt}}(\omega)$  is not physically realizable since it violates the Paley–Wiener criterion given by Eq. (21.3.4):

$$\int_{-\infty}^{\infty} \frac{\ln |H(\omega)|}{1 + \omega^2} d\omega < \infty \quad (21.3.4)$$

Taking the inverse FT, we obtain the optimum impulse response  $h_{\text{opt}}(t)$  as

$$h_{\text{opt}}(t) = \frac{1}{5} \delta(t) + \frac{3}{10} e^{-|t-\alpha|}$$

$h_{\text{opt}}(t)$  is shown in Fig. 22.2.4 for values of  $\alpha = 6, 3, 0, -3$ .

The impulse response is also noncausal for any value of the lag  $\alpha$  because of the presence of the impulse function.

The minimum mean-square error is obtained from Eq. (22.2.9) and is given by

$$J_{\min} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\frac{4}{1+4\omega^2} \cdot \frac{4}{4+\omega^2}}{\frac{4}{1+4\omega^2} + \frac{4}{4+\omega^2}} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{4}{5(1+\omega^2)} d\omega = \frac{2}{5}$$

#### Causal Weiner Filtering (Realizable Filter)

In the last section we derived an integral equation [Eq. (22.2.2)] for an unrealizable filter. We now impose the condition that the system is causal, that is,  $h(t) = 0$ ,  $t < 0$ . The estimation interval is the real line  $T = \{-\infty < t < \infty\}$ , and the observation interval is now  $U = \{u \leq t\}$ . The ranges for  $S(t)$  and  $N(t)$  are  $T$ . Under this condition, Eq. (22.1.13) becomes

$$\int_{-\infty}^t h_{\text{opt}}(t-u) R_X(u-v) du = R_{XS}(t-v), \quad v \in U \quad (22.2.11)$$

Making the change of variables  $t-u = \xi$  and  $t-v = \tau$  in Eq. (22.2.11), we obtain

$$\int_0^\infty h_{\text{opt}}(\xi) R_X(\tau - \xi) d\xi = R_{XS}(\tau), \quad \tau \geq 0 \quad (22.2.12)$$

Equation (22.2.12) is called the *Wiener–Hopf integral* equation. This equation is a necessary and sufficient condition for the causal optimum filter  $h_{\text{opt}}(t) = 0$ ,  $t < 0$  to exist.

The corresponding mean-square error is obtained from Eq. (22.1.14) as

$$J_{\min} = R_S(0) - \int_0^\infty h_{\text{opt}}(\tau) R_{XS}(\tau) d\tau \quad (22.2.13)$$

Even though Eq. (22.2.12) is similar to Eq. (22.2.2) with  $\alpha = 0$ , the solution is far more complex because  $h(\tau)$  was required to vanish along the negative time axis.

If we take the FT of both sides of Eq. (22.1.12), we obtain

$$\int_0^\infty R_{XS}(\tau) e^{-j\omega\tau} d\tau = \int_0^\infty \left[ \int_0^\infty h_{\text{opt}}(\xi) R_X(\tau - \xi) d\xi \right] e^{-j\omega\tau} d\tau$$

or

$$\begin{aligned} \int_0^\infty R_{XS}(\tau) e^{-j\omega\tau} d\tau &= \int_0^\infty \left[ \int_0^\infty h_{\text{opt}}(\xi) e^{-j\omega\xi} R_X(\tau - \xi) d\xi \right] \\ &\quad \times e^{-j\omega(\tau - \xi)} d\tau \end{aligned} \quad (22.2.14)$$

Substituting  $\tau - \xi = \beta$  in Eq. (22.2.14), we obtain

$$\int_0^\infty R_{XS}(\tau) e^{-j\omega\tau} d\tau = \int_0^\infty h_{\text{opt}}(\xi) e^{-j\omega\xi} d\xi \int_{-\xi}^\infty R_X(\beta) e^{-j\omega\beta} d\beta \quad (22.2.15)$$

The presence of  $\xi$  in the lower limit of the RHS of Eq. (22.2.15) poses a serious problem since it is a variable involved in the integration. We have to seek other methods for solving the *Weiner–Hopf integral* equation.

We note that Eq. (22.2.11) or (22.2.12) is not dependent on the exact nature of the signal  $X(t)$  and  $S(t)$  but only on their autocorrelation functions, which are not unique.

Thus, other signals that have the same autocorrelation functions will give the same result.

Equation (22.2.12) will be solved by *spectral factorization* techniques for which we require some intermediate results.

### Whitening Filter

We are given an input stationary random process  $X(t)$  and want to find a causal impulse response  $g(t)u(t)$  such that the output process  $Z(t)$  given by

$$Z(t) = \int_0^\infty g(\tau) X(t - \tau) d\tau \quad (22.2.16)$$

is a white-noise process with autocorrelation  $R_Z(\tau) = \delta(\tau)$ .

We will make the assumption that the psd  $S_X(\omega)$  of  $X(t)$  is a real rational function of  $\omega^2$  for real  $\omega$  so that all the poles and zeros of  $S_X(\omega)$  in the complex  $\omega$  plane occur as complex conjugate pairs. From Eq. (21.2.9) we have

$$S_Z(\omega) = 1 = S_X(\omega) |G(\omega)|^2 \quad (22.2.17)$$

Hence the required transfer function is given by

$$\frac{1}{G(\omega)} \cdot \frac{1}{G^*(\omega)} = S_X(\omega) \quad (22.2.18)$$

where  $G(\omega)$  has poles in the top half and  $G^*(\omega)$  has poles in the bottom half of the  $\omega$  plane. The corresponding impulse responses  $g(t)$  and  $g^{-1}(t)$ , given by

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) e^{j\omega t} d\omega; \quad g^{-1}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{G(\omega)} e^{j\omega t} d\omega \quad (22.2.19)$$

satisfy the causality and stability criteria:

$$g(t) \text{ and } g^{-1}(t) = 0, \quad t < 0; \quad g(t) \text{ and } g^{-1}(t) \rightarrow 0 \text{ as } t \rightarrow \infty.$$

Under these conditions  $X(t)$  can be generated from the following equation:

$$X(t) = \int_0^{\infty} g^{-1}(\tau) Z(t - \tau) d\tau \quad (22.2.20)$$

### Spectral Factorization

As mentioned in the previous section, the psd  $S_X(\omega)$  is a rational function of  $\omega^2$ , and some properties are listed in Chapter 19, Eqs. (19.5.8)–(19.5.11). Additional properties relevant to Wiener filtering will be enumerated:

1.  $S_X(\omega)$  is bounded because  $X(t)$  is assumed to be a zero mean stationary process:

$$S_X(\omega) < \infty \quad (22.2.21)$$

2. As a consequence of property 1, poles of  $S_X(\omega)$  cannot be real.
3. Since  $R_X(\tau)$  is real,  $S_X(\omega) = S_X^*(\omega)$  and hence it has complex conjugate poles and zeros (symmetry about the real axis).
4. From Eq. (19.5.8),  $S_X(\omega) = S_X(-\omega)$  and hence it has symmetry about the imaginary axis.
5. As a consequence of properties 3 and 4, the poles and zeros have symmetry about both axes as shown in Fig. 22.2.5.

Two methods of factoring  $S_X(\omega)$  will be discussed. The first is called *multiplicative factoring*, given by

$$S_X(\omega) = k \frac{(\omega - z_1)(\omega - z_2) \cdots (\omega - z_M)}{(\omega - p_1)(\omega - p_2) \cdots (\omega - p_N)} \cdot k \frac{(\omega - z_1^*)(\omega - z_2^*) \cdots (\omega - z_M^*)}{(\omega - p_1^*)(\omega - p_2^*) \cdots (\omega - p_N^*)} \quad (22.2.22)$$

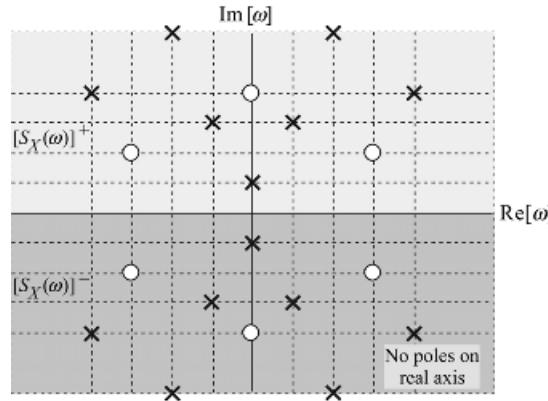


FIGURE 22.2.5

where  $p_n$ ,  $n = 1, \dots, N$  are poles with positive imaginary parts and  $z_n$ ,  $n = 1, \dots, M$  are zeros with positive imaginary parts. Thus,  $S_X(\omega)$  can be represented by

$$S_X(\omega) = [S_X(\omega)]^+ [S_X(\omega)]^- \quad (22.2.23)$$

where  $[S_X(\omega)]^+$  has poles and zeros on the upper and  $[S_X(\omega)]^-$  has poles and zeros in the lower half of the  $\omega$  plane as shown in Fig. 22.2.5. The second one is *additive factoring*, where we have a partial fraction expansion and separate the positive and negative roots in the top and bottom halves of the  $\omega$  plane. In this case  $S_X(\omega)$  can be represented by

$$S_X(\omega) = [S_X(\omega)]_+ + [S_X(\omega)]_- \quad (22.2.24)$$

In multiplicative factoring poles and zeros are involved, whereas in additive factoring only the poles are involved.

We can now identify from Eq. (22.2.18) the transfer functions  $G(\omega)$  and  $G^*(\omega)$  as follows:

$$G(\omega) = \frac{1}{[S_X(\omega)]^+}; \quad G^*(\omega) = \frac{1}{[S_X(\omega)]^-} \quad (22.2.25)$$

**Example 22.2.3** The psd  $S_X(\omega)$  is given by

$$S_X(\omega) = 25 \frac{(\omega^2 + 4)}{(\omega^2 + 9)(\omega^2 + 16)}$$

The multiplicative factoring is obtained from the form of  $S_X(\omega)$  and is given by

$$[S_X(\omega)]^+ = 5 \frac{(\omega + 2j)}{(\omega + 3j)(\omega + 4j)}; \quad [S_X(\omega)]^- = 5 \frac{(\omega - 2j)}{(\omega - 3j)(\omega - 4j)}$$

The additive factoring is obtained by obtaining a partial fraction expansion and expressing  $S_X(\omega)$  as

$$\begin{aligned} S_X(\omega) &= \frac{25}{7} \left( \frac{12}{\omega^2 + 16} - \frac{5}{\omega^2 + 9} \right) \\ &= \frac{25}{7} \left\{ \frac{3}{2(4 + j\omega)} - \frac{5}{6(3 + j\omega)} + \frac{3}{2(4 - j\omega)} - \frac{5}{6(3 - j\omega)} \right\} \end{aligned}$$

and identifying

$$\begin{aligned} [S_X(\omega)]_+ &= \frac{25}{7} \left( \frac{3}{2(4 + j\omega)} - \frac{5}{6(3 + j\omega)} \right); \\ [S_X(\omega)]_- &= \frac{25}{7} \left( \frac{3}{2(4 - j\omega)} - \frac{5}{6(3 - j\omega)} \right) \end{aligned}$$

The differences between these two factorings are clear.

### **Solution for Wiener–Hopf Filter Equation [Eq. (22.2.12)]**

The method for finding the causal solution is to pass the input process  $X(t)$  through the causal whitening filter  $g(t)$  of Eq. (22.2.16), and then through another optimal causal filter  $g_{\text{opt}}(t)$  so that these two filters are in tandem, will yield the desired  $h_{\text{opt}}(t)$ , which will yield

the optimum  $Y_o(t)$  as shown in Fig. 22.2.6:

$$Y_o(t) = g_{\text{opt}}(t) * g(t) * X(t) \quad (22.2.26)$$

From Eq. (22.2.26) the filter  $g_{\text{opt}}(t)$  can be obtained as

$$h_{\text{opt}}(t) = g(t) * g_{\text{opt}}(t) \quad \text{or} \quad g_{\text{opt}}(t) = h_{\text{opt}}(t) * g^{-1}(t) \quad (22.2.27)$$

Taking FT on both sides of Eq. (22.2.27), we obtain

$$H_{\text{opt}}(\omega) = G(\omega)G_{\text{opt}}(\omega) \quad \text{or} \quad G_{\text{opt}}(\omega) = \frac{H_{\text{opt}}(\omega)}{G(\omega)} \quad (22.2.28)$$

The filter  $g_{\text{opt}}(t)$  is the optimal filter obtained by passing an input white noise to obtain the desired optimal response  $Y_o(t)$ . The Wiener–Hopf equation corresponding to  $g_{\text{opt}}(t)$  is obtained from Eq. (22.2.2) with  $\alpha = 0$ :

$$\int_0^\infty g_{\text{opt}}(\xi)R_Z(\tau - \xi)d\xi = R_{ZS}(\tau), \quad \tau \geq 0 \quad (22.2.29)$$

In Eq. (22.2.29), we have

$$R_Z(\tau - \xi) = \delta(\tau - \xi); \quad R_{ZS}(\tau) = R_{XS}(\tau) * g(-\tau) \quad (22.2.30)$$

Substituting Eq. (22.2.30) into Eq. (22.2.29) yields

$$g_{\text{opt}}(\tau) = R_{XS}(\tau) * g(-\tau), \quad \tau \geq 0 \quad (22.2.31)$$

In other words,  $g_{\text{opt}}(\tau)$  is the causal part of the noncausal term  $R_{XS}(\tau) * g(-\tau)$ . We can define the FT of  $R_{XS}(\tau) * g(-\tau)$  in Eq. (22.2.31) as  $K(\omega)$ . From the partial fraction expansion,  $K(\omega)$  can be additively factored by using Eq. (22.2.24) as  $K(\omega) = G_{\text{opt}}(\omega) + G_{\text{opt,L}}(\omega)$ , where  $G_{\text{opt}}(\omega)$  has poles in the upper and  $G_{\text{opt,L}}(\omega)$  has poles in the lower half of the  $\omega$  plane, or

$$K(\omega) = S_{XS}(\omega)G^*(\omega) = G_{\text{opt}}(\omega) + G_{\text{opt,L}}(\omega) \quad (22.2.32)$$

where

$$G_{\text{opt}}(\omega) = [S_{XS}(\omega)G^*(\omega)]_+ = \left[ \frac{S_{XS}(\omega)}{[S_X(\omega)]^-} \right]_+ \quad (22.2.33a)$$

$$G_{\text{opt,L}}(\omega) = [S_{XS}(\omega)G^*(\omega)]_- = \left[ \frac{S_{XS}(\omega)}{[S_X(\omega)]^-} \right]_- \quad (22.2.33b)$$

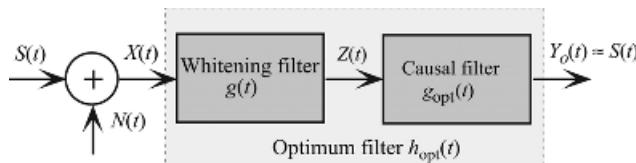


FIGURE 22.2.6

and

$$\text{IFT}[K(\omega)] = k(t) = g_{\text{opt}}(t)u(t) + g_{\text{opt,L}}(-t)u(-t) \quad (22.2.34)$$

Substituting Eqs. (22.2.25) and (22.2.33a) in Eq. (22.2.28), we obtain the final form for the optimal causal transfer function  $H_{\text{opt}}(\omega)$  as

$$H_{\text{opt}}(\omega) = \frac{1}{[S_X(\omega)]^+} \left[ \frac{S_{XS}(\omega)}{[S_X(\omega)]^-} \right]_+ \quad (22.2.35)$$

The minimum mean-square error associated with the causal optimal impulse response  $h_{\text{opt}}(t)$  is given by Eq. (22.2.13):

$$J_{\min} = R_S(0) - \int_0^\infty h_{\text{opt}}(\tau)R_{XS}(\tau)d\tau \quad (22.2.13)$$

The integral in the minimum mean-square error [Eq. (21.2.13)] can be written in the frequency domain using Parseval's theorem:

$$\begin{aligned} \int_0^\infty h_{\text{opt}}(\tau)R_{XS}(\tau)d\tau &= \frac{1}{2\pi} \int_{-\infty}^\infty H_{\text{opt}}(\omega)S_{XS}(\omega)d\omega \quad (\text{Parseval's theorem}) \\ &= \frac{1}{2\pi} \int_{-\infty}^\infty G_{\text{opt}}(\omega)[G(\omega)S_{XS}(\omega)]d\omega \quad [\text{Eq. (22.2.33)}] \\ &= \frac{1}{2\pi} \int_{-\infty}^\infty G_{\text{opt}}(\omega) \left[ \frac{S_{XS}(\omega)}{[S_X(\omega)]^+} \right]_- d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^\infty G_{\text{opt}}(\omega)G_{\text{opt}}^*(\omega)d\omega \\ &= \int_0^\infty g_{\text{opt}}^2(\tau)d\tau \quad (\text{Parseval's theorem}) \end{aligned} \quad (22.2.36)$$

Substituting Eq. (22.2.36) in Eq. (22.2.13), the minimum mean-square error  $J_{\min}$  can be written in terms of the optimal filter  $g_{\text{opt}}(t)$  for a white-noise input as

$$J_{\min} = R_S(0) - \int_0^\infty g_{\text{opt}}^2(\tau)d\tau \quad (22.2.37)$$

If signal  $S(t)$  and noise  $N(t)$  are independent, then Eqs. (22.2.35) and (22.2.13) can be further simplified using Eq. (22.2.4). Equation (22.2.35) can be written as

$$H_{\text{opt}}(\omega) = \frac{1}{[S_S(\omega) + S_N(\omega)]^+} \left[ \frac{S_S(\omega)}{[S_S(\omega) + S_N(\omega)]^-} \right]_+ \quad (22.2.38)$$

Equation (22.2.13) can be written as

$$J_{\min} = R_S(0) - \int_0^\infty h_{\text{opt}}(\tau)R_S(\tau)d\tau \quad (22.2.39)$$

Further simplifications are possible [21] if additionally  $N(t)$  is white noise with psd  $N_0$ . Under this condition and from Eq. (22.2.4) we have

$$S_{XS}(\omega) = S_S(\omega) = S_X(\omega) - N_0 \quad (22.2.40)$$

Substituting Eq. (22.2.40) in filter equation (22.2.35), we can write

$$\begin{aligned}
 H_{\text{opt}}(\omega) &= \frac{1}{[S_X(\omega)]^+} \left[ \frac{S_X(\omega)}{[S_X(\omega)]^-} - \frac{N_0}{[S_X(\omega)]^-} \right]_+ \\
 &= \frac{1}{[S_X(\omega)]^+} \left[ \frac{[S_X(\omega)]^+ [S_X(\omega)]^-}{[S_X(\omega)]^-} - \frac{[N_0]^+ [N_0]^-}{[S_X(\omega)]^-} \right]_+ \\
 &= \frac{1}{[S_X(\omega)]^+} \left[ [S_X(\omega)]^+ - \frac{[N_0]^+ [N_0]^-}{[S_X(\omega)]^-} \right]_+ \\
 &= 1 - \frac{[N_0]^+}{[S_S(\omega) + N_0]^+} \left[ \left( \frac{N_0}{S_S(\omega) + N_0} \right)^- \right]_+ \quad (22.2.41)
 \end{aligned}$$

The multiplicative factoring yields  $[N_0]^+ = \sqrt{N_0}$  and the factor within the braces in Eq. (22.2.41) can be simplified as follows:

$$\frac{N_0}{S_S(\omega) + N_0} = \frac{N_0}{\frac{N(\omega)}{D(\omega)} + N_0} = \frac{N_0 D(\omega)}{N(\omega) + N_0 D(\omega)} \quad (22.2.42)$$

In Eq. (22.2.42) the degree of  $N(\omega)$  is less than or equal to the degree of  $D(\omega)$ , and hence

$$\left[ \left( \frac{N_0 D(\omega)}{N(\omega) + N_0 D(\omega)} \right)^- \right]_+ = 1 + [(A(\omega))^-]_+ = 1 \quad (22.2.43)$$

Since all the poles of  $(A(\omega))^-$  are in the lower half of the  $\omega$  plane, the factor  $[(A(\omega))^-]_+ = 0$  in Eq. (22.2.43). Substituting Eq. (22.2.43) in Eq. (22.2.41), we obtain a simplified expression for the optimal filter:

$$H_{\text{opt}}(\omega) = 1 - \frac{\sqrt{N_0}}{[S_S(\omega) + N_0]^+} \quad (22.2.44)$$

In a similar manner, Eq. (22.2.39) can also be simplified. Since  $R_X(\tau) = R_S(\tau) + N_0 \delta(\tau)$  the Wiener–Hopf equation (22.2.12) can be written as

$$\int_0^\infty h_{\text{opt}}(\xi) R_S(\tau - \xi) d\xi + N_0 h_{\text{opt}}(\tau) = R_S(\tau), \quad \tau \geq 0 \quad (22.2.45)$$

and Eq. (22.2.45) evaluated at  $\tau = 0$  becomes

$$\int_0^\infty h_{\text{opt}}(\xi) R_S(\xi) d\xi = R_S(0) - N_0 h_{\text{opt}}(0) \quad (22.2.46)$$

Substituting Eq. (22.2.46) in Eq. (22.2.39) we obtain

$$J_{\text{min}} = N_0 h_{\text{opt}}(0) \quad (22.2.47)$$

### Solution for Wiener–Hopf Prediction–Smoothing Equation

The form of the Wiener–Hopf (W–H) equation for prediction and smoothing is obtained by replacing  $R_{XS}(\tau)$  with  $R_{XS}(\tau + \alpha)$  in Eq. (22.2.12):

$$\int_0^\infty h_{\text{opt}}(\xi) R_X(\tau - \xi) d\xi = R_{XS}(\tau + \alpha), \quad \tau \geq 0 \quad (22.2.48)$$

As mentioned earlier, if  $\alpha = 0$ , we have a filtering problem whose solution is given by Eq. (22.2.35). If  $\alpha > 0$ , we have a prediction problem, and if  $\alpha < 0$ , we have a smoothing problem. Corresponding to the factor  $\alpha$ , the optimal transfer function will be given by

$$H_{\text{opt}}(\omega) = \frac{1}{[S_X(\omega)]^+} \left[ \frac{e^{j\omega\alpha} S_{XS}(\omega)}{[S_X(\omega)]^-} \right]_+ = \frac{1}{[S_X(\omega)]^+} [K(\omega) e^{j\omega\alpha}]_+ \quad (22.2.49)$$

and the minimum mean-square error by

$$J_{\text{min}} = R_S(0) - \int_0^\infty h_{\text{opt}}(\tau) R_{XS}(\tau + \alpha) d\tau \quad (22.2.50)$$

Similar to Eq. (22.2.34), we can define the following IFT

$$\begin{aligned} \text{IFT} \left[ \frac{e^{j\omega\alpha} S_{XS}(\omega)}{[S_X(\omega)]^-} \right] &= \text{IFT}[e^{j\omega\alpha} K(\omega)] \\ &= k(t + \alpha) = g_{\text{opt}}(t + \alpha)u(t + \alpha) + g_{\text{opt,L}}(-t - \alpha)u(-t - \alpha) \end{aligned} \quad (22.2.51)$$

and the minimum mean-square error  $J_{\text{min}}$  will be given by an equation similar to Eq. (22.2.37) as

$$J_{\text{min}} = R_S(0) - \int_0^\infty k^2(\tau + \alpha) d\tau = R_S(0) - \int_0^\infty k^2(\tau) d\tau \quad (22.2.52)$$

From Eq. (22.2.52) we can observe that the error is a maximum when the prediction time  $\alpha = \infty$ , and a minimum for infinite lag time  $\alpha = -\infty$ . When  $\alpha = 0$ , we have Eq. (22.2.37).

We will now determine a realizable filter for Examples 22.2.1 and 22.2.2 with the same assumption that  $S(t)$  and the noise  $N(t)$  are independent and both are zero mean stationary random processes.

**Example 22.2.4** We will consider Example 22.2.1, where the autocorrelation functions of the input signal  $S(t)$  and the noise  $N(t)$  to a system are

$$R_S(\tau) = e^{-(|\tau|/2)} \quad \text{and} \quad R_N(\tau) = 4\delta(\tau)$$

The power spectral densities (psd's) of signal and noise are

$$S_S(\omega) = \frac{4}{1 + 4\omega^2} \quad \text{and} \quad S_N(\omega) = 4$$

**1. Filtering Problem:**  $\alpha = 0$ . We will first estimate the signal  $S(t)$  by a realizable filter  $H_{\text{opt}}(\omega)$  using Eq. (22.2.38):

$$\begin{aligned} S_S(\omega) + S_N(\omega) &= \frac{8 + 16\omega^2}{1 + 4\omega^2} = 2 \frac{\sqrt{2} + 2j\omega}{1 + 2j\omega} \cdot 2 \frac{\sqrt{2} - 2j\omega}{1 - 2j\omega} \\ G(\omega) &= \frac{1}{[S_S(\omega) + S_N(\omega)]^+} = \frac{1 + 2j\omega}{2(\sqrt{2} + 2j\omega)} = \frac{1}{2} - \frac{\sqrt{2} - 1}{2(\sqrt{2} - 2j\omega)} \\ G^*(\omega) &= \frac{1}{[S_S(\omega) + S_N(\omega)]^-} = \frac{1 - 2j\omega}{2(\sqrt{2} - 2j\omega)} = \frac{1}{2} - \frac{\sqrt{2} - 1}{2(\sqrt{2} - 2j\omega)} \end{aligned}$$

We find  $K(\omega)$  as follows:

$$\begin{aligned} K(\omega) &= \frac{S_S(\omega)}{[S_S(\omega) + S_N(\omega)]^-} = S_S(\omega)G^*(\omega) \\ &= \frac{1 - 2j\omega}{2(\sqrt{2} - 2j\omega)} \cdot \frac{4}{1 + 4\omega^2} = \frac{2}{(\sqrt{2} - 2j\omega)(1 + 2j\omega)} \end{aligned}$$

Partial fraction expansion of  $K(\omega)$  yields

$$K(\omega) = \frac{2(\sqrt{2} - 1)}{(1 + 2j\omega)} + \frac{2(\sqrt{2} - 1)}{(\sqrt{2} - 2j\omega)} \quad (\text{A})$$

Taking the IFT of Eq. (A) above, we obtain

$$k(t) = g_{\text{opt}}(t)u(t) + g_{\text{opt,L}}(-t)u(-t) = (\sqrt{2} - 1) \left[ e^{-(t/2)}u(t) + e^{(t/\sqrt{2})}u(-t) \right] \quad (\text{B})$$

We can identify  $G_{\text{opt}}(\omega)$  and  $G_{\text{opt,L}}(\omega)$  as follows:

$$\begin{aligned} G_{\text{opt}}(\omega) &= \left[ \frac{S_S(\omega)}{[S_S(\omega) + S_N(\omega)]^-} \right]_+ = [K(\omega)]_+ = \frac{2(\sqrt{2} - 1)}{(1 + 2j\omega)} \\ G_{\text{opt,L}}(\omega) &= \left[ \frac{S_S(\omega)}{[S_S(\omega) + S_N(\omega)]^-} \right]_- = [K(\omega)]_- = \frac{2(\sqrt{2} - 1)}{(\sqrt{2} - 2j\omega)} \end{aligned}$$

Hence  $H_{\text{opt}}(\omega)$  is given by

$$H_{\text{opt}}(\omega) = G(\omega)G_{\text{opt}}(\omega) = \frac{1 + 2j\omega}{2(\sqrt{2} + 2j\omega)} \cdot \frac{2(\sqrt{2} - 1)}{(1 + 2j\omega)} = \frac{\sqrt{2} - 1}{\sqrt{2} + 2j\omega}$$

This result can also be obtained from Eq. (22.2.44) as follows. Substituting  $[S_S(\omega) + N_0]^+ = 2[(\sqrt{2} + 2j\omega)/(1 + 2j\omega)]$  and  $\sqrt{N_0} = 2$  in Eq. (22.2.44), we obtain

$$H_{\text{opt}}(\omega) = 1 - \frac{2(1 + 2j\omega)}{2(\sqrt{2} + 2j\omega)} = \frac{\sqrt{2} + 2j\omega - 1 - 2j\omega}{\sqrt{2} + 2j\omega} = \frac{\sqrt{2} - 1}{\sqrt{2} + 2j\omega}$$

which is the same as before.

The inverse Fourier transforms of  $G(\omega)$ ,  $G^*(\omega)$ , and  $H_{\text{opt}}(\omega)$  are given by

$$\begin{aligned} g(t) &= \frac{1}{2} \delta(t) - \frac{\sqrt{2} - 1}{4} e^{-(t/\sqrt{2})}u(t); \quad g(-t) = \frac{1}{2} \delta(t) - \frac{\sqrt{2} - 1}{4} e^{-(t/\sqrt{2})}u(-t) \\ h_{\text{opt}}(t) &= \frac{\sqrt{2} - 1}{2} e^{-(t/\sqrt{2})}u(t) \end{aligned}$$

The time function  $k(t)$  in Eq. (B), and  $g(t)$  and  $g(-t)$  corresponding to the poles and zeros of  $[S_S(\omega)]^+$  and  $[S_S(\omega)]^-$ , and  $h_{\text{opt}}(t)$ , the optimum realizable impulse response, are shown in Fig. 22.2.7. To determine the minimum mean-square error  $J_{\min}$ , we need the function  $k^2(t)$ , which is given by

$$k^2(t) = (\sqrt{2} - 1)^2 [e^{-t}u(t) + e^{\sqrt{2}t}u(-t)] \quad (\text{C})$$

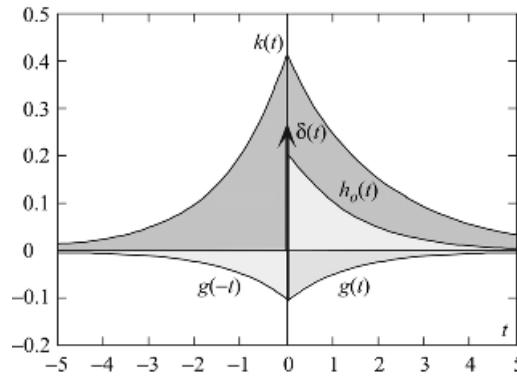


FIGURE 22.2.7

The minimum mean-square error is calculated by substituting the causal part of Eq. (C) in Eq. (22.2.52):

$$J_{\min} = R_S(0) - \int_0^{\infty} k^2(\tau) d\tau = 1 - (\sqrt{2} - 1)^2 \int_0^{\infty} e^{-\tau} d\tau = 1 - (\sqrt{2} - 1)^2 = 2(\sqrt{2} - 1)$$

We can also calculate  $J_{\min}$  using Eqs. (22.2.39) or (22.2.47):

$$\begin{aligned} J_{\min} &= R_S(0) - \int_0^{\infty} h_{\text{opt}}(\tau) R_S(\tau) d\tau = 1 - \int_0^{\infty} \frac{1}{2 + \sqrt{8}} e^{-(t/\sqrt{2})} e^{-(t/2)} dt \\ &= 1 - \left[ (2\sqrt{2} - 3) e^{-(t/\sqrt{2})} e^{-(t/2)} \right]_0^{\infty} = 2(\sqrt{2} - 1) \\ &= N_0 h_{\text{opt}}(0) = \frac{4}{2 + \sqrt{8}} = \frac{2}{1 + \sqrt{2}} = 2(\sqrt{2} - 1) \end{aligned}$$

All three equations yield the same result:  $J_{\min} = 2(\sqrt{2} - 1) = 0.8284$ .

The causal part of the noncausal solution of Example 22.2.1 given by  $h_{\text{opt}} \text{noncausal}(t) = \frac{\sqrt{2}}{8} e^{-(t/\sqrt{2})} u(t)$  and the causal solution of this example,  $h_{\text{opt}} \text{causal}(t) = [(\sqrt{2} - 1/2)] e^{-(1/\sqrt{2})} u(t)$ , are shown in Fig. 22.2.8. The minimum mean-square errors in

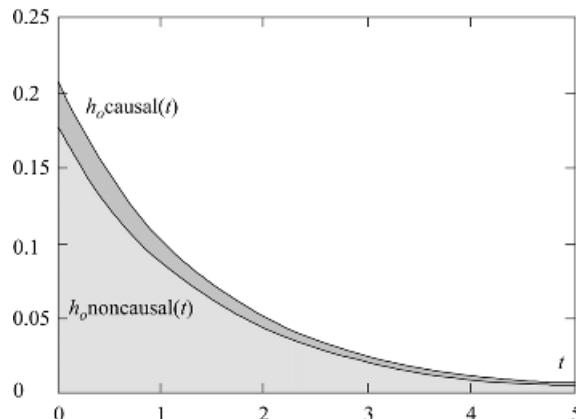


FIGURE 22.2.8

both examples are given by  $J_{\min \text{ noncausal}} = 0.7071$  and  $J_{\min \text{ causal}} = 0.8284$ . The differences between the causal and noncausal solutions are not that significant for this problem, but the complexity of getting a causal solution is much greater. From  $K(\omega)$  in Eq. (A) for a delay  $\alpha$ , we can write

$$\begin{aligned} K(\omega)e^{j\omega\alpha} &= \left[ \frac{2(\sqrt{2}-1)}{(1+2j\omega)} + \frac{2(\sqrt{2}-1)}{(\sqrt{2}-2j\omega)} \right] e^{j\omega\alpha} \\ \iff (\sqrt{2}-1) &\left[ e^{-(t+\alpha)/2}u(t+\alpha) + e^{(t+\alpha)/\sqrt{2}}u(-t-\alpha) \right] \end{aligned}$$

and the time function  $k(t+\alpha)$  is shown for both  $\alpha > 0$  and  $\alpha < 0$  in Fig. 22.2.9.

**2. Prediction Problem:**  $\alpha > 0$ . For the prediction problem of  $\alpha > 0$ , the causal part of  $k(t+\alpha)$  from Fig. 22.2.9 is given by

$$k(t+\alpha)_{\text{causal}} = (\sqrt{2}-1)e^{-(t+\alpha)/2}u(t), \quad \alpha > 0$$

and the FT of this equation yields,

$$[K(\omega)e^{j\omega\alpha}]_+ = \frac{(\sqrt{2}-1)e^{-(\alpha/2)}}{\frac{1}{2} + j\omega} = \frac{2(\sqrt{2}-1)e^{-(\alpha/2)}}{1 + 2j\omega}$$

and substituting the preceding equation in Eq. (22.2.49), we obtain the optimum transfer function for the prediction problem

$$\begin{aligned} H_{\text{opt}}(\omega) &= G(\omega)[K(\omega)e^{j\omega\alpha}]_+ = \frac{1+2j\omega}{2(\sqrt{2}+2j\omega)} \cdot \frac{2(\sqrt{2}-1)e^{-(\alpha/2)}}{1+2j\omega} \\ &= \frac{(\sqrt{2}-1)e^{-(\alpha/2)}}{\sqrt{2}+2j\omega} \end{aligned}$$

and the optimum impulse response given by

$$h_{\text{opt}}(t) = \frac{\sqrt{2}-1}{2} e^{-(1/\sqrt{2})[t+(\alpha/2)]} u(t), \quad \alpha > 0$$

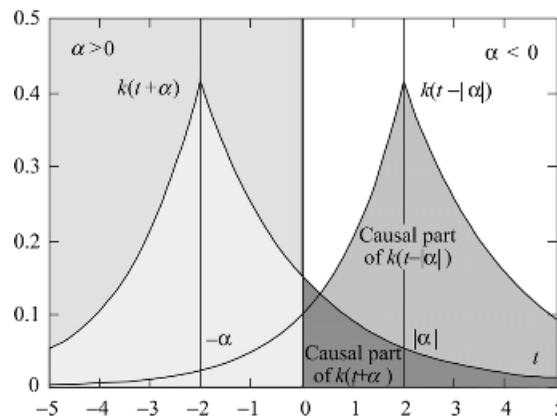


FIGURE 22.2.9

Since  $\alpha > 0$ , the minimum mean-square error is obtained by substituting the causal part of Eq. (C),  $k^2(t) = (\sqrt{2} - 1)^2 e^{-t} u(t)$ , in Eq. (22.2.52) to obtain

$$J_{\min} = R_S(0) - \int_{\alpha}^{\infty} k^2(\tau) d\tau = 1 - \int_{\alpha}^{\infty} (\sqrt{2} - 1)^2 e^{-\tau} d\tau = 1 - (\sqrt{2} - 1)^2 e^{-\alpha}, \quad \alpha > 0$$

**3. Smoothing Problem:**  $\alpha < 0$ . For the smoothing problem of  $\alpha < 0$ , the causal part of  $k(t + \alpha)$  from Fig. 22.2.9 is

$$k(t + \alpha)_{\text{causal}} = (\sqrt{2} - 1) \left[ e^{-[(t+\alpha)/2]} u(t + \alpha) + e^{[(t+\alpha)/\sqrt{2}]} [u(-t - \alpha) - u(-t)] \right], \quad \alpha < 0$$

and the FT of this equation yields

$$\begin{aligned} [K(\omega) e^{j\omega\alpha}]_+ &= 2(\sqrt{2} - 1) \left[ \frac{e^{j\omega\alpha}}{1 + 2j\omega} + \frac{e^{j\omega\alpha} - e^{\alpha/\sqrt{2}}}{\sqrt{2} - 2j\omega} \right] \\ &= 2(\sqrt{2} - 1) \left[ \frac{(\sqrt{2} + 1)e^{j\omega\alpha} - (1 + 2j\omega)e^{\alpha/\sqrt{2}}}{(1 + 2j\omega)(\sqrt{2} - 2j\omega)} \right] \end{aligned}$$

and substituting this equation in Eq. (22.2.49), we obtain the optimum transfer function for the smoothing problem:

$$\begin{aligned} H_{\text{opt}}(\omega) &= \frac{1 + 2j\omega}{2(\sqrt{2} + 2j\omega)} 2(\sqrt{2} - 1) \left[ \frac{(\sqrt{2} + 1)e^{j\omega\alpha} - (1 + 2j\omega)e^{\alpha/\sqrt{2}}}{(1 + 2j\omega)(\sqrt{2} - 2j\omega)} \right] \\ &= (\sqrt{2} - 1) \left[ \frac{(\sqrt{2} + 1)e^{j\omega\alpha} - (1 + 2j\omega)e^{\alpha/\sqrt{2}}}{(2 + 4\omega^2)} \right] \end{aligned}$$

To obtain the minimum mean-square error, we note that since  $\alpha < 0$  we have to use the entire Eq. (C) in Eq. (22.2.52) and write

$$\begin{aligned} J_{\min} &= R_S(0) - \int_{\alpha}^{\infty} k^2(\tau) d\tau = 1 - (\sqrt{2} - 1)^2 \left[ \int_{\alpha}^0 e^{\sqrt{2}\tau} d\tau + \int_0^{\infty} e^{-\tau} d\tau \right] \\ &= 1 - (\sqrt{2} - 1)^2 \left[ \frac{1}{\sqrt{2}} (1 - e^{\sqrt{2}\alpha}) + 1 \right], \quad \alpha < 0 \end{aligned}$$

The graph of the minimum mean-square error  $J_{\min}(\alpha)$  for values of  $\alpha$  ranging from  $-\infty$  to  $\infty$  is shown in Fig. 22.2.10. The MSE  $J_{\min}$  reaches a minimum value for infinite lag, and this value is given by

$$\begin{aligned} J_{\min}(-\infty) &= 1 - (\sqrt{2} - 1)^2 \left[ \frac{1}{\sqrt{2}} (1 - e^{\sqrt{2}\alpha}) + 1 \right] \Big|_{\alpha \rightarrow -\infty} \\ &= 1 - (\sqrt{2} - 1)^2 \left[ \frac{\sqrt{2} + 1}{\sqrt{2}} \right] = \frac{1}{\sqrt{2}} \end{aligned}$$

and the maximum error for infinite prediction time is

$$J_{\min}(\infty) = 1 - [(\sqrt{2} - 1)^2 e^{-\alpha}] \Big|_{\alpha \rightarrow \infty} = 1$$

**Example 22.2.5** This example will find a causal filter for Example 22.2.2, where the autocorrelation functions of the input signal  $S(t)$  and the independent noise  $N(t)$  to a

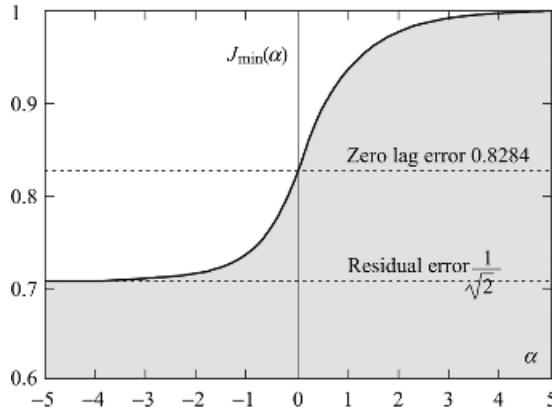


FIGURE 22.2.10

system are given by

$$R_S(\tau) = e^{-(|\tau|/2)} \quad \text{and} \quad R_N(\tau) = e^{-2|\tau|}$$

The power spectral densities of signal and noise are

$$S_S(\omega) = \frac{4}{1 + 4\omega^2} \quad \text{and} \quad S_N(\omega) = \frac{4}{4 + \omega^2}$$

**1. Filtering Problem:**  $\alpha > 0$ . From Example 22.2.4 we can write

$$\begin{aligned} S_X(\omega) &= S_S(\omega) + S_N(\omega) = 20 \frac{1 + \omega^2}{(1 + 4\omega^2)(4 + \omega^2)} \\ &= \frac{2\sqrt{5}(1 + j\omega)}{(1 + 2j\omega)(2 + j\omega)} \frac{2\sqrt{5}(1 - j\omega)}{(1 - 2j\omega)(2 - j\omega)} \\ G(\omega) &= \frac{(1 + 2j\omega)(2 + j\omega)}{2\sqrt{5}(1 + j\omega)}; \quad G^*(\omega) = \frac{(1 - 2j\omega)(2 - j\omega)}{2\sqrt{5}(1 - j\omega)} \end{aligned}$$

We find  $K(\omega)$  as follows:

$$K(\omega) = S_S(\omega)G^*(\omega) = \frac{4}{1 + 4\omega^2} \frac{(1 - 2j\omega)(2 - j\omega)}{2\sqrt{5}(1 - j\omega)} = \frac{2(2 - j\omega)}{\sqrt{5}(1 + 2j\omega)(1 - j\omega)}$$

Partial fraction expansion of  $K(\omega)$  yields

$$K(\omega) = \frac{2\sqrt{5}}{3} \left[ \frac{1}{(1 + 2j\omega)} + \frac{1}{5(1 - j\omega)} \right] \quad (\text{A})$$

Taking the IFT of Eq. (A), we obtain

$$k(t) = g_{\text{opt}}(t)u(t) + g_{\text{opt,L}}(-t)u(-t) = \frac{\sqrt{5}}{3} \left[ e^{-(t/2)}u(t) + \frac{2}{5}e^t u(-t) \right] \quad (\text{B})$$

$k(t)$  is shown in Fig. 22.2.11.

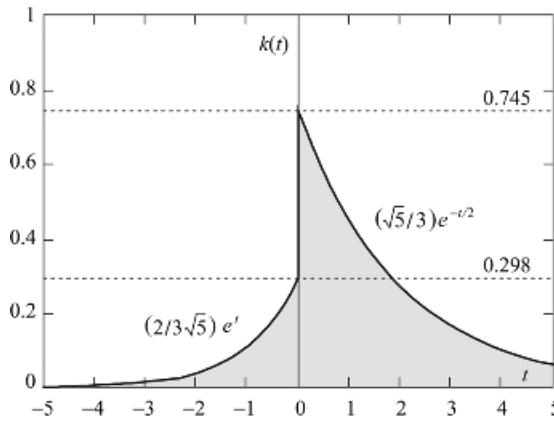


FIGURE 22.2.11

We can identify  $G_{\text{opt}}(\omega)$  and  $G_{\text{opt,L}}(\omega)$  as

$$G_{\text{opt}}(\omega) = \left[ \frac{S_S(\omega)}{[S_S(\omega) + S_N(\omega)]^-} \right]_+ = [K(\omega)]_+ = \frac{2\sqrt{5}}{3(1+2j\omega)}$$

$$G_{\text{opt,L}}(\omega) = \left[ \frac{S_S(\omega)}{[S_S(\omega) + S_N(\omega)]^-} \right]_- = [K(\omega)]_- = \frac{2\sqrt{5}}{15(1-j\omega)}$$

Hence  $H_{\text{opt}}(\omega)$  is given by

$$H_{\text{opt}}(\omega) = G(\omega)G_{\text{opt}}(\omega) = \frac{(1+2j\omega)(2+j\omega)}{2\sqrt{5}(1+j\omega)} \cdot \frac{2\sqrt{5}}{3(1+2j\omega)} = \frac{(2+j\omega)}{3(1+j\omega)}$$

The inverse Fourier transforms of  $G(\omega)$ ,  $G^*(\omega)$ , and  $H_{\text{opt}}(\omega)$  are given by

$$g(t) = \frac{-\sqrt{5}}{10} e^{-t} u(t) - 3\delta(t) + 2\delta'(t); \quad g(-t) = \frac{-\sqrt{5}}{10} e^t u(-t) - 3\delta(t) - 2\delta'(t)$$

$$h_{\text{opt}}(t) = \frac{1}{3} e^{-t} u(t) + \frac{1}{3} \delta(t)$$

The optimum impulse response may not be realizable because of the presence of the impulse function.

To determine the minimum mean-square error  $J_{\min}$ , we need the function  $k^2(t)$ , which is given by

$$k^2(t) = \frac{5}{9} \left[ e^{-t} u(t) + \frac{4}{25} e^{2t} u(-t) \right] \quad (\text{C})$$

The minimum mean-square error is calculated by substituting the causal part of Eq. (C) in Eq. (22.2.52):

$$J_{\min} = R_S(0) - \int_0^\infty k^2(\tau) d\tau = 1 - \frac{5}{9} \int_0^\infty e^{-\tau} d\tau = \frac{4}{9}$$

We can also calculate  $J_{\min}$  using Eqs. (22.2.39) or (22.2.47):

$$\begin{aligned} J_{\min} &= R_S(0) - \int_0^{\infty} h_{\text{opt}}(\tau)R_S(\tau)d\tau = 1 - \int_0^{\infty} \left( \frac{1}{3}e^{-t} + \frac{1}{3}\delta(t) \right) e^{-(t/2)}dt \\ &= 1 - \left( \frac{1}{3} + \frac{2}{9} \right) = 1 - \frac{5}{9} = \frac{4}{9} \end{aligned}$$

The two equations yield the same result  $J_{\min} = \frac{4}{9} = 0.4444$ .

The causal part of the noncausal solution of Example 22.2.2 for  $\alpha = 0$  is given by  $h_{\text{opt noncausal}}(t) = \frac{1}{5}\delta(t) + \frac{3}{10}e^{-t}u(t)$ , and the causal solution of this example is  $h_{\text{opt causal}}(t) = \frac{1}{3}\delta(t) + \frac{1}{3}e^{-t}u(t)$ , shown in Fig. 22.2.12. The minimum mean-square errors in both examples are given by  $J_{\min \text{noncausal}} = 0.4$  and  $J_{\min \text{causal}} = 0.4444$ . The differences between the causal and noncausal solutions in this example also are not that significant for this problem, but the complexity of getting a causal solution is much greater.

From  $K(\omega)$  in Eq. (A) for a delay  $\alpha$ , we can write

$$\begin{aligned} K(\omega)e^{j\omega\alpha} &= \frac{2\sqrt{5}}{3} \left[ \frac{1}{(1+2j\omega)} + \frac{1}{5(1-j\omega)} \right] e^{j\omega\alpha} \\ &\Leftrightarrow \frac{\sqrt{5}}{3} \left[ e^{-(t+\alpha)/2}u(t+\alpha) + \frac{2}{5}e^{[t+\alpha]}u(-t-\alpha) \right] \end{aligned}$$

**2. Prediction Problem:**  $\alpha > 0$ . For the prediction problem of  $\alpha > 0$ , the causal part of  $k(t + \alpha)$  is given by

$$k(t + \alpha)_{\text{causal}} = \frac{\sqrt{5}}{3} e^{-(t+\alpha)/2}u(t), \quad \alpha > 0$$

and the FT of the preceding equation yields

$$[K(\omega)e^{j\omega\alpha}]_+ = \frac{\sqrt{5}e^{-(\alpha/2)}}{3\left(\frac{1}{2} + j\omega\right)} = \frac{2\sqrt{5}e^{-(\alpha/2)}}{3(1 + 2j\omega)}$$

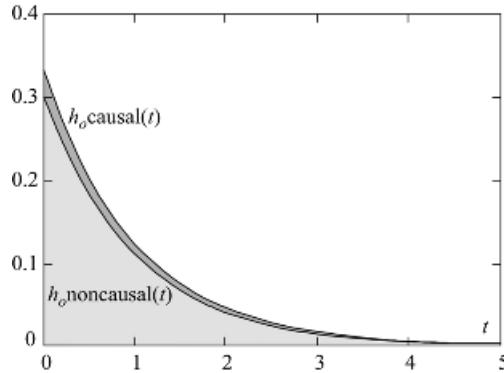


FIGURE 22.2.12

and substituting the preceding equation in Eq. (22.2.49), we obtain the optimum transfer function for the prediction problem

$$\begin{aligned} H_{\text{opt}}(\omega) &= G(\omega)[K(\omega)e^{j\omega\alpha}]_+ = \frac{(1+2j\omega)(2+j\omega)}{2\sqrt{5}(1+j\omega)} \cdot \frac{2\sqrt{5}e^{-(\alpha/2)}}{3(1+2j\omega)} \\ &= \frac{(2+j\omega)}{3(1+j\omega)} \cdot e^{-(\alpha/2)} \end{aligned}$$

and the impulse response, given by

$$h_{\text{opt}}(t) = \frac{1}{3}e^{-[t+(\alpha/2)]}u(t) + \delta(t), \quad \alpha > 0$$

Since  $\alpha > 0$ , the minimum mean-square error is obtained by substituting the causal part of Eq. (C),  $k^2(t) = \frac{5}{9}e^{-t}u(t)$ , in Eq. (22.2.52) to obtain

$$J_{\text{min}} = R_S(0) - \int_{\alpha}^{\infty} k^2(\tau)d\tau = 1 - \int_{\alpha}^{\infty} \frac{5}{9}e^{-\tau}d\tau = 1 - \frac{5}{9}e^{-\alpha}, \quad \alpha > 0$$

**3. Smoothing Problem:  $\alpha < 0$ .** For the smoothing problem of  $\alpha < 0$ , the causal part of  $k(t + \alpha)$  is

$$k(t + \alpha)_{\text{causal}} = \frac{\sqrt{5}}{3} \left[ e^{-[(t+\alpha)/2]}u(t + \alpha) + \frac{2}{5}e^{t+\alpha}[u(-t - \alpha) - u(-t)] \right], \quad \alpha < 0$$

and the FT of the equation above yields

$$\begin{aligned} [K(\omega)e^{j\omega\alpha}]_+ &= \frac{2\sqrt{5}}{3} \left[ \frac{e^{j\omega\alpha}}{1+2j\omega} + \frac{1}{5} \frac{e^{j\omega\alpha} - e^{\alpha}}{(1-j\omega)} \right] \\ &= \frac{2\sqrt{5}}{3} \left[ \frac{3e^{j\omega\alpha}(2-j\omega) - (1+2j\omega)e^{\alpha}}{5(1+2j\omega)(1-j\omega)} \right] \end{aligned}$$

and substituting the preceding equation in Eq. (22.2.49), we obtain the optimum transfer function for the smoothing problem:

$$\begin{aligned} H_{\text{opt}}(\omega) &= \frac{(1+2j\omega)(2+j\omega)}{2\sqrt{5}(1+j\omega)} \frac{2\sqrt{5}}{3} \left[ \frac{3e^{j\omega\alpha}(2-j\omega) - (1+2j\omega)e^{\alpha}}{5(1+2j\omega)(1-j\omega)} \right] \\ &= \frac{1}{3} \left[ \frac{3e^{j\omega\alpha}(4+\omega^2) - (2+j\omega)(1+2j\omega)e^{\alpha}}{5(1+\omega^2)} \right] \end{aligned}$$

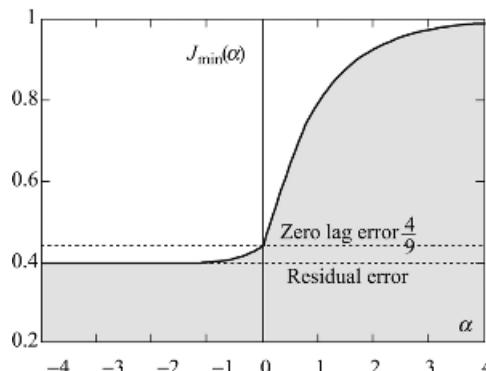


FIGURE 22.2.13

For obtaining the minimum mean-square error, we note that since  $\alpha < 0$ , we have to use the entire Eq. (C) in Eq. (22.2.52) and write

$$\begin{aligned} J_{\min} &= R_S(0) - \int_{\alpha}^{\infty} k^2(\tau) d\tau = 1 - \frac{5}{9} \left[ \frac{4}{25} \int_{\alpha}^0 e^{2\tau} d\tau + \int_0^{\alpha} e^{-\tau} d\tau \right] \\ &= 1 - \frac{5}{9} \left[ \frac{2}{25} (1 - e^{2\alpha}) + 1 \right] = \frac{4}{9} - \frac{2}{45} (1 - e^{2\alpha}), \quad \alpha < 0 \end{aligned}$$

The graph of the minimum mean-square error  $J_{\min}(\alpha)$  for values of  $\alpha$  ranging from  $-\infty$  to  $\infty$  is shown in Fig. 22.2.13.

$J_{\min}$  reaches a minimum value for infinite lag, and this value is given by

$$J_{\min}(-\infty) = \frac{4}{9} - \frac{2}{45} [(1 - e^{2\alpha})] \Big|_{\alpha \rightarrow -\infty} = \frac{4}{9} - \frac{2}{49} = \frac{2}{5}$$

and the maximum error for infinite prediction time is

$$J_{\min}(\infty) = 1 - \frac{5}{9} e^{-\alpha} \Big|_{\alpha \rightarrow \infty} = 1$$

## 22.3 DISCRETE KALMAN FILTER\*

The Wiener filter discussed in the previous sections was derived under the imposition of the physical realizability condition on the filter. As a consequence of this condition, Eq. (22.2.12) is not a convolution, and transform techniques could not be directly used, so we had to resort to spectral factorization. Further, the input processes are assumed to be stationary, and the system is linear time-invariant. To circumvent these and other difficulties, Kalman and Bucy [27] proposed a recursive linear digital filter that accommodates nonstationary processes. We will motivate the concept of recursive estimation.

\*Sections 22.3 and 22.4 are adapted with permission from Chapter 9 of the book, *Nonlinear Filtering and Smoothing*, by Venkatarama Krishnan, originally published by John Wiley and Sons in 1984, and republished by Dover Publications in 2005.

### Recursive Estimation

We will assume that  $\{X_i, i = 1, 2, \dots, k\}$  is a discrete signal random process and that  $\{Y_i, i = 1, 2, \dots, k\}$  is an observation or measurement random process related to the signal process by the equation

$$Y_i = X_i + V_i, \quad i = 1, 2, \dots, k \quad (22.3.1)$$

where  $\{V_i, i = 1, 2, \dots, k\}$  is a zero mean white-noise sequence with  $E[V_i V_j] = \sigma_V^2 \delta_{ij}$ ,  $i, j = 1, 2, \dots, k$ . We will further assume that  $X_i = X$  for all  $i$ . The schematic is shown in Fig. 22.3.1.

We want to estimate the process  $\{X_i = X\}$  given the observation process  $\{Y_i, i = 1, 2, \dots, k\}$ . From Eq. (18.3.10) the minimum variance unbiased estimator is  $\hat{X}$ , the conditional expectation of  $X$  given  $\{Y_i, i = 1, 2, \dots, k\}$ , which can be written as

$$\hat{X}_k = \frac{1}{k} \sum_{i=1}^k Y_i \quad (22.3.2)$$

An additional observation  $Y_{k+1}$  is available at time  $k+1$ . The new estimate is given by

$$\hat{X}_{k+1} = \frac{1}{k+1} \sum_{i=1}^{k+1} Y_i \quad (22.3.3)$$

However, the estimate given by Eq. (22.3.3) processes all observations  $\{Y_i, i = 1, \dots, k+1\}$  again, and this is unacceptable for machine computation. We want to use the prior estimate  $\hat{X}$  from Eq. (22.3.2) to save on computation time. Hence we can rewrite Eq. (22.3.3) as follows:

$$\begin{aligned} \hat{X}_{k+1} &= \frac{k}{k+1} \frac{1}{k} \sum_{i=1}^{k+1} Y_i + \frac{Y_{k+1}}{k+1} = \frac{k}{k+1} \hat{X}_k + \frac{1}{k+1} Y_{k+1} \\ &= \left(1 - \frac{1}{k+1}\right) \hat{X}_k + \frac{1}{k+1} Y_{k+1} = \hat{X}_k + \frac{1}{k+1} (Y_{k+1} - \hat{X}_k) \end{aligned} \quad (22.3.4)$$

In Eq. (22.3.4) the term  $(Y_{k+1} - \hat{X}_k)$  is the *innovations process*, so called because it contains new information. We have utilized the past estimate to update the present estimate, thus achieving recursive action.

More generally, the first equation in Eq. (22.3.4) can be expressed in the form

$$\hat{X}_{k+1} = K'_k \hat{X}_k + K_k Y_{k+1} \quad (22.3.5)$$

where the weighting coefficients  $K'_k$  and  $K_k$  have to be determined using the optimality criteria of unbiasedness and minimum variance. We will reformulate this problem so that the solution in the general case will be clear. The time just *before the observation*

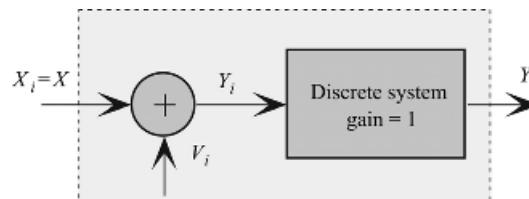


FIGURE 22.3.1

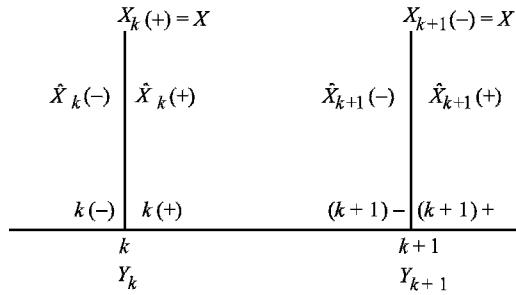


FIGURE 22.3.2

$Y_{k+1}$  is denoted by  $(k+1)-$ , and the time just *after the observation* is denoted by  $(k+1)+$ . Hence  $\hat{X}_{k+1}(-)$  is the estimate just before the observation, and  $\hat{X}_{k+1}(+)$  is the estimate just after the observation. The timing diagram of Fig. 22.3.2 illustrates these quantities.

The evolution of the state between the times  $k+$  and  $(k+1)-$  is governed by the dynamics of the system, and in this case, since  $X_k$  is a constant  $X$  for all times, we have

$$X_k(+) = X \text{ and } X_{k+1}(-) = X \text{ and hence } \hat{X}_k(+) = \hat{X}_{k+1}(-) \quad (22.3.6)$$

Thus there are two distinct zones: (1) *between observations*, where the estimate is governed by the dynamics of the system; and (2) *across observations*, where the estimate is updated by the new observation. The recursive formulation of Eq. (22.3.4) can be written in two stages as

$$\begin{aligned} \text{Between observations: } \hat{X}_{k+1}(-) &= \hat{X}_k(+) \\ \text{Across observations: } \hat{X}_{k+1}(+) &= \hat{X}_{k+1}(-) + \frac{1}{k+1} [Y_{k+1} - \hat{X}_{k+1}(-)] \end{aligned} \quad (22.3.7)$$

The two equations in Eq. (22.3.7) can be combined into a single equation as

$$\hat{X}_{k+1}(+) = \hat{X}_k(+) + \frac{1}{k+1} [Y_{k+1} - \hat{X}_k(+)] \quad (22.3.8)$$

which is Eq. (22.3.4) in a more precise form.

### Formulation of Discrete Kalman Filter

We can now formulate the recursive discrete nonstationary filtering problem. **Boldface** will represent both matrices and vectors. A discrete linear nonstationary signal process is represented by the vector difference equation

$$\text{Signal process: } \mathbf{X}_k = \Phi_{k,k-1} \mathbf{X}_{k-1} + \Gamma_{k-1} \mathbf{W}_{k-1}, \quad k \in \{1, 2, \dots, N\}, \quad \mathbf{X}_0 \quad (22.3.9)$$

The observation process is another discrete linear process represented by

$$\text{Observation process: } \mathbf{Y}_k = \mathbf{H}_k \mathbf{X}_k + \mathbf{V}_k, \quad k \in \{1, 2, \dots, N\} \quad (22.3.10)$$

Given the observations  $\{\mathbf{Y}_k, k \in \{1, 2, \dots, N\}\}$ , we have to find the estimate  $\hat{\mathbf{X}}_k$  for the signal  $\mathbf{X}_k$  under the optimality criteria of unbiasedness and minimum variance, and this estimate is given by the conditional expectation  $\hat{\mathbf{X}}_k = E[\mathbf{X} | \mathbf{Y}_k, k \in \{1, 2, \dots, N\}]$ .

The following assumptions are made regarding Eqs. (22.3.9) and (22.3.10).

### Assumptions

1. The signal process  $\{\mathbf{X}_k, k = [1, 2, \dots, N]\}$  with initial condition  $\mathbf{X}_0$  is an  $n$ -vector, and the observation process  $\{\mathbf{Y}_k, k = [1, 2, \dots, N]\}$  with initial condition  $\mathbf{Y}_0 = \mathbf{0}$  is an  $m$ -vector at time  $t = t_k$ .
2.  $\Phi_{k,k-1}$  is an  $n \times n$  nonsingular discrete signal transition matrix from time  $t = t_{k-1}$  to time  $t = t_k$ .
3.  $\Gamma_k$  is an  $n \times r$  discrete nonrandom rectangular matrix at time  $t = t_k$ .
4.  $\mathbf{W}_k$  is an  $r$ -vector of zero mean white Gaussian signal noise sequence with covariance matrix  $E[\mathbf{W}_k \mathbf{W}_k^T] = \mathbf{Q}_k$ .
5.  $\mathbf{V}_k$  is an  $m$ -vector of zero mean white Gaussian observation noise sequence with covariance matrix  $E[\mathbf{V}_k \mathbf{V}_k^T] = \mathbf{R}_k$  where  $\mathbf{W}_k$  and  $\mathbf{V}_k$  are independent,  $E[\mathbf{W}_k \mathbf{V}_k^T] = \mathbf{0}$ . All quantities are evaluated at time  $t = t_k$ .
6.  $\mathbf{H}_k$  is an  $m \times n$  discrete nonrandom rectangular weighting matrix at time  $t = t_k$ .
7. The Gaussian initial condition vector  $\mathbf{X}_0$  is assumed independent of both the noise vector sequences  $\{\mathbf{W}_k\}$  and  $\{\mathbf{V}_k\}$ .
8. If  $\mathbf{U}_k$  is any random vector sequence, then the estimate  $\hat{\mathbf{U}}_k$  is the expectation of  $\mathbf{U}_k$  conditioned on the observations  $\{\mathbf{Y}_k, k = [1, 2, \dots, N]\}$  and will be represented by  $E^k[\mathbf{U}_k] = \hat{\mathbf{U}}_k = E[\mathbf{U}_k | \mathbf{Y}_k, k = 1, 2, \dots, N]$ .
9. The estimation error vector sequence  $\mathbf{X}_k = (\hat{\mathbf{X}}_k - \mathbf{X}_k)$  and the noise vector sequences  $\{\mathbf{W}_k\}$  and  $\{\mathbf{V}_k\}$  are independent.
10. The innovations sequence  $\{\mathbf{v}_k, k = [1, 2, \dots, N]\}$  is also an  $m$ -vector, given by

$$\mathbf{v}_k = \mathbf{Y}_k - \mathbf{H}_k \hat{\mathbf{X}}_k, \quad k = 1, 2, \dots, N \quad (22.3.11)$$

11. The  $n \times n$  error covariance matrix defined by,  $\mathbf{P}_k = E[\hat{\mathbf{X}}_k - \mathbf{X}_k][\hat{\mathbf{X}}_k - \mathbf{X}_k]^T$  is positive definite.

Under the these assumptions, the discrete Kalman filter takes the form as shown:

*Across Observations.* At the observation time  $t_k$ , the conditional expectations  $E^k[\mathbf{X}_k] \big|_{t_k(+)} = \hat{\mathbf{X}}_k(+)$  and  $E^k[\mathbf{X}_k] \big|_{t_k(-)} = \hat{\mathbf{X}}_k(-)$  satisfy

$$\begin{aligned} \hat{\mathbf{X}}_k(+) &= \hat{\mathbf{X}}_k(-) + \mathbf{K}_k[\mathbf{Y}_k - \mathbf{H}_k \hat{\mathbf{X}}_k(-)], \\ k \in \{1, 2, \dots, N\}, \quad \hat{\mathbf{X}}_0 &= \mathbf{X}_0 \end{aligned} \quad (22.3.12a)$$

and the covariance matrix  $\mathbf{P}_k(+)$  satisfies

$$\mathbf{P}_k(+) = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k(-), \quad k \in \{1, 2, \dots, N\} \quad (22.3.12b)$$

where the  $n \times m$  matrix  $\mathbf{K}_k$ , called the *Kalman gain*, is given by

$$\mathbf{K}_k = \mathbf{P}_k(-) \mathbf{H}_k^T [\mathbf{H}_k \mathbf{P}_k(-) \mathbf{H}_k^T + \mathbf{R}_k]^{-1}, \quad k \in \{1, 2, \dots, N\} \quad (22.3.13)$$

*Between Observations.* Between observation times  $t_{k-1}$  and  $t_k$  the estimates satisfy

$$\hat{\mathbf{X}}_k(-) = \Phi_{k,k-1} \hat{\mathbf{X}}_{k-1}(+), \quad k \in \{1, 2, \dots, N\} \quad (22.3.14a)$$

and the covariance matrix satisfies

$$\mathbf{P}_k(-) = \Phi_{k,k-1} \mathbf{P}_{k-1}(+) \Phi_{k,k-1}^T + \Gamma_{k-1} \mathbf{Q}_{k-1} \Gamma_{k-1}^T, \quad k \in \{1, 2, \dots, N\} \quad (22.3.14b)$$

We can gain insight into the problem by deriving these equations. Let  $\hat{\mathbf{X}}_k(-)$  be the estimate determined at time  $t_k(-)$  and  $Y_k$ , the observation at time  $t_k$ . We will use the prior estimate  $\hat{\mathbf{X}}_k(-)$  and the observation  $Y_k$  to get a new estimate to  $\hat{\mathbf{X}}_k(+)$  at time  $t_k(+)$ . We can write an equation analogous to the scalar Eq. (22.3.5) as

$$\hat{\mathbf{X}}_k(+) = \mathbf{K}'_k \hat{\mathbf{X}}_k(-) + \mathbf{K}_k \mathbf{Y}_k \quad (22.3.15)$$

where the weighting  $n \times n$  matrix  $\mathbf{K}'_k$  and weighting  $n \times m$  matrix  $\mathbf{K}_k$  have to be determined using the two criteria of unbiasedness and minimum variance.

If  $\mathbf{X}_k$  is the true unobservable value of the state at time  $t_k$ , we can define estimation errors  $\tilde{\mathbf{X}}_k(+)$  at time  $t_k(+)$  and  $\tilde{\mathbf{X}}_k(-)$  at time  $t_k(-)$  as

$$\begin{aligned} \tilde{\mathbf{X}}_k(+) &= \hat{\mathbf{X}}_k(+) - \mathbf{X}_k \\ \tilde{\mathbf{X}}_k(-) &= \hat{\mathbf{X}}_k(-) - \mathbf{X}_k \end{aligned} \quad (22.3.16)$$

We can convert Eq. (22.3.15) into an error equation by subtracting  $\mathbf{X}_k$  from both sides:

$$\hat{\mathbf{X}}_k(+) - \mathbf{X}_k = \mathbf{K}'_k [\hat{\mathbf{X}}_k(-) - \mathbf{X}_k] + \mathbf{K}'_k \mathbf{X}_k - \mathbf{X}_k + \mathbf{K}_k \mathbf{Y}_k \quad (22.3.17)$$

Substituting for  $\mathbf{Y}_k$  from Eq. (22.3.10) and error matrices from Eq. (22.3.16) in Eq. (22.3.17), we obtain

$$\tilde{\mathbf{X}}_k(+) = \mathbf{K}'_k \tilde{\mathbf{X}}_k(-) + [\mathbf{K}'_k + \mathbf{K}_k \mathbf{H}_k - \mathbf{I}] \mathbf{X}_k + \mathbf{K}_k \mathbf{V}_k \quad (22.3.18)$$

Taking expected value of both sides of Eq. (22.3.18), we obtain

$$E[\tilde{\mathbf{X}}_k(+)] = \mathbf{K}'_k E[\tilde{\mathbf{X}}_k(-)] + [\mathbf{K}'_k + \mathbf{K}_k \mathbf{H}_k - \mathbf{I}] E[\mathbf{X}_k] + \mathbf{K}_k E[\mathbf{V}_k] \quad (22.3.19)$$

If the estimator  $\hat{\mathbf{X}}_k$  is to be unbiased, then we must have

$$E[\tilde{\mathbf{X}}_k(+)] = \mathbf{0} = E[\tilde{\mathbf{X}}_k(-)] \quad (22.3.20)$$

Hence, with  $E[\mathbf{V}_k] = \mathbf{0}$  and  $E[\mathbf{X}_k] \neq \mathbf{0}$  in Eq. (22.3.19), we deduce

$$\mathbf{K}'_k + \mathbf{K}_k \mathbf{H}_k - \mathbf{I} = \mathbf{0} \quad \text{or} \quad \mathbf{K}'_k = \mathbf{I} - \mathbf{K}_k \mathbf{H}_k \quad (22.3.21)$$

We have thus expressed the matrix  $\mathbf{K}'_k$  in terms of the matrix  $\mathbf{K}_k$ .

*Across Observations.* Substituting Eq. (22.3.21) in Eq. (22.3.15), we establish the signal update equation:

$$\hat{\mathbf{X}}_k(+) = \hat{\mathbf{X}}_k(-) + \mathbf{K}_k [\mathbf{Y}_k - \mathbf{H}_k \hat{\mathbf{X}}_k(-)], \quad k \in \{1, 2, \dots, N\}, \quad \hat{\mathbf{X}}_0 = \hat{\mathbf{X}}_0 \quad (22.3.22)$$

Thus yet unknown matrix  $\mathbf{K}_k$  is determined using the minimum variance criterion. The estimation error equation [Eq. (22.3.18)] can be rewritten by substituting  $\mathbf{K}'_k$  as

$$\begin{aligned} \tilde{\mathbf{X}}_k(+) &= [\mathbf{I} - \mathbf{K}_k \mathbf{H}_k] \tilde{\mathbf{X}}_k(-) + \mathbf{K}_k \mathbf{V}_k \\ &= \tilde{\mathbf{X}}_k(-) + \mathbf{K}_k [\mathbf{V}_k - \mathbf{H}_k \tilde{\mathbf{X}}_k(-)] \end{aligned} \quad (22.3.23)$$

Equation (22.3.12a) represents the estimate update, whereas Eq. (22.3.23) represents the estimate error update. We can now define the error covariance matrices  $\mathbf{P}_k(+)$  and  $\mathbf{P}_k(-)$  as

$$\mathbf{P}_k(+) = E[\tilde{\mathbf{X}}_k(+) \tilde{\mathbf{X}}_k^T(+)]; \quad \mathbf{P}_k(-) = E[\tilde{\mathbf{X}}_k(-) \tilde{\mathbf{X}}_k^T(-)] \quad (22.3.24)$$

Substituting Eq. (22.3.23) in Eq. (22.3.24), we obtain

$$\begin{aligned} \mathbf{P}_k(+) &= E\{\tilde{\mathbf{X}}_k(-) + \mathbf{K}_k[\mathbf{V}_k - \mathbf{H}_k \tilde{\mathbf{X}}_k(-)] \\ &\quad \times (\tilde{\mathbf{X}}_k(-) + \mathbf{K}_k[\mathbf{V}_k - \mathbf{H}_k \tilde{\mathbf{X}}_k(-)])^T\} \\ &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k(-) (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T \end{aligned} \quad (22.3.25)$$

The variance function that should be minimized is

$$E[\tilde{\mathbf{X}}_k^T(+) \tilde{\mathbf{X}}_k(+)] = \text{Tr } \mathbf{P}_k(+) \quad (22.3.26)$$

where  $\text{Tr } \mathbf{P}_k(+)$  is the trace of the covariance matrix  $\mathbf{P}_k(+)$ . We can minimize Eq. (22.3.26) with respect to the matrix  $\mathbf{K}_k$  by solving

$$\begin{aligned} \frac{\partial}{\partial \mathbf{K}_k} \text{Tr } \mathbf{P}_k(+) &= \frac{\partial}{\partial \mathbf{K}_k} \text{Tr}[(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k(-) (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T] \\ &= \mathbf{0} \end{aligned} \quad (22.3.27)$$

Equation (22.3.27) is solved by using the result from Eq. (16.3.23) stating that if a matrix  $\mathbf{B}$  is symmetric and  $\mathbf{A}$  is any other matrix, then

$$\frac{\partial}{\partial \mathbf{A}} \text{Tr}[\mathbf{ABA}^T] = 2\mathbf{AB} \quad (16.3.23)$$

Applying Eq. (16.3.23) to Eq. (22.3.27) results in

$$\frac{\partial}{\partial \mathbf{K}_k} \text{Tr } \mathbf{P}_k(+) = -2(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k(-) \mathbf{H}_k^T + \mathbf{K}_k \mathbf{R}_k = \mathbf{0} \quad (22.3.29)$$

Solving for  $\mathbf{K}_k$  in Eq. (22.3.29), we obtain the equation for the Kalman gain matrix

$$\mathbf{K}_k = \mathbf{P}_k(-) \mathbf{H}_k^T [\mathbf{H}_k \mathbf{P}_k(-) \mathbf{H}_k^T + \mathbf{R}_k]^{-1}, \quad k \in \{1, 2, \dots, N\} \quad (22.3.30)$$

which is Eq. (22.3.13). Substituting  $\mathbf{K}_k$  from Eq. (22.3.30), in Eq. (22.3.25), we obtain Eq. (22.3.12b):

$$\mathbf{P}_k(+) = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k(-), \quad k \in \{1, 2, \dots, N\} \quad (22.3.31)$$

We have shown the result for “across observations”.

*Between Observations.* The conditional expectation of the signal process shown below

$$\mathbf{X}_k = \Phi_{k, k-1} \mathbf{X}_{k-1} + \Gamma_{k-1} \mathbf{W}_{k-1}, \quad k \in \{1, 2, \dots, N\}, \quad \mathbf{X}_0 \quad (22.3.32)$$

conditioned on the observations  $\{\mathbf{Y}_{k-1}\}$ , is given by

$$\hat{\mathbf{X}}_k(-) = \Phi_{k, k-1} \hat{\mathbf{X}}_{k-1}(+) + \Gamma_{k-1} \hat{\mathbf{W}}_{k-1}, \quad k \in \{1, 2, \dots, N\} \quad (22.3.33)$$

where  $\hat{\mathbf{X}}_k(-) = E[\mathbf{X}_k | \{\mathbf{Y}_{k-1}\}]$ . Since  $\mathbf{W}_{k-1}$  is independent of the observations  $\{\mathbf{Y}_{k-1}\}$ ,  $\hat{\mathbf{W}}_{k-1} = E[\mathbf{W}_{k-1} | \{\mathbf{Y}_{k-1}\}] = E[\mathbf{W}_{k-1}] = \mathbf{0}$ , and Eq. (22.3.33) reduces to Eq. (22.3.14a):

$$\hat{\mathbf{X}}_k(-) = \Phi_{k, k-1} \hat{\mathbf{X}}_{k-1}(+), \quad k \in \{1, 2, \dots, N\} \quad (22.3.34)$$

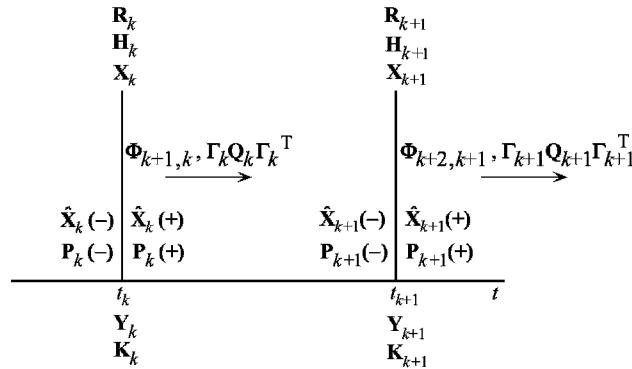


FIGURE 22.3.3

To demonstrate Eq. (22.3.14b) we subtract Eq. (22.3.32) from Eq. (22.3.34) to obtain the error equation

$$\tilde{\mathbf{X}}_k(-) = \Phi_{k,k-1} \tilde{\mathbf{X}}_{k-1}(+) - \Gamma_{k-1} \mathbf{W}_{k-1} \quad (22.3.35)$$

and form the error covariance matrix  $\mathbf{P}_k(-) = E[\mathbf{X}_k(-)\mathbf{X}_k(-)^T]$ , given by

$$\begin{aligned} \mathbf{P}_k(-) &= E\{[\Phi_{k,k-1} \tilde{\mathbf{X}}_{k-1}(+) - \Gamma_{k-1} \mathbf{W}_{k-1}][\Phi_{k,k-1} \tilde{\mathbf{X}}_{k-1}(+) - \Gamma_{k-1} \mathbf{W}_{k-1}]^T\} \\ &= E\{\Phi_{k,k-1} \tilde{\mathbf{X}}_{k-1}(+) \tilde{\mathbf{X}}_{k-1}^T (+) \Phi_{k,k-1}^T + \Gamma_{k-1} \mathbf{W}_{k-1} \mathbf{W}_{k-1}^T \Gamma_{k-1}^T \\ &\quad - \Phi_{k,k-1} \tilde{\mathbf{X}}_{k-1}(+) \mathbf{W}_{k-1}^T \Gamma_{k-1}^T - \Gamma_{k-1} \mathbf{W}_{k-1} \tilde{\mathbf{X}}_{k-1}^T (+) \Phi_{k,k-1}^T\} \end{aligned} \quad (22.3.36)$$

From assumption 9 (above) that the estimation errors are independent of the noise sequences, Eq. (22.3.36) can be simplified to Eq. (22.3.14b):

$$\mathbf{P}_k(-) = \Phi_{k,k-1} \mathbf{P}_{k-1}(+) \Phi_{k,k-1}^T + \Gamma_{k-1} \mathbf{Q}_{k-1} \Gamma_{k-1}^T, \quad k \in \{1, 2, \dots, N\} \quad (22.3.37)$$

It is to be particularly noted that the variance update equation [Eq. (22.3.31)] does not depend on either the estimate or the observations, whereas the signal update [Eq. (22.3.22)] is dependent on the covariance matrix via the Kalman gain.

The timing diagram similar to Fig. 22.3.2 for the realization of the Kalman filter is shown in Fig. 22.3.3.

The block diagram for the implementation of the discrete Kalman filter is shown in Fig. 22.3.4.

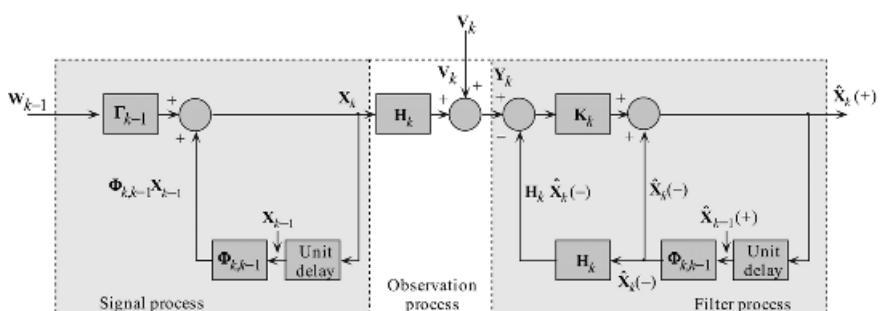


FIGURE 22.3.4

### Comments

1. The equations for across and between measurements can be combined into a single equation for both the estimated mean and the unconditional variance, and written as follows:

$$\hat{\mathbf{X}}_k(-) = \Phi_{k,k-1} \{ \hat{\mathbf{X}}_{k-1}(-) + \mathbf{K}_{k-1} [\mathbf{Y}_{k-1} - \mathbf{H}_{k-1} \hat{\mathbf{X}}_{k-1}(-)] \}, \quad k \in \{1, 2, \dots, N\} \quad (22.3.38)$$

$$\begin{aligned} \mathbf{P}_k(-) = & \Phi_{k,k-1} [(\mathbf{I} - \mathbf{K}_{k-1} \mathbf{H}_{k-1}) \mathbf{P}_{k-1}(-) \Phi_{k,k-1}^T \\ & + \Gamma_{k-1} \mathbf{Q}_{k-1} \Gamma_{k-1}^T], \quad k \in \{1, 2, \dots, N\} \end{aligned} \quad (22.3.39)$$

where, from Eq. (22.3.30), we obtain

$$\mathbf{K}_{k-1} = \mathbf{P}_{k-1}(-) \mathbf{H}_{k-1}^T [\mathbf{H}_{k-1} \mathbf{P}_{k-1}(-) \mathbf{H}_{k-1}^T + \mathbf{R}_{k-1}]^{-1}, \quad k \in \{1, 2, \dots, N\} \quad (22.3.40)$$

2. The error covariance update equation [Eq. (22.3.31)]

$$\mathbf{P}_k(+) = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k(-), \quad k \in \{1, 2, \dots, N\} \quad (22.3.31)$$

can be expressed in an alternate form as

$$\mathbf{P}_k^{-1}(+) = \mathbf{P}_k^{-1}(-) + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \quad (22.3.41)$$

3. The Kalman gain matrix  $\mathbf{K}_k$  given by Eq. (22.3.30)

$$\mathbf{K}_k = \mathbf{P}_k(-) \mathbf{H}_k^T [\mathbf{H}_k \mathbf{P}_k(-) \mathbf{H}_k^T + \mathbf{R}_k]^{-1}, \quad k \in \{1, 2, \dots, N\} \quad (22.3.30)$$

can also be expressed in an alternate form as

$$\mathbf{K}_k = \mathbf{P}_k(+) \mathbf{H}_k^T \mathbf{R}_k^{-1} \quad (22.3.42)$$

4. If the signal and observation noises are not assumed Gaussian, then the variance equation will be a conditional variance that will be dependent on the third conditional moment. The Gaussian assumption renders the third moments zero, resulting in an unconditional variance. Hence the signal estimate equation [Eq. (22.3.38)] and the variance equation [Eq. (22.3.39)] are not coupled as stated earlier. As a consequence, the Kalman gain and covariance matrix can be

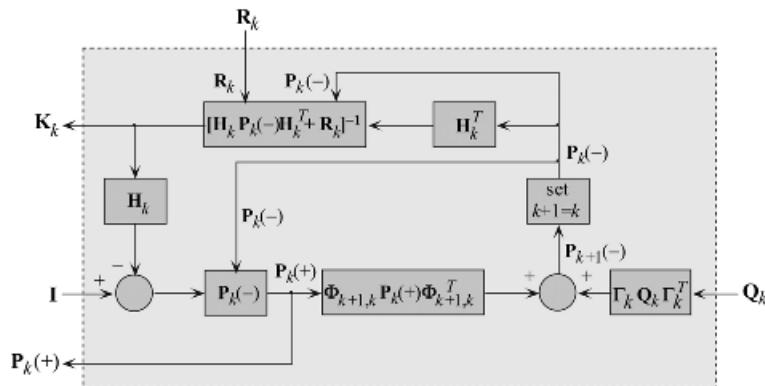


FIGURE 22.3.5

processed independently of the observations or the state estimate. This is of immense value in navigation problems. The propagation of the covariance matrix  $\mathbf{P}_k(+)$  and the evaluation of the Kalman gain  $\mathbf{K}_k$  are shown in Fig. 22.3.5.

The Kalman filter algorithm is summarized in the following list:

### Summary of Discrete Kalman Filter Algorithm

In this list  $N$  stands for the discrete data set  $N = \{1, 2, \dots, N\}$ .

*Signal or State Process:*

$$\mathbf{X}_k = \Phi_{k, k-1} \mathbf{X}_{k-1} + \Gamma_{k-1} \mathbf{W}_{k-1}, \quad k \in N, \mathbf{X}_0 \quad (22.3.9)$$

*Observation or Measurement Process:*

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X}_k + \mathbf{V}_k, \quad k \in N, \mathbf{Y}_0 = 0 \quad (22.3.10)$$

*Signal Error Process:*

$$\mathbf{X}_k = \hat{\mathbf{X}}_k - \mathbf{X}_k$$

*Prior Statistics:*

$$\begin{aligned} E[\mathbf{W}_k] &= E[\mathbf{V}_k] = \mathbf{0}; E[\mathbf{W}_k \mathbf{W}_k^T] = \mathbf{Q}_k, E[\mathbf{V}_k \mathbf{V}_k^T] = \mathbf{R}_k \\ E[\mathbf{W}_k \mathbf{V}_k^T] &= E[\mathbf{W}_k \mathbf{X}_0^T] = E[\mathbf{V}_k \mathbf{X}_0^T] = \mathbf{0} \\ \text{cov}[\mathbf{X}_0] &= E[\mathbf{X}_0 \mathbf{X}_0^T] = \mathbf{P}_0; \text{cov}[\mathbf{X}_k] = E[\mathbf{X}_k \mathbf{X}_k^T] = \mathbf{P}_k \end{aligned} \quad (22.3.43)$$

### Filter Algorithm

*Across Observations*

*Variance Update Equation:*

$$\begin{aligned} \mathbf{P}_k(+) &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k(-1) \quad \text{or} \quad \mathbf{P}_k^{-1}(+) \\ &= \mathbf{P}_k^{-1}(-) + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k, \quad k \in N \end{aligned} \quad (22.3.12b)$$

*Kalman Gain Update Equation:*

$$\mathbf{K}_k = \mathbf{P}_k(-) \mathbf{H}_k^T [\mathbf{H}_k \mathbf{P}_k(-) \mathbf{H}_k^T + \mathbf{R}_k]^{-1} = \mathbf{P}_k(+) \mathbf{H}_k^T \mathbf{R}_k^{-1}, \quad k \in N \quad (22.3.13)$$

*Signal Estimate Update Equation:*

$$\hat{\mathbf{X}}_k(+) = \hat{\mathbf{X}}_k(-) + \mathbf{K}_k [\mathbf{Y}_k - \mathbf{H}_k \hat{\mathbf{X}}_k(-)], \quad k \in N, \quad \hat{\mathbf{X}}_0 = \mathbf{X}_0 \quad (22.3.12a)$$

*Between Observations*

*Variance Extrapolation:*

$$\mathbf{P}_k(-) = \Phi_{k, k-1} \mathbf{P}_{k-1}(+) \Phi_{k, k-1}^T + \Gamma_{k-1} \mathbf{Q}_{k-1} \Gamma_{k-1}^T, \quad k \in N \quad (22.3.14b)$$

*Signal Estimate Extrapolation:*

$$\hat{\mathbf{X}}_k(-) = \Phi_{k, k-1} \hat{\mathbf{X}}_{k-1}(+), \quad k \in N \quad (22.3.14a)$$

**Example 22.3.1** The observation process is given by a scalar equation:

$$Y_{k-1} = X_{k-1} + V_{k-1}, \quad k \in \{1, 2, \dots, N\}$$

where the variance of the noise sequence  $\{V_k\}$  is  $\mathbf{R}_k = \text{var}[V_k] = \sigma_V^2$  and the scalar covariance matrices  $\mathbf{P}_k(+) = p_k(+)$  and  $\mathbf{P}_k(-) = p_k(-)$ . We will investigate the propagation of the error variance when the next observation  $Y_k$  is available. Using the more convenient covariance equation [Eq. (22.3.40)], we obtain

$$\frac{1}{p_k(+)} = \frac{1}{p_k(-)} + \frac{1}{\sigma_V^2} \quad \text{or} \quad p_k(+) = \frac{p_k(-)}{1 + p_k(-)/\sigma_V^2}$$

We observe that if the observation uncertainty  $\sigma_V^2$  is low, then the additional observation  $Y_k$  reduces the estimation uncertainty  $p_k(+)$  drastically, or  $p_k(+) \ll p_k(-)$ . On the other hand, if  $\sigma_V^2$  is high, then the estimation uncertainty remains practically the same:  $p_k(+) \approx p_k(-)$ . An important fact to note is that an additional observation does not degrade the estimation accuracy! The Kalman gain  $k_k$  is computed from Eq. (22.3.41) as follows:

$$k_k = p_k(+) \cdot 1 \cdot \frac{1}{\sigma_V^2} = \frac{p_k(-)}{1 + p_k(-)/\sigma_V^2} \cdot \frac{1}{\sigma_V^2} = \frac{p_k(-)}{\sigma_V^2 + p_k(-)} = \frac{1}{1 + \sigma_V^2/p_k(-)}$$

For low measurement variance, the Kalman gain is approximately 1; for high measurement variance, the gain becomes 0. As stated earlier, the variance and the gain can be processed independently of the observations.

**Example 22.3.2** A zero mean signal and observation processes are given by

$$X_k = X_{k-1}, \quad k \in \{1, 2, \dots, N\}; \quad Y_k = X_k + V_k, \quad k \in \{1, 2, \dots, N\}$$

where  $\{V_k\}$  is a zero mean Gaussian white noise sequence with variance  $\sigma_V^2$ . The initial error variance is given by  $E[\hat{X}_0 - X_0]^2 = p_0$ .

*Across Observations.* The Kalman gain  $k_k$  is obtained from Eq. (22.3.13) as

$$k_k = p_k(-) \cdot 1 \cdot [1 \cdot p_k(-) \cdot 1 + \sigma_V^2]^{-1} = \frac{p_k(-)}{p_k(-) + \sigma_V^2}$$

and from Eq. (22.3.41) it can also be given by

$$k_k = p_k(+) \cdot 1 \cdot \frac{1}{\sigma_V^2} = \frac{p_k(+)}{\sigma_V^2}$$

The variance is obtained from Eq. (22.3.12b) as

$$p_k(+) = \left[ 1 - \frac{p_k(-)}{p_k(-) + \sigma_V^2} \right] \quad p_k(-) = \frac{p_k(-)\sigma_V^2}{p_k(-) + \sigma_V^2} = \frac{p_k(-)}{1 + p_k(-)/\sigma_V^2} \quad (\text{A})$$

which is the same as obtained in the previous example. The signal estimate is obtained from Eq. (22.3.12a) as

$$\hat{X}_k(+) = \hat{X}_k(-) + \frac{p_k(+)}{\sigma_V^2} [Y_k - \hat{X}_k(-)] = \hat{X}_k(-) + \frac{p_k(-)}{p_k(-) + \sigma_V^2} [Y_k - \hat{X}_k(-)]$$

*Between Observations.* From Eqs. (22.3.14), we obtain

$$\hat{X}_{k+1}(-) = \hat{X}_k(+); \quad p_{k+1}(-) = p_k(+) \quad (\text{B})$$

*Closed-Form Solution.* Substituting  $p_k(-) = p_{k-1}(+)$  in Eq. (A), we obtain

$$p_k(+) = \frac{p_{k+1}(+)}{1 + p_{k-1}(+)/\sigma_V^2}$$

The preceding difference equation can be solved as follows:

$$p_1(+) = \frac{p_0}{1 + p_0/\sigma_V^2}$$

$$p_2(+) = \frac{p_1(+) \frac{p_0}{1 + p_0/\sigma_V^2}}{1 + p_1(+)/\sigma_V^2} = \frac{\frac{p_0}{1 + p_0/\sigma_V^2}}{1 + \frac{p_0}{\sigma_V^2(1 + p_0/\sigma_V^2)}} = \frac{p_0}{1 + 2p_0/\sigma_V^2}$$

and by induction the solution is

$$p_k(+) = \frac{p_0}{1 + kp_0/\sigma_V^2} \quad \text{and from Eq. (B)} \quad p_{k+1}(-) = \frac{p_0}{1 + kp_0/\sigma_V^2}$$

Hence the Kalman gain will be given by

$$k_k = \frac{p_k(+)}{\sigma_V^2} = \frac{p_0}{\sigma_V^2 + kp_0}$$

The signal estimate equation can be rewritten as

$$\hat{X}_k(+) = \hat{X}_k(-) + \frac{p_0}{\sigma_V^2 + kp_0} [Y_k - \hat{X}_k(-)]$$

and from Eq. (B)

$$\hat{X}_{k+1}(-) = \hat{X}_k(-) + \frac{p_0}{\sigma_V^2 + kp_0} [Y_k - \hat{X}_k(-)]$$

*Remarks.* As the number  $k$  of observations increases,  $p_{k+1}(-)$  decreases and the estimation accuracy improves.

### Kalman Filter—Scalar Case

For a clearer understanding of the Kalman filter, we shall take a scalar case and derive the filter equations. Let  $N$  represent the observation interval given by  $N = \{1, 2, \dots, N\}$ .

The signal and the observation processes for a scalar case are given by

$$\begin{aligned} X_k &= \phi X_{k-1} + W_k, & k \in N \\ &= h_k X_k + V_k, & k \in N \end{aligned} \quad (22.3.44)$$

where  $\{W_k\}$  and  $\{V_k\}$  are zero mean Gaussian white-noise sequences with variances  $\sigma_W^2$  and  $\sigma_V^2$  respectively. The initial error variance is given by  $E[\hat{X}_0 - X_0]^2 = p_0$ .

*Across Observations.* The Kalman gain  $k_k$  is obtained from Eq. (22.3.13) as,

$$k_k = p_k(-) h_k [h_k^2 p_k(-) + \sigma_V^2]^{-1} = \frac{p_k(-) h_k}{h_k^2 p_k(-) + \sigma_V^2} = \frac{p_k(+) h_k}{\sigma_V^2} \quad (22.3.45)$$

The variance is obtained from Eq. (22.3.12b) as follows:

$$\begin{aligned} p_k(+) &= \left[ 1 - \frac{p_k(-)h_k \cdot h_k}{h_k^2 p_k(-) + \sigma_V^2} \right] p_k(-) = \frac{p_k(-)\sigma_V^2}{h_k^2 p_k(-) + \sigma_V^2} \\ &= \frac{p_k(-)}{1 + h_k^2 p_k(-)/\sigma_V^2} \end{aligned} \quad (22.3.46)$$

The signal estimate is obtained from Eq. (22.3.12a) as follows:

$$\begin{aligned} \hat{X}_k(+) &= \hat{X}_k(-) + k_k [Y_k - h_k \hat{X}_k(-)] \\ &= \hat{X}_k(-) + \frac{p_k(-)h_k}{h_k^2 p_k(-) + \sigma_V^2} [Y_k - h_k \hat{X}_k(-)] \end{aligned} \quad (22.3.47)$$

*Between Observations.* From Eqs. (22.3.14a) and (22.3.14b), we obtain

$$\hat{X}_k(-) = \phi \hat{X}_{k+1}(+); \quad p_k(-) = \phi^2 p_{k-1}(+) + \sigma_W^2 \quad (22.3.48)$$

Substituting for  $p_k(-)$  from Eq. (22.3.48) in Eq. (22.3.46), we obtain

$$p_k(+) = \frac{\phi^2 p_{k-1}(+) + \sigma_W^2}{1 + h_k^2 [\phi^2 p_{k-1}(+) + \sigma_W^2]/\sigma_V^2} \quad (22.3.49)$$

Since the first observation comes at  $t = 1$ , we have  $p_0(-) = p_0(+) = p_0$  and  $p_1(+)$  is

$$p_1(+) = \frac{\phi^2 p_0 + \sigma_W^2}{1 + h_1^2 [\phi^2 p_0 + \sigma_W^2]/\sigma_V^2} \quad (22.3.50)$$

Hence  $p_k(+)$  can be solved recursively. The Kalman gain  $k_k$  is given by Eq. (22.3.45). The signal estimate extrapolation is obtained from Eq. (22.3.14a) as

$$\hat{X}_k(-) = \phi \hat{X}_{k-1}(+) \quad (22.3.51)$$

**Example 22.3.3** The signal and observation processes are given by Eq. (22.3.44) with  $\phi = \frac{4}{5}$  and  $h_k = \frac{6}{5}$ . The zero mean white-noise signals have variances  $\sigma_W^2 = 1$  and  $\sigma_V^2 = 1$ . The initial variance  $p_0 = \frac{4}{5}$ . The initial value  $X_0 = 2.5403$ , and the observations  $\{Y_k\}$  for  $N = 20$  are given in Table 22.3.1.

TABLE 22.3.1

$k$	$Y_k$	$k$	$Y_k$
0	5.5629	11	0.7010
1	0.5982	12	4.2405
2	2.9623	13	0.9243
3	3.4911	14	1.3428
4	2.3531	15	3.5725
5	-0.3591	16	1.3205
6	-0.3427	17	0.9066
7	2.4036	18	2.9516
8	1.3624	19	2.7597
9	2.8077	20	1.8991
10	3.2718		

We will formulate the Kalman filter for two cycles and then use the recursive formulas. Since the first observation arrives at  $t = 1$ , we have  $p_0(-) = p_0(+) = p_0 = \frac{4}{5}$ , and substituting this value of  $p_0$  in Eq. (22.3.50), we obtain

*Cycle 1:*

$$p_1(+) = \frac{\left(\frac{4}{5}\right)^2 \cdot \frac{4}{5} + 1}{1 + \left(\frac{6}{5}\right)^2 \cdot \frac{\left(\frac{4}{5}\right)^2 \cdot \frac{4}{5} + 1}{1}} = \frac{4725}{9929} = 0.4759$$

The Kalman gain  $k_1$  is obtained from Eq. (22.3.45) as

$$k_1 = \frac{p_1(+)h_k}{1} = \frac{4725}{9929} \cdot \frac{6}{5} = \frac{5670}{9929} = 0.5711$$

Since  $\hat{X}_0(+) = X_0 = 2.5403$ , we can determine  $\hat{X}_1(-)$  from Eq. (22.3.48) as  $\hat{X}_1(-) = 0.8 \times 2.5403 = 2.0322$ . Hence  $\hat{X}_1(+)$  is obtained from Eq. (22.3.47) as follows:

$$\begin{aligned} \hat{X}_1(+) &= \hat{X}_1(-) + k_1[Y_1 - h_1\hat{X}_1(-)] = 2.0322 + 0.5711 \\ &\quad \times (0.5982 - 1.2 \times 2.0322) = 0.9811 \end{aligned}$$

*Cycle 2.* From Eq. (22.3.49), we have

$$p_2(+) = \frac{\left(\frac{4}{5}\right)^2 \cdot \frac{4725}{9929} + 1}{1 + \left(\frac{6}{5}\right)^2 \cdot \frac{\left(\frac{4}{5}\right)^2 \cdot \frac{4725}{9929} + 1}{1}} = \frac{323,825}{714,533} = 0.4532$$

The Kalman gain  $k_2$  is obtained from Eq. (22.3.45) as

$$k_2 = \frac{p_2(+)h_k}{1} = \frac{323,825}{714,533} \cdot \frac{6}{5} = \frac{388,590}{714,533} = 0.5438$$

Since  $\hat{X}_1(+) = 0.9811$ , we can determine  $\hat{X}_2(-)$  from Eq. (22.3.48) as

$$\hat{X}_2(-) = 0.8 \times 0.9811 = 0.7849$$

Hence  $\hat{X}_2(+)$  is obtained from Eq. (22.3.47) as follows:

$$\begin{aligned} \hat{X}_2(+) &= \hat{X}_2(-) + k_2[Y_2 - h_2\hat{X}_2(-)] \\ &= 0.7849 + 0.5438 \times (2.9623 - 1.2 \times 0.7849) = 1.8836 \end{aligned}$$

We can now use the recursive formulas [Eqs. (22.3.49), (22.3.45), and (22.3.47)] to obtain  $p_k(+)$ ,  $k_k$  and  $\hat{X}_k(+)$  respectively for observation interval  $N = 20$ . Figure 22.3.6 shows the simulated signal process  $X_k$ , the observation data  $Y_k$ , and the estimated data  $\hat{X}_k(+)$  for  $k = 0, \dots, 20$ .

Figure 22.3.7 shows the Kalman gain  $k_k$  for  $k = 0, \dots, 10$ .

The figure shows that from the fifth observation the gain is a constant at 0.5415 and further observations do not improve the estimation accuracy.

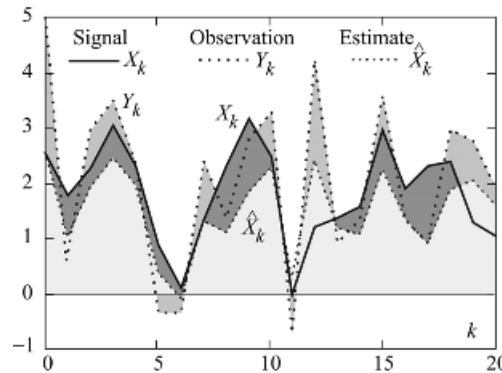


FIGURE 22.3.6

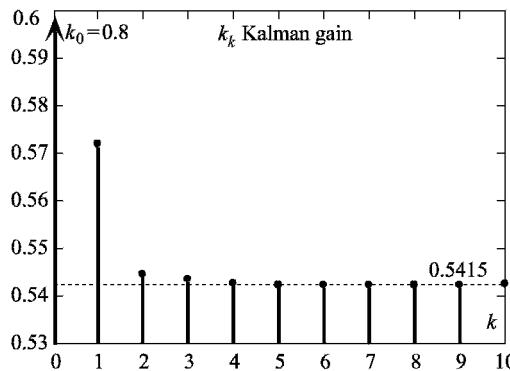


FIGURE 22.3.7

## 22.4 CONTINUOUS KALMAN FILTER

We can now derive the analogous Kalman filter for the continuous-time problem. The continuous linear signal process can be represented in the time interval  $T$  by an  $n$ -vector stochastic differential equation

$$\frac{d\mathbf{X}(t)}{dt} = \mathbf{F}(t)\mathbf{X}(t) + \mathbf{G}(t)\mathbf{W}(t), \quad t \in T, \quad \mathbf{X}_0 \quad (22.4.1)$$

The corresponding linear observation process is given by

$$\mathbf{Y}(t) = \mathbf{H}(t)\mathbf{X}(t) + \mathbf{V}(t), \quad t \in T \quad (22.4.2)$$

Given the observations  $\{\mathbf{Y}(t), t \in T\}$ , we have to find the estimate  $\hat{\mathbf{X}}(t)$  for the signal  $\mathbf{X}(t)$  under the optimality criteria of unbiasedness and minimum variance, and this estimate is given by the conditional expectation  $\hat{\mathbf{X}}(t) = E[\mathbf{X}(t) | \mathbf{Y}(t), t \in T]$ .

### Assumptions

1. The  $n$ -vector process  $\{\mathbf{X}(t), t \in T\}$  and the  $m$ -vector process  $\{\mathbf{Y}(t), t \in T\}$  are random processes defined on a suitable probability space.

2.  $\mathbf{F}(t)$  is a nonsingular  $n \times n$  matrix,  $\mathbf{G}(t)$  is an  $n \times r$  rectangular matrix, and  $\mathbf{H}(t)$  is an  $m \times n$  rectangular matrix such that

$$\begin{aligned} \int_{t \in T} |F_{ij}(t)| dt &< \infty, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, n \\ \int_{t \in T} |G_{ij}(t)|^2 dt &< \infty, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, r \\ \int_{t \in T} |H_{ij}(t)| dt &< \infty, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n \end{aligned} \quad (22.4.3)$$

3.  $\mathbf{W}(t)$  is an  $r$ -vector of zero mean Gaussian white-noise processes with covariance parameter matrix  $\mathbf{Q}(t)$ , given by  $\mathbf{Q}(t) = E[\mathbf{W}(t)\mathbf{W}^T(t)]dt$ .
4.  $\mathbf{V}(t)$  is an  $m$ -vector of zero mean Gaussian white-noise processes with covariance parameter matrix  $\mathbf{R}(t)$ , given by  $\mathbf{R}(t) = E[\mathbf{V}(t)\mathbf{V}^T(t)]dt$ .
5. The noise vectors  $\mathbf{W}(t)$  and  $\mathbf{V}(t)$  are independent, or  $E[\mathbf{W}(t)\mathbf{V}^T(t)] = \mathbf{0}$ .
6. The Gaussian initial condition vector  $\mathbf{X}_0$  is independent of both the noise vectors  $\mathbf{W}(t)$  and  $\mathbf{V}(t)$ , or  $E[\mathbf{W}(t)\mathbf{X}_0^T] = \mathbf{0}$  and  $E[\mathbf{V}(t)\mathbf{X}_0^T] = \mathbf{0}$ .
7. If  $\mathbf{U}(t)$  is any random process vector, then the estimate  $\hat{\mathbf{U}}(t)$  is the expectation of  $\mathbf{U}(t)$  conditioned on the observation  $\{\mathbf{Y}(s), s \leq t\}$  and will be represented by

$$E^t[\mathbf{U}(t)] = \hat{\mathbf{U}}(t) = E[\mathbf{U}(t) | \mathbf{Y}(s), \quad s \leq t].$$

8. The innovations process  $\{\mathbf{v}(t), t \in T\}$  is an  $m$ -vector white-noise process given by

$$\mathbf{v}(t) = \mathbf{Y}(t) - \mathbf{H}(t)\hat{\mathbf{X}}(t), \quad t \in T \quad (22.4.4)$$

9. The error covariance matrix  $\mathbf{P}(t) = E\{\hat{\mathbf{X}}(t) - \mathbf{X}(t)\}[\hat{\mathbf{X}}(t) - \mathbf{X}(t)]^T$  is positive definite.

Under these assumptions the continuous Kalman filter can be given by the following equations. Since this is a continuous-time process, we cannot distinguish between across and between observations. The conditional expectation  $E^t[\mathbf{X}(t)] = \hat{\mathbf{X}}(t)$  satisfies the stochastic differential equation

$$\frac{d}{dt}\hat{\mathbf{X}}(t) = \mathbf{F}(t)\hat{\mathbf{X}}(t) + \mathbf{K}(t)[\mathbf{Y}(t) - \mathbf{H}(t)\hat{\mathbf{X}}(t)], \quad t \in T \quad (22.4.5)$$

and the covariance matrix  $\mathbf{P}(t)$  satisfies the stochastic matrix Riccati equation

$$\frac{d}{dt}\mathbf{P}(t) = \mathbf{F}(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{F}^T(t) + \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}^T(t) - \mathbf{K}(t)\mathbf{H}(t)\mathbf{P}(t), \quad t \in T \quad (22.4.6)$$

where the Kalman gain matrix  $\mathbf{K}(t)$  satisfies

$$\mathbf{K}(t) = \mathbf{P}(t)\mathbf{H}^T(t)\mathbf{R}^{-1}(t), \quad t \in T \quad (22.4.7)$$

We can show these results by discretizing the differential equations and comparing them to the discrete Kalman filter. Assuming a uniform discretization interval of  $\Delta t = t_k - t_{k-1}$ ,  $k = 1, \dots, N$ , we can write Eq. (22.4.1) as

$$\mathbf{X}_k = (\mathbf{I} + \mathbf{F}_{k-1}\Delta t)\mathbf{X}_{k-1} + \mathbf{G}_{k-1}\mathbf{W}_{k-1}\Delta t, \quad k \in \{1, 2, \dots, N\}, \quad \mathbf{X}_0 \quad (22.4.8)$$

where the subscript  $k$  stands for the value of the processes at time  $t = t_k$ . Comparing Eq. (22.4.8) to Eq. (22.3.9), we can identify

$$\mathbf{I} + \mathbf{F}_{k-1}\Delta t = \Phi_{k,k-1}; \quad \mathbf{G}_{k-1}\Delta t = \Gamma_{k-1} \quad (22.4.9)$$

From assumptions 3 and 4, the noise covariance matrices are given by

$$E[\mathbf{W}(t)\mathbf{W}^T(t)] = \frac{\mathbf{Q}(t)}{dt}; \quad E[\mathbf{V}(t)\mathbf{V}^T(t)] = \frac{\mathbf{R}(t)}{dt} \quad (22.4.10)$$

As the discretization interval is  $\Delta t$ , the covariance matrices of  $\mathbf{W}(t)$  and  $\mathbf{V}(t)$  at time  $t_k$  can be given in terms of the discrete covariance matrices  $\mathbf{Q}_k$  and  $\mathbf{R}_k$  as

$$\frac{\mathbf{Q}(t_k)}{\Delta t} = \mathbf{Q}_k; \quad \frac{\mathbf{R}(t_k)}{\Delta t} = \mathbf{R}_k \quad (22.4.11)$$

Since the limit will be taken as  $\Delta t \rightarrow 0$ , the distinction between  $t_k(+)$  and  $t_k(-)$  will be eliminated and written as  $t_k(+) = t_k(-) = t_k$ . After substituting Eqs. (22.4.9) and (22.4.11), the Kalman gain equation [Eq. (22.3.40)] can be rewritten for  $k = 1, \dots, N$ :

$$\begin{aligned} \mathbf{K}_{k-1} &= \mathbf{P}_{k-1}\mathbf{H}_{k-1}^T \left[ \mathbf{H}_{k-1}\mathbf{P}_{k-1}\mathbf{H}_{k-1}^T + \frac{\mathbf{R}(t_{k-1})}{\Delta t} \right]^{-1} \\ &= \mathbf{P}_{k-1}\mathbf{H}_{k-1}^T \left[ \mathbf{H}_{k-1}\mathbf{P}_{k-1}\mathbf{H}_{k-1}^T \Delta t + \mathbf{R}_{k-1} \right]^{-1} \Delta t \end{aligned} \quad (22.4.12)$$

Dividing Eq. (22.4.12) by  $\Delta t$  and taking the limit as  $\Delta t \rightarrow 0$ , we obtain

$$\lim_{\Delta t \rightarrow 0} \frac{\mathbf{K}_{k-1}}{\Delta t} = \mathbf{K}(t) = \mathbf{P}(t)\mathbf{H}^T(t)\mathbf{R}^{-1}(t), \quad t \in T \quad (22.4.13)$$

which is Eq. (22.4.7).

Similarly, substituting Eqs. (22.4.9) and (22.4.11) in the covariance equation [Eq. (22.3.39)], we can write

$$\begin{aligned} \mathbf{P}_k &= [\mathbf{I} + \mathbf{F}_{k-1}\Delta t][(\mathbf{I} - \mathbf{K}_{k-1}\mathbf{H}_{k-1})\mathbf{P}_{k-1}][\mathbf{I} + \mathbf{F}_{k-1}\Delta t]^T + \mathbf{G}_{k-1}\Delta t \frac{\mathbf{Q}(t_k)}{\Delta t} \mathbf{G}_{k-1}^T \Delta t \\ &= [\mathbf{I} + \mathbf{F}_{k-1}\Delta t]\mathbf{P}_{k-1}[\mathbf{I} + \mathbf{F}_{k-1}\Delta t]^T - [\mathbf{I} + \mathbf{F}_{k-1}\Delta t][\mathbf{K}_{k-1}\mathbf{H}_{k-1}\mathbf{P}_{k-1}] \\ &\quad \times [\mathbf{I} + \mathbf{F}_{k-1}\Delta t]^T + \mathbf{G}_{k-1}\mathbf{Q}(t_k)\mathbf{G}_{k-1}^T \Delta t, \quad k \in \{1, 2, \dots, N\} \end{aligned} \quad (22.4.14)$$

Substituting for  $\mathbf{K}_{k-1} = \mathbf{K}(t)\Delta t$  from Eq. (22.4.13) in Eq. (22.4.14), we obtain

$$\begin{aligned} \mathbf{P}_k - \mathbf{P}_{k-1} &= \mathbf{P}_{k-1}\mathbf{F}_{k-1}^T \Delta t + \mathbf{F}_{k-1}\mathbf{P}_{k-1}\Delta t + \mathbf{G}_{k-1}\mathbf{Q}(t_k)\mathbf{G}_{k-1}^T \Delta t \\ &\quad - \mathbf{K}(t)\mathbf{H}_{k-1}\mathbf{P}_{k-1}\Delta t + o(\Delta t) \end{aligned} \quad (22.4.15)$$

where  $\lim_{\Delta t \rightarrow 0} o(\Delta t) \rightarrow 0$ . Dividing Eq. (22.4.15) by  $\Delta t$  and taking the limit as  $\Delta t \rightarrow 0$ , we obtain

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} \frac{\mathbf{P}_k - \mathbf{P}_{k-1}}{\Delta t} &= \frac{d}{dt}\mathbf{P}(t) = \mathbf{P}(t)\mathbf{F}^T(t) + \mathbf{F}(t)\mathbf{P}(t) + \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}^T(t) \\ &\quad - \mathbf{K}(t)\mathbf{H}(t)\mathbf{P}(t), \quad t \in T \end{aligned} \quad (22.4.16)$$

which is Eq. (22.4.6).

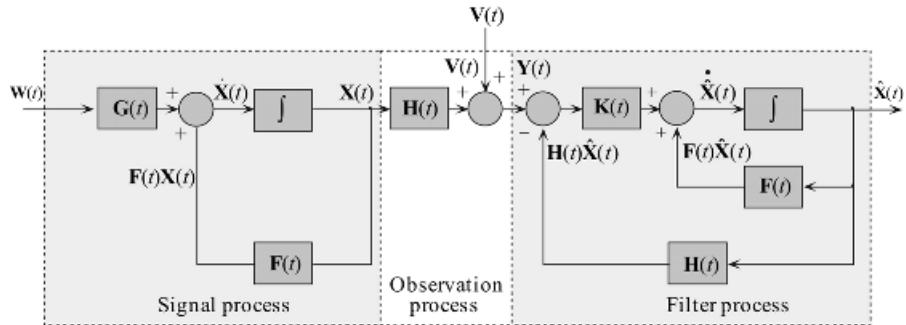


FIGURE 22.4.1

Finally, substituting the same Eqs. (22.4.9) and (22.4.11) in the state estimate equations [Eq. (22.3.38)], we can write

$$\begin{aligned}\hat{\mathbf{X}}_k &= [\mathbf{I} + \mathbf{F}_{k-1}\Delta t]\{\hat{\mathbf{X}}_{k-1} + \mathbf{K}_{k-1}[\mathbf{Y}_{k-1} - \mathbf{H}_{k-1}\hat{\mathbf{X}}_{k-1}]\} \\ &= \hat{\mathbf{X}}_{k-1} + \mathbf{F}_{k-1}\hat{\mathbf{X}}_{k-1}\Delta t + \mathbf{K}_{k-1}[\mathbf{Y}_{k-1} - \mathbf{H}_{k-1}\hat{\mathbf{X}}_{k-1}] \\ &\quad + \mathbf{F}_{k-1}\Delta t\mathbf{K}_{k-1}[\mathbf{Y}_{k-1} - \mathbf{H}_{k-1}\hat{\mathbf{X}}_{k-1}], \quad k \in \{1, 2, \dots, N\}\end{aligned}\quad (22.4.17)$$

Substituting for  $\mathbf{K}_{k-1} = \mathbf{K}(t)\Delta t$  from Eq. (22.4.13) in Eq. (22.4.17), we obtain

$$\hat{\mathbf{X}}_k - \hat{\mathbf{X}}_{k-1} = \mathbf{F}_{k-1}\hat{\mathbf{X}}_{k-1}\Delta t + \mathbf{K}(t)\Delta t[\mathbf{Y}_{k-1} - \mathbf{H}_{k-1}\hat{\mathbf{X}}_{k-1}] + o(\Delta t) \quad (22.4.18)$$

Dividing Eq. (22.4.18) by  $\Delta t$  and taking the limit as  $\Delta t \rightarrow 0$ , we obtain

$$\lim_{\Delta t \rightarrow 0} \frac{\hat{\mathbf{X}}_k - \hat{\mathbf{X}}_{k-1}}{\Delta t} = \frac{d}{dt}\hat{\mathbf{X}}(t) = \mathbf{F}(t)\hat{\mathbf{X}}(t) + \mathbf{K}(t)[\mathbf{Y}(t) - \mathbf{H}(t)\hat{\mathbf{X}}(t)], \quad t \in T \quad (22.4.19)$$

which is Eq. (22.4.5). The block diagram implementing the continuous Kalman filter is shown in Fig. 22.4.1.

**Example 22.4.1** This example is the continuous analog of the discrete example (Example 22.3.2). The zero mean signal and observation processes are given by

$$\frac{d}{dt}X(t) = 0, \quad t \in T; \quad Y(t) = X(t) + V(t), \quad t \in T$$

where  $V(t)$  is a zero mean white-noise process with variance  $\sigma_V^2$ . The variance of the initial state is  $p_0$ . In this example  $\mathbf{F}(t) = 0$ ,  $\mathbf{G}(t) = 0$  and  $\mathbf{H}(t) = 1$ .

The Kalman gain  $k(t)$  is obtained from Eq. (22.4.7):

$$k(t) = \frac{p(t)}{\sigma_V^2}$$

The variance is obtained from Eq. (22.4.6):

$$\frac{d}{dt}p(t) = -\frac{p^2(t)}{\sigma_V^2} \quad \text{or} \quad \frac{dp(t)}{p^2(t)} = -\frac{dt}{\sigma_V^2}$$

Integrating both sides of the equation from  $p_0$  to  $p(t)$ , we have

$$\int_{p_0}^{p(t)} \frac{dp(t)}{p^2(t)} = -\frac{1}{\sigma_V^2} \int_0^t dt \quad \text{or} \quad p(t) = \frac{p_0}{1 + p_0 t / \sigma_V^2}$$

The signal estimate from Eq. (22.4.5) is given by

$$\frac{d}{dt} \hat{X}(t) = \frac{p_0}{\sigma_V^2 + p_0 t} [Y(t) - \hat{X}(t)]$$

**Example 22.4.2** We will formulate a continuous version of Example 22.3.3. The signal and observation processes are given by

$$\frac{d}{dt} X(t) = -X(t) + W(t), \quad X_0; \quad Y(t) = 2X(t) + V(t)$$

The zero mean noise signals  $W(t)$  and  $V(t)$  are independent Gaussian-distributed processes with the variance parameters  $E[W^2(t)]dt = 1$  and  $E[V^2(t)]dt = 1$ .

The error variance from Eq. (22.4.6) is

$$\frac{d}{dt} p(t) = -2p(t) - 2p^2(t) + 1, \quad p_0 \quad (\text{A})$$

This equation is a Riccati equation, and one method of solving it is to convert it to a linear differential equation. Otherwise, it can be solved numerically. The Kalman gain and the state estimate hinge on finding this solution. Following the procedure in Ref. 57, we will assume a solution to the preceding equation in the form

$$p(t) = a + b \tanh(ct + d)$$

where  $a, b, c, d$  are constants to be determined. Differentiating  $p(t)$  and substituting in Eq. (A), we obtain

$$\begin{aligned} bc[1 - \tanh^2(ct + d)] &= 1 - 2[a + b \tanh(ct + d)] - 2[a + b \tanh(ct + d)]^2 \\ &= 1 - 2a - 2a^2 - 2b(1 + 2a) \tanh(ct + d) - 2b^2 \tanh^2(ct + d) \end{aligned}$$

Equating like coefficients of  $\tanh(ct + d)$  on both sides of the equation above, we obtain

$$a = -\frac{1}{2}; \quad b = \pm \frac{\sqrt{3}}{2}; \quad c = \pm \sqrt{3}$$

Since  $p(t)$  as a variance function has to be positive, we use the positive values for  $b$  and  $c$  in  $p(t)$  and write

$$p(t) = -\frac{1}{2} + \frac{\sqrt{3}}{2} \tanh(\sqrt{3}t + d)$$

where the unknown constant  $d$  has to be obtained from the initial condition  $p_0$ . Thus

$$\tanh(d) = \frac{2p_0 + 1}{\sqrt{3}}; \quad d = \tanh^{-1} \left( \frac{2p_0 + 1}{\sqrt{3}} \right)$$

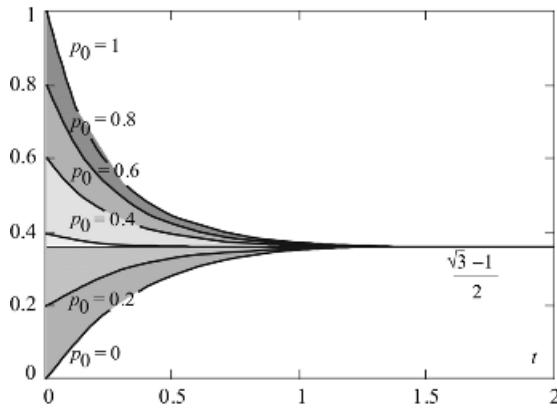


FIGURE 22.4.2

Hence

$$p(t) = -\frac{1}{2} + \frac{\sqrt{3}}{2} \tanh \left[ \sqrt{3}t + \tanh^{-1} \left( \frac{2p_0 + 1}{\sqrt{3}} \right) \right], \quad 0 \leq t < \infty \quad (\text{B})$$

$p(t)$  is shown in Fig. 22.4.2 for values of the initial condition  $p_0 = 0, 0.2, 0.4, 0.6, 0.8, 1$ .

From Eq. (B) and from Fig. 22.4.2 we see that as  $t \rightarrow \infty$ ,  $p(t)$  tends to constant given by

$$\lim_{t \rightarrow \infty} p(t) = -\frac{1}{2} + \frac{\sqrt{3}}{2} \tanh[\infty] = \frac{\sqrt{3} - 1}{2}$$

which is independent of the initial condition. This is a desirable situation where the final value of the uncertainty is a constant independent of the initial conditions because they may not be known a priori to any degree of certainty.

From Eq. (22.4.7), the Kalman gain  $k(t)$  is given by

$$k(t) = 2p(t)$$

The signal estimate is obtained from Eq. (22.4.5):

$$\frac{d}{dt} \hat{X}(t) = -\hat{X}(t) + 2p(t)[Y(t) - 2\hat{X}(t)]$$

Here also as  $t \rightarrow \infty$ , the final value of the estimate  $\hat{X}(t)$  will be independent of the initial condition.

# Probabilistic Methods in Transmission Tomography\*

Probabilistic methods are used extensively in medical imaging. In this chapter Bayesian reconstruction algorithms for transmission tomographic images are presented. The reconstruction is based on maximum a posteriori (MAP) estimation using the expectation–maximization (EM) algorithm. Gibbs’ density is used as the prior distribution with a sigmoidal potential function. The reconstruction problem itself is solved through a one-step-late expectation–maximization algorithm (EM-OSL) proposed by Green [9].<sup>†</sup> Suitable approximations are made to render the problem more tractable for parallel computation. Computer-simulated phantoms are used for image reconstruction. These images are compared to those reconstructed from currently available reconstruction algorithms in terms of image quality and convergence rate. This algorithm works quite well for low-photon-count and low-contrast cases in which the widely used convolution backprojection algorithm performs rather poorly.

## 23.1 INTRODUCTION

In transmission tomography (TT) the acquired projection data contain high statistical variations due to limited dose of radiation introduced in patients. Deterministic reconstruction techniques such as convolution backprojection and other Fourier-based methods introduce various types of artifacts. Transmission tomography models for reconstruction takes into account the probabilistic nature of the problem by incorporating statistical variations and other physical phenomena in a stochastic model. Application of the maximum-likelihood estimation to image reconstruction problem was recognized by Rockmore and Macovski [1,2]. An alternative to the maximum-likelihood approach was provided by Dempster et al. [3] with a general iterative method known as the *expectation–maximization* (EM) algorithm. The relevance of the EM algorithm to maximum-likelihood image reconstruction

\*This chapter is based on the work [14] carried out by the author and his student at the Center of Advanced Computation and Telecommunications, Univ. Massachusetts Lowell and reported in an internal memorandum, *Modified EM Algorithm for Transmission Tomography*, T94-I, 1994.

<sup>†</sup>References cited in this chapter are listed in Section 23.8 (not in the References list at the end of the book).

was noticed by Shepp and Vardi [4], who applied it to a stochastic model of positron emission tomography (PET), and independently by Lange and Carson [5], who applied it to general models of emission and transmission tomography.

The initial optimism of the EM method of reconstruction in emission tomography has been tempered by the grainy, speckled appearance of the reconstructed images. This graininess is a manifestation of low correlations between parameter estimates for neighboring pixels. A satisfying solution is to incorporate some smoothness criterion in the reconstruction process itself.

In emission tomography, this perspective has been advocated by Geman and McClure [6,7], Hebert and Leahy [8], and Green [9], who build on previous works of Geman and Geman [10] and Besag [11]. In particular, Geman and McClure [6,7] introduce Gibbs' priors with nearest-neighbor interactions. Green [9] devises an approximate EM algorithm incorporating nearest-neighbor interactions with the one-step-late expectation–maximization (EM-OSL) algorithm. Lange [12] has demonstrated the conditions under which the EM-OSL algorithm converges to the unique maximum posterior point in image space for both emission and transmission cases.

In this chapter Gibbs' prior with sigmoidal potential function is developed that applies the EM-OSL scheme of Green [9] to the transmission algorithm proposed by Lange [12]. In simplifying the resultant transcendental equation, three terms have been used in the series expansion rather than the two in Lange [12]. The solution of the resulting quadratic equation yields more accurate results. The sigmoidal potential function smooths out significant differences in neighboring pixel values while retaining sharp-edged features of images.

## 23.2 STOCHASTIC MODEL

### Notation

Before describing the model in detail, some of the notations are defined in Fig. 23.2.1:

$i$	Projection subscript
$K$	Number of projection paths
$j$	Ordered pixel subscript for each projection path $i$ starting from the pixel farthest from the detector $i$
$\{I_i\}$	Set of $(m^* - 1)$ pixels contributing to the projection path $i$
$\{J_j\}$	Set of projections to which pixel $j$ contributes
$Y_i$	Random variable for detected photons from projection $i$ , or $Y_i = X_{im}$
$X_{ij}$	Random variable for photons from projection $i$ <i>entering</i> pixel $j$
$\hat{X}_{ij}$	Random variable for photons from projection $i$ <i>leaving</i> pixel $j$ , or entering pixel $j + 1$ , that is, $\hat{X}_{ij} = X_{ij+1}$
$M_j$	Random variable for attenuation coefficient of pixel $j$
$\mu_j$	Realization of $M_j$
$l_{ij}$	Pathlength of projection $i$ through pixel $j$
$N$	Neighboring pixels

\*The subscript  $i$  in  $m_i$  has been omitted for clarity of notation.

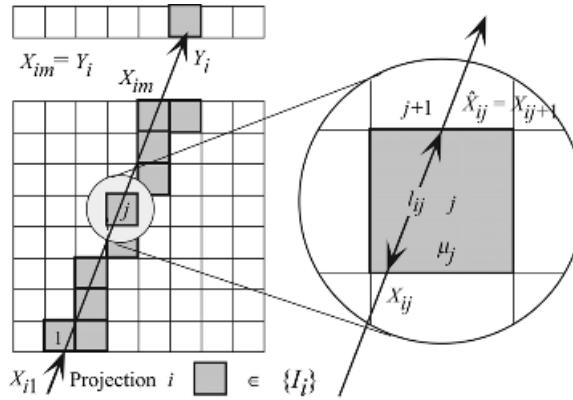


FIGURE 23.2.1

### Probability Model

Corresponding to each projection  $i$ , the source photons are assumed to travel along a thin path en route to the detector. This path cuts through a set  $\{I_i\}$  of  $(m-1)$  pixels ordered by the index  $j$ ; that is,  $j = 1$  is the pixel nearest to the source, and pixel  $j = m-1$  is the last pixel along the path. The attenuation of pixel  $j$  is a random variable  $M_j$  taking the value  $\mu_j$  and the length of the path inside pixel  $j$  is  $l_{ij}$ . Photons traveling along path  $i$  are counted as  $X_{ij}$  at the entrance to pixel  $j$  and as  $\hat{X}_{ij} = X_{ij+1}$  at the exit, with  $X_{ij} - \hat{X}_{ij}$  photons becoming attenuated in pixel  $j$ . The number of photons  $X_{im}$  leaving the last pixel  $(m-1)$  is the same as the number of photons  $Y_i$  being detected by the detector  $i$ , or  $X_{im} = Y_i$ . The attenuation coefficients  $\{\mu_j\}$  have to be estimated from the observed photon count  $X_{im} = Y_i$ .

The random variable  $X_{i1}$  of photons leaving the source at projection  $i$  is assumed to be Poisson-distributed as

$$P(X_{i1} = x_{i1}) = e^{-d_i} \frac{d_i^{x_{i1}}}{x_{i1}!} \quad (23.2.1)$$

where  $d_i = E[X_{i1}]$  is the mean value of photons leaving the source in projection  $i$ .

The probability that a photon entering pixel  $j$  along path  $i$  will survive the attenuation at pixel  $j$ , conditioned on the attenuation coefficient  $M_j = \mu_j$ , is assumed to be  $P_j = e^{-\mu_j l_{ij}}$ . The probability that  $\hat{X}_{ij}$  photons will emanate from pixel  $j$  conditioned on both  $X_{ij} = x_{ij}$  photons entering pixel  $j$ , and on the attenuation coefficient  $M_j = \mu_j$ , can be characterized by a binomial distribution:

$$P\{\hat{X}_{ij} = \hat{x}_{ij} | X_{ij} = x_{ij} | M_j = \mu_j\} = \binom{x_{ij}}{\hat{x}_{ij}} (e^{-\mu_j l_{ij}})^{\hat{x}_{ij}} (1 - e^{-\mu_j l_{ij}})^{(x_{ij} - \hat{x}_{ij})} \quad (23.2.2)$$

From Eq. (23.2.1) the conditional mean value,  $\gamma_{ij} = E[X_{ij} | \mu_1, \mu_2, \dots, \mu_{j-1}]$  of the photons,  $X_{ij}$  entering pixel  $j$  along the projection  $i$  can be given by

$$\gamma_{ij} = E[X_{ij} | \mu_{j-1}] = d_i (e^{\mu_1 l_{i1}}) (e^{\mu_2 l_{i2}}) \cdots (e^{\mu_{j-1} l_{ij-1}}) = d_i e^{-\sum_{k=1}^{j-1} \mu_k l_{ik}} \quad (23.2.3)$$

where  $\mu_{j-1}$  represents the attenuation coefficient vector  $\mu_{j-1} = [\mu_1, \mu_2, \dots, \mu_{j-1}]^T$ . Similarly, the conditional mean value,  $\gamma_{im} = E[X_{im} | \mu_1, \mu_2, \dots, \mu_{m-1}]$  of the detected photons  $X_{im}$  is given by

$$\gamma_{im} = E[X_{im} | \mu_{m-1}] = d_i (e^{\mu_1 l_{i1}}) (e^{\mu_2 l_{i2}}) \cdots (e^{\mu_{m-1} l_{im-1}}) = d_i e^{-\sum_{k=1}^{m-1} \mu_k l_{ik}} \quad (23.2.4)$$

where  $\boldsymbol{\mu}_{m-1}$  represents the attenuation coefficient vector  $\boldsymbol{\mu}_{m-1} = [\mu_1, \mu_2, \dots, \mu_{m-1}]^T$ . The ratio  $\gamma_{im}/\gamma_{ij}$  can be expressed as follows:

$$\frac{\gamma_{im}}{\gamma_{ij}} = \frac{E[X_{im} | \boldsymbol{\mu}_{m-1}]}{E[X_{ij} | \boldsymbol{\mu}_{j-1}]} = \frac{d_i e^{-\sum_{k=1}^{m-1} \mu_k l_{ik}}}{d_i e^{-\sum_{k=1}^{j-1} \mu_k l_{ik}}} = e^{-\sum_{k=j}^{m-1} \mu_k l_{ik}} \quad (23.2.5)$$

Since the input random variable  $X_{ij}$  to pixel  $j$  conditioned on  $\boldsymbol{\mu}_{j-1} = [\mu_1, \mu_2, \dots, \mu_{j-1}]^T$  is Poisson-distributed with mean  $\gamma_{ij}$  given by Eq. (23.2.3), the conditional distribution of  $X_{ij}$  is

$$P[X_{ij} = x_{ij} | \boldsymbol{\mu}_{j-1}] = e^{-\gamma_{ij}} \frac{\gamma_{ij}^{x_{ij}}}{x_{ij}!} \quad (23.2.6)$$

The probability of the detected photons  $X_{im}$  conditioned on  $X_{ij}$  photons entering pixel  $j$ , and the attenuation vector  $\boldsymbol{\mu}_{m-1} = [\mu_1, \mu_2, \dots, \mu_{m-1}]^T$ , can again be given by a binomial distribution similar to Eq. (23.2.2):

$$\begin{aligned} P\{X_{im} = x_{im} | X_{ij} = x_{ij} | \boldsymbol{\mu}_{m-1}\} \\ = \binom{x_{ij}}{x_{im}} \left( e^{-\sum_{k=j}^{m-1} \mu_k l_{ik}} \right)^{x_{im}} \left( 1 - e^{-\sum_{k=j}^{m-1} \mu_k l_{ik}} \right)^{(x_{ij} - x_{im})} \end{aligned} \quad (23.2.7)$$

The ratio  $\gamma_{im}/\gamma_{ij}$  from Eq. (23.2.5) is substituted in Eq. (23.2.7) and rewritten as

$$P\{X_{im} = x_{im} | X_{ij} = x_{ij} | \boldsymbol{\mu}_{m-1}\} = \binom{x_{ij}}{x_{im}} \left( \frac{\gamma_{im}}{\gamma_{ij}} \right)^{x_{im}} \left( 1 - \frac{\gamma_{im}}{\gamma_{ij}} \right)^{(x_{ij} - x_{im})} \quad (23.2.8)$$

Using Bayes' theorem [Eq. (2.4.4)] the probability of  $X_{ij}$  conditioned on the detected photons  $X_{im} = x_{im} = y_i$ , and the attenuation coefficient vector  $\boldsymbol{\mu}_{m-1}$  can be written as

$$\begin{aligned} P\{X_{ij} = x_{ij} | X_{im} = x_{im} | \boldsymbol{\mu}_{m-1}\} \\ = \frac{P\{X_{im} = x_{im} | X_{ij} = x_{ij} | \boldsymbol{\mu}_{m-1}\} P\{X_{ij} = x_{ij} | \boldsymbol{\mu}_{m-1}\}}{P\{X_{im} = x_{im} | \boldsymbol{\mu}_{m-1}\}} \\ = \frac{\left( \frac{x_{ij}!}{x_{im}!(x_{ij} - x_{im})!} \right) \left( \frac{\gamma_{im}}{\gamma_{ij}} \right)^{x_{im}} \left( 1 - \frac{\gamma_{im}}{\gamma_{ij}} \right)^{(x_{ij} - x_{im})} e^{-\gamma_{ij}} \frac{\gamma_{ij}^{x_{ij}}}{x_{ij}!}}{e^{-\gamma_{im}} \frac{\gamma_{im}^{x_{im}}}{x_{im}!}} \end{aligned} \quad (23.2.9)$$

Simplifying Eq. (23.2.9), we obtain a conditional Poisson distribution

$$\begin{aligned} P\{X_{ij} = x_{ij} | X_{im} = x_{im} | \boldsymbol{\mu}_{m-1}\} \\ = \frac{1}{(x_{ij} - x_{im})!} (\gamma_{ij} - \gamma_{im})^{(x_{ij} - x_{im})} e^{-(\gamma_{ij} - \gamma_{im})} \end{aligned} \quad (23.2.10)$$

The mean of the photons  $X_{ij}$  conditioned on the observed detected photons  $X_{im} = x_{im} = y_i$ , and attenuation coefficient vector  $\boldsymbol{\mu}_{m-1}$  can be obtained from by taking the expected

value of Eq. (23.2.10):

$$\begin{aligned}
 E[X_{ij} | X_{im} = Y_i = x_{im} | \boldsymbol{\mu}_{m-1}] &= \sum_{x_{ij}=0}^{\infty} \frac{x_{ij}}{(x_{ij} - x_{im})!} (\gamma_{ij} - \gamma_{im})^{(x_{ij} - x_{im})} e^{-(\gamma_{ij} - \gamma_{im})} \\
 &= \sum_{x_{ij}=0}^{\infty} \frac{x_{ij} - x_{im} + x_{im}}{(x_{ij} - x_{im})!} (\gamma_{ij} - \gamma_{im})^{(x_{ij} - x_{im})} e^{-(\gamma_{ij} - \gamma_{im})} \\
 &= \gamma_{ij} - \gamma_{im} + x_{im} = \gamma_{ij} - \gamma_{im} + y_i \stackrel{\Delta}{=} N_{ij}
 \end{aligned} \quad (23.2.11)$$

We will now find the joint conditional probability of the photons from projection  $i$  entering pixels  $\{j, j = m, m-1, \dots, 1\}$  conditioned on the attenuation coefficient vector  $\boldsymbol{\mu}_{m-1}$ . Defining an  $m$ -vector  $\mathbf{X}_i = \{X_{im}, X_{im-1}, \dots, X_{i1}\}^T$ , the joint conditional probability,  $P\{\mathbf{X}_i | \boldsymbol{\mu}_{m-1}\}$  of  $\mathbf{X}_i$  conditioned on the attenuation coefficient vector  $\boldsymbol{\mu}_{m-1}$ , can be evaluated by using the Markov property of  $\mathbf{X}_i$ :

$$\begin{aligned}
 P\{\mathbf{X}_i | \boldsymbol{\mu}_{m-1}\} &= P\{X_{im}, X_{im-1}, \dots, X_{i1} | \boldsymbol{\mu}_{m-1}\} \\
 &= P\{X_{im} | X_{im-1} | \boldsymbol{\mu}_{m-1}\} P\{X_{im-1} | X_{im-2} | \boldsymbol{\mu}_{m-2}\} \\
 &\quad \times \cdots P\{X_{i2} | X_{i1} | \boldsymbol{\mu}_1\} P\{X_{i1}\}
 \end{aligned} \quad (23.2.12)$$

But by analogy to Eq. (23.2.2), the conditional probability of  $X_{ik+1}$  conditioned on  $X_{ik}$  and  $\mu_k$  is given by a binomial distribution:

$$\begin{aligned}
 P\{X_{ik+1} = x_{ik+1} | X_{ik} = x_{ik} | M_k = \mu_k\} \\
 = \binom{x_{ik}}{x_{ik+1}} (e^{-\mu_k l_{ik}})^{x_{ik+1}} (1 - e^{-\mu_k l_{ik}})^{(x_{ik} - x_{ik+1})}
 \end{aligned} \quad (23.2.13)$$

Substituting Eqs. (23.2.13) and (23.2.1) into Eq. (23.2.12), we obtain

$$P\{\mathbf{X}_i | \boldsymbol{\mu}_{m-1}\} = \left[ \prod_{k=1}^{m-1} \binom{x_{ik}}{\hat{x}_{ik}} (e^{-\mu_k l_{ik}})^{\hat{x}_{ik}} (1 - e^{-\mu_k l_{ik}})^{(x_{ik} - \hat{x}_{ik})} \right] e^{-d_i} \frac{d_i^{x_{i1}}}{x_{i1}!} \quad (23.2.14)$$

where we have used  $\hat{x}_{ik} = x_{ik+1}$ .

Equation (23.2.14) represents the conditional probability for the projection path  $i$ . Similar conditional probabilities can be formulated for all the  $K$  projection paths.

We define matrices  $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K\}^T$ ,  $\mathbf{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K\}^T$ , and  $\mathbf{M} = \{\text{the set of attenuation coefficient vectors } \boldsymbol{\mu}_{m_i-1} \text{ for all projections } i = 1, \dots, K\}$ . Summing Eq. (23.2.14) over all projections  $i$ , the following conditional probability can be written as

$$P\{\mathbf{X} | \mathbf{M}\} = \prod_{i=1}^K \left[ \prod_{k=1}^{m-1} \binom{x_{ik}}{\hat{x}_{ik}} (e^{-\mu_k l_{ik}})^{\hat{x}_{ik}} (1 - e^{-\mu_k l_{ik}})^{(x_{ik} - \hat{x}_{ik})} \right] e^{-d_i} \frac{d_i^{x_{i1}}}{x_{i1}!} \quad (23.2.15)$$

and the loglikelihood function  $\ln P\{\mathbf{X} | \mathbf{M}\}$  can be written as

$$\ln P\{\mathbf{X} | \mathbf{M}\} = \sum_{i=1}^K \sum_{k=1}^{m-1} [\hat{x}_{ik} \ln (e^{-\mu_k l_{ik}}) + (x_{ik} - \hat{x}_{ik}) \ln (1 - e^{-\mu_k l_{ik}})] + C \quad (23.2.16)$$

where the constant  $C$  represents the logarithm of terms independent of  $\mu_k$ .

The estimate of the attenuation matrix  $\mathbf{M}$  is obtained by maximizing the posterior probability of  $\mathbf{M}$  conditioned on the observed but incomplete data  $\mathbf{Y}$ , that is  $P\{\mathbf{M} | \mathbf{Y}\}$

or the corresponding loglikelihood function  $\ln P\{\mathbf{M} | \mathbf{Y}\}$ . However, maximizing  $\ln P\{\mathbf{M} | \mathbf{Y}\}$  is an intractable problem. The problem is mitigated by embedding the observed but incomplete space  $\mathbf{Y}$  into the complete but unobservable space  $\mathbf{X}$ . Since  $\mathbf{X}$  is connected to  $\mathbf{Y}$  by a many-to-one mapping, the richer space  $\mathbf{X}$  provides more flexibility in data manipulation, resulting in an easier maximization of  $\ln P\{\mathbf{M} | \mathbf{X}\}$ . Using Bayes' theorem, the likelihood function  $P\{\mathbf{M} | \mathbf{X}\}$  will be given by

$$P\{\mathbf{M} | \mathbf{X}\} = \frac{P\{\mathbf{X} | \mathbf{M}\}P\{\mathbf{M}\}}{P\{\mathbf{X}\}} \quad (23.2.17)$$

and the corresponding loglikelihood function is

$$\ln P\{\mathbf{M} | \mathbf{X}\} = \ln P\{\mathbf{X} | \mathbf{M}\} + \ln P\{\mathbf{M}\} - \ln P\{\mathbf{X}\} \quad (23.2.18)$$

Equations (23.2.17) or (23.2.18) will form the basis for the estimation procedure.

### 23.3 STOCHASTIC ESTIMATION ALGORITHM

The maximum-likelihood estimator for  $\mathbf{M}$  is obtained by assuming that the attenuation coefficients  $\{\mu_k\}$  are equiprobable, in which case  $P\{\mathbf{M}\}$  is constant. Since  $P\{\mathbf{X}\}$  is independent of  $\mathbf{M}$ , maximizing  $P\{\mathbf{M} | \mathbf{X}\}$  amounts to maximizing the loglikelihood function  $\ln P\{\mathbf{X} | \mathbf{M}\}$  given in Eq. (23.2.16) with respect to  $\mathbf{M}$ . Since it contains the unknown values of random variables  $X_{ik}$  and  $\hat{X}_{ik}$ , direct maximization of Eq. (23.2.16) will not give meaningful results. Hence its expected value will be maximized with respect to  $\{\mu_k\}$ . From Eq. (23.2.11), the conditional expectations of  $X_{ik}$  and  $\hat{X}_{ik}$  are functions of  $\{\mu_k\}$  via  $\{\gamma_{ik}\}$ , and hence an iterative method of maximizing Eq. (23.2.18) has to be developed.

The projection matrix  $\mathbf{Y}$  is some function  $\mathbf{g}(\mathbf{X})$  where  $\mathbf{g}(\cdot)$  is a many-to-one mapping from the space of  $\mathbf{X}$  to the space of  $\mathbf{Y}$ . Hence the iterative solution to the maximum-likelihood estimation problem will proceed in alternating steps of expectation and maximization. First the expectation of  $\ln P\{\mathbf{X} | \mathbf{M}\}$  is taken conditioned on the observation matrix  $\mathbf{Y}$  and the current estimate  $\mathbf{M}^{(n)}$ . This conditional expected value  $E[\ln P\{\mathbf{X} | \mathbf{M}\} | \mathbf{Y}, \mathbf{M}^{(n)}]$  is maximized with respect to  $\mathbf{M}$  to yield an updated estimate  $\mathbf{M}^{(n+1)}$ . The updated conditional expected value  $E[\ln P\{\mathbf{X} | \mathbf{M}\} | \mathbf{Y}, \mathbf{M}^{(n+1)}]$  is again maximized and the process repeated. According to Dempster [3], the loglikelihood function  $\ln P\{\mathbf{M} | \mathbf{Y}\}$  or equivalently  $\ln P\{\mathbf{Y} | \mathbf{M}\}$  will increase in value at each iteration step with the result that

$$\ln P\{\mathbf{Y} | \mathbf{M}^{(n+1)}\} \geq \ln P\{\mathbf{Y} | \mathbf{M}^{(n)}\} \quad (23.3.1)$$

and the equality holds only when the maximum-likelihood solution  $\mathbf{M}_{\max}$  is reached and the maximum value is  $\ln P\{\mathbf{Y} | \mathbf{M}_{\max}\}$ .

#### Expectation Step

If  $\mathbf{M}^{(n)}$  is the estimate of the attenuation matrix  $\mathbf{M}$  at the  $n$ th iteration, the expectation of  $\ln P\{\mathbf{X} | \mathbf{M}\}$  at the  $(n+1)$ st iteration, conditioned on the observation  $\mathbf{Y}$  and  $\mathbf{M}^{(n)}$ , can be written as

$$E[\ln P\{\mathbf{X} | \mathbf{M}\} | \mathbf{Y}, \mathbf{M}^{(n)}] = \sum_{i=1}^K \sum_{k=1}^{m-1} \left[ \hat{N}_{ik} \ln(e^{-\mu_k l_{ik}}) + (N_{ik} - \hat{N}_{ik}) \ln(1 - e^{-\mu_k l_{ik}}) \right] + C_1 \quad (23.3.2)$$

where the conditional expectations  $N_{ik}$  and  $\hat{N}_{ik}$  conditioned on  $\mathbf{Y}$  and  $\mathbf{M}^{(n)}$  are obtained from Eq. (23.2.11) as

$$\begin{aligned} N_{ik} &= E[X_{ik} | Y_i | \mathbf{M}^{(n)}] = \gamma_{ik} - \gamma_{im} + y_i \\ \hat{N}_{ik} &= E[\hat{X}_{ik} | Y_i | \mathbf{M}^{(n)}] = \gamma_{ik+1} - \gamma_{im} + y_i \end{aligned} \quad (23.3.3)$$

where  $\hat{X}_{ik} = X_{ik+1}$  and  $C_1$  is yet another constant independent of  $\{\mu_k\}$ .

### Maximization Step

Since  $N_{ik}$  and  $\hat{N}_{ik}$  are known,  $E[\ln P\{\mathbf{X} | \mathbf{M}\} | \mathbf{Y}, \mathbf{M}^{(n)}]$  can now be maximized by differentiating Eq. (23.3.2) with respect to  $\mu_k$  and setting it equal to 0:

$$\frac{\partial}{\partial \mu_k} E[\ln P\{\mathbf{X} | \mathbf{M}\} | \mathbf{Y}, \mathbf{M}^{(n)}] = \sum_{i \in J_k} \left[ -\hat{N}_{ik} l_{ik} + (N_{ik} - \hat{N}_{ik}) \frac{l_{ik}}{e^{\mu_k l_{ik}} - 1} \right] = 0 \quad (23.3.4)$$

Equation (23.3.4) is a transcendental equation that can be solved by approximating the exponential term by using the first three terms in the series expansion:

$$\frac{1}{e^{\mu_k l_{ik}} - 1} = \frac{1}{\mu_k l_{ik}} - \frac{1}{2} + \frac{1}{12} \mu_k l_{ik} + o[(\mu_k l_{ik})^2] \quad (23.3.5)$$

### Expectation–Maximization Algorithm

Substitution of Eq. (23.3.5) in Eq. (23.3.4) results in

$$\left[ \frac{1}{12} \sum_{i \in J_k} (N_{ik} - \hat{N}_{ik}) l_{ik}^2 \right] \mu_k^2 - \left[ \frac{1}{2} \sum_{i \in J_k} (N_{ik} + \hat{N}_{ik}) l_{ik} \right] \mu_k + \left[ \sum_{i \in J_k} (N_{ik} - \hat{N}_{ik}) \right] = 0$$

or

$$A \mu_k^2 - B \mu_k + C = 0 \quad \text{for all } k \quad (23.3.6)$$

Solving for  $\mu_k$  in the quadratic equation [Eq. (23.3.6)] yields the  $(n+1)$ st estimate  $\mu_k^{(n+1)}$

$$\mu_k^{(n+1)} = \frac{B \pm \sqrt{B^2 - 4AC}}{2A} \Big|_{\mu_k^{(n)}} \quad \text{for all } k \quad (23.3.7)$$

where the coefficients  $A$ ,  $B$ , and  $C$  are dependent on  $\mu_k^{(n)}$ . The smaller root has to be chosen as the solution in this algorithm so that for  $l_{ik} \mu_k < 6$  this solution will always fall inside the solution bounds established by Lange [5] for all values of  $k$  and  $i$ .

As  $n \rightarrow \infty$  in Eq. (23.3.7),  $\mu_k^{(n)} \rightarrow \mu_k^{\max}$ , and hence it is the desired expectation–maximization (EM) algorithm.

### EM-One-Step-Late (OSL) Algorithm

Equation (23.3.7) represents the maximum-likelihood solution with equiprobable attenuation coefficients  $\mu_k^{(n+1)}$ . This solution can be considerably refined by using prior information  $P\{\mathbf{M}\}$  about the attenuation coefficients  $\mathbf{M}$ , giving rise to the maximum

a posteriori solution. In this case, in Eq. (23.2.18) only  $P\{\mathbf{X}\}$  does not depend on  $\mathbf{M}$ . Applying the same iterative techniques to Eq. (23.2.18) as used in deriving Eq. (23.3.4), there results an equation similar to Eq. (23.3.4) as shown below:

$$\begin{aligned}\frac{\partial}{\partial \mu_k} E[\ln P\{\mathbf{M} | \mathbf{X}\} | \mathbf{Y}, \mathbf{M}^{(n)}] &= \frac{\partial}{\partial \mu_k} E[\ln P\{\mathbf{X} | \mathbf{M}\} | \mathbf{Y}, \mathbf{M}^{(n)}] + \frac{\partial}{\partial \mu_k} \ln P\{\mathbf{M}\} \\ &= \sum_{i \in J_k} \left[ -\hat{N}_{ik} l_{ik} + (N_{ik} - \hat{N}_{ik}) \frac{l_{ik}}{e^{\mu_k l_{ik}} - 1} \right] + \frac{\partial}{\partial \mu_k} \ln P\{\mathbf{M}\} = 0\end{aligned}\quad (23.3.8)$$

Using the same series approximation for the exponential in Eq. (23.3.8) as in Eq. (23.3.5) leads to the equation

$$\begin{aligned}A &= \left[ \frac{1}{12} \sum_{i \in J_k} (N_{ik} - \hat{N}_{ik}) l_{ik}^2 \right] \mu_k^2 - \left[ \frac{1}{2} \sum_{i \in J_k} (N_{ik} - \hat{N}_{ik}) l_{ik} - \frac{\partial}{\partial \mu_k} \ln P\{\mathbf{M}\} \right] \mu_k \\ C &= \left[ \sum_{i \in J_k} (N_{ik} - \hat{N}_{ik}) \right] = 0\end{aligned}\quad (23.3.9)$$

where  $A$ ,  $B$ , and  $C$  are the same quantities defined in Eq. (23.3.6), Eq. (23.3.9) can be rewritten as follows:

$$A \mu_k^2 - \left( B - \frac{\partial}{\partial \mu_k} \ln P\{\mathbf{M}\} \right) \mu_k + C = 0 \quad \text{for all } k \quad (23.3.10)$$

Solving for  $\mu_k$  in the preceding quadratic yields the result

$$\mu_k^{(n+1)} = \frac{\left( B - \frac{\partial}{\partial \mu_k^{(n+1)}} \ln P\{\mathbf{M}\} \right) \pm \sqrt{\left( B - \frac{\partial}{\partial \mu_k^{(n+1)}} \ln P\{\mathbf{M}\} \right)^2 - 4AC}}{2A} \Bigg|_{\mu_k^{(n)}} \quad (23.3.11)$$

Equation (23.3.11) cannot be solved since it involves the derivative of the log-prior distribution at  $\mu_k^{(n+1)}$ . Green [9] resolves the problem by using the derivative at  $\mu_k^{(n)}$  yielding the EM one-step-late algorithm (EM-OSL). Additionally, the derivative of the prior  $P(\mathbf{M})$  is multiplied with a parameter  $\beta$  to add flexibility on the influence of the prior on the algorithm. With these adjustments, Eq. (23.3.11) can be rewritten as follows:

$$\mu_k^{(n+1)} = \frac{\left( B - \beta \frac{\partial}{\partial \mu_k^{(n)}} \ln P\{\mathbf{M}\} \right) \pm \sqrt{\left( B - \beta \frac{\partial}{\partial \mu_k^{(n)}} \ln P\{\mathbf{M}\} \right)^2 - 4AC}}{2A} \Bigg|_{\mu_k^{(n)}} \quad (23.3.12)$$

In Eq. (23.3.12) the smaller root of Eq. (23.3.12) is chosen in this iterative algorithm for the same convergence reasons as in Eq. (23.3.7). The optimum value of  $\beta$  will depend on both the characteristics of the image and the prior used. If  $\beta = 0$ , then Eq. (23.3.12) is the EM algorithm of Eq. (23.3.7).

## 23.4 PRIOR DISTRIBUTION $P\{\mathbf{M}\}$

The prior distribution  $P\{\mathbf{M}\}$  in Eq. (23.3.12) has to be carefully chosen. From consideration of Markovian fields, a proper choice is the exponential distribution given by  $P\{\boldsymbol{\mu}\} = (1/z)e^{-V(\boldsymbol{\mu})}$ ,  $\boldsymbol{\mu} \geq 0$ , where  $z = \int_{\boldsymbol{\mu} \geq 0} e^{-V(\boldsymbol{\mu})} d\boldsymbol{\mu}$  is the normalization constraint. This is known as the *Gibbs distribution* discussed in Section 7.15. The potential function  $V(\boldsymbol{\mu})$  is designed to penalize large differences in estimated parameters for neighboring pixels. The constraints on the potential functions in Section 7.15 are discussed by Lange [12]. In this application, a first-order neighborhood interaction of the form

$$V(\boldsymbol{\mu}) = \sum_{\substack{(j,k) \in N \\ N \sim \text{neighboring pixels}}} w_{jk} v(\mu_j - \mu_k), \quad (23.4.1)$$

is considered where the weighting functions  $w_{jk}$  for a first-order Markovian field are defined by

$$w_{jk} = \begin{cases} 1: & \text{orthogonal pixel} \\ 1/\sqrt{2}: & \text{diagonal pixel} \\ 0: & \text{otherwise} \end{cases}$$

and are shown in Fig. 23.4.1.

With Eq. (23.4.1), the prior distribution  $P\{\boldsymbol{\mu}\}$  is of the form

$$P\{\boldsymbol{\mu}\} = \frac{1}{z} e^{-\sum_{(j,k) \in N} w_{jk} v(\mu_j - \mu_k)}, \quad \boldsymbol{\mu} \geq 0 \quad (23.4.2)$$

Two potential functions  $v(\mu)$  are considered here and their performances compared. The first is the proposed logarithm of hyperbolic cosine proposed by Green [9] with a parameter  $\xi$  given by

$$v(\mu) = \ln \cosh(\xi\mu) = \ln \frac{e^{\xi\mu} + e^{-\xi\mu}}{2} \quad (23.4.3)$$

The second is the proposed sigmoidal function with a positive parameter  $\xi$  given by

$$v(\xi\mu) = \frac{2}{1 + e^{-\xi\mu^2}} - 1 \quad (23.4.4)$$

	0	0	0	
0	$1/\sqrt{2}$	1	$1/\sqrt{2}$	0
0	1		1	0
0	$1/\sqrt{2}$	1	$1/\sqrt{2}$	0
	0	0	0	

FIGURE 23.4.1

The parameter  $\xi = 0$  in both of these potential functions amounts to a straight EM algorithm without any prior.

Figures 23.4.2 show the sigmoid function and its derivatives for values of the adjustable parameter  $\xi = 2000, 3000, 4000, 5000, 6000$ . By varying this parameter, it is possible to smooth out speckles in the reconstructed images while retaining the sharp edges. This fact will be demonstrated analytically from the series expansion of Eq. (23.3.12).

Defining, for convenience

$$B_1 = B - \beta \frac{\partial}{\partial \mu_k^{(n)}} \ln P\{\mathbf{M}\} \quad (23.4.5)$$

Eq. (23.3.12) can be expanded in series in  $AC/B_1^2$  and written as follows:

$$\begin{aligned} \mu_k^{(n+1)} &= \frac{B_1 \pm \sqrt{B_1^2 - 4AC}}{2A} = \frac{B_1}{2A} \left[ 1 - \left( 1 - \frac{4AC}{B_1^2} \right)^{1/2} \right] \\ &= \frac{C}{B_1} \left[ 1 + \frac{AC}{B_1^2} + \frac{2A^2C^2}{B_1^4} + \frac{5A^3C^3}{B_1^6} + \frac{14A^4C^4}{B_1^8} + \frac{42A^5C^5}{B_1^{10}} + \dots \right] \end{aligned} \quad (23.4.6)$$

Real roots of Eq. (23.4.6) mandate that  $(AC/B_1^2) < \frac{1}{4}$ , and the first-order approximation for  $\mu_k^{(n+1)}$  can be given by

$$\mu_k^{(n+1)} = \frac{C}{B_1} = \frac{\sum_{i \in J_k} (N_{ik} - \hat{N}_{ik})}{\frac{1}{2} \sum_{i \in J_k} (N_{ik} + \hat{N}_{ik}) l_{ik} - \beta \frac{\partial}{\partial \mu_k} \ln P\{\mathbf{M}\}} \Bigg|_{\mu_k^{(n)}} \quad (23.4.7)$$

Taking the derivative of the logarithm of Eq. (23.4.2) with respect to  $\mu_k^{(n)}$  and substituting it in Eq. (23.4.7), we obtain

$$\mu_k^{(n+1)} = \frac{\sum_{i \in J_k} (N_{ik} - \hat{N}_{ik})}{\frac{1}{2} \sum_{i \in J_k} (N_{ik} + \hat{N}_{ik}) l_{ik} - \beta \sum_{j \in N} w_{jk} \frac{\partial}{\partial \mu_k} \left[ v(\mu_j^{(n)} - \mu_k^{(n)}) \right]} \Bigg|_{\mu_k^{(n)}} \quad (23.4.8)$$

The denominator of Eq. (23.4.8) involves the *sum* of two terms, one of which is the derivative of the potential function. This sum directly determines the penalty imposed on successive iterations in this algorithm. The relatively higher intensity of pixel  $k$  than those of its nearest neighbors is reflected by a positive value of the derivative and hence a larger value of the denominator. As a consequence, the penalty for the next iteration of  $\mu_k^{(n+1)}$  is higher, resulting in a lower estimate for the intensity of pixel  $k$ . On the other hand, the intensity of pixel  $k$  lower than those of its nearest neighbors results in a higher estimate of the intensity of pixel  $k$ . Figure 23.4.2b shows that the derivative of the sigmoidal function exhibits a band of high (0–0.05) and low (−0.05–0) values dependent on the parameter  $\xi$ . Because of this characteristic, the penalty imposed is effective only when intensity differences fall into this band. As a consequence, the unusually high differences in intensity such as sharp-edged features of the image are preserved while the speckled appearance of the image due to Poisson noise is smoothed. As shown in Fig. 23.4.2, the degree of edge separation needed can be adjusted with the parameter  $\xi$  for any particular image reconstruction.

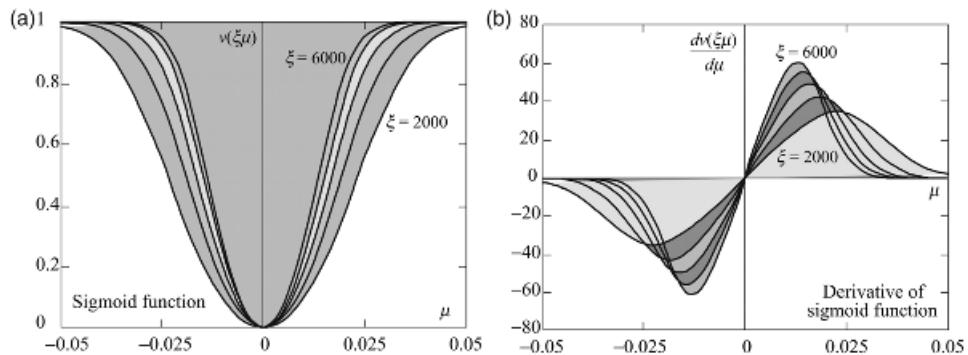


FIGURE 23.4.2

The derivative of the Incosh function, on the other hand, monotonically increases. As a consequence, there is no discrimination between edges and artifacts, resulting in uniformly smoothed edges. Hence the sigmoidal potential function should give a better reconstruction of the images.

## 23.5 COMPUTER SIMULATION

A computer simulation was performed to investigate the quality of reconstructed images using the sigmoid potential function and to compare it with Incosh potential function. Comparisons are also made with the conventional convolution backprojection (CBP) and the straight expectation–maximization algorithm. The simulation block diagram is shown in Fig. 23.5.1a. The simulation to obtain the four images for comparison is shown in Fig. 23.5.1b.

A phantom was created with parameters as shown in Fig. 23.5.2. The numerical values show the uniform attenuation coefficients in that region. This phantom is fairly complicated in detail, including a high-contrast outer shielding ring. In the usual reconstructions, especially from projections with a low number of photon counts, this shielding ring obscures the delicate features present inside it. The intention in this study was to assess how well the sigmoidal potential function handled the presence of the shielding ring. The photon source was assumed to emit a relatively low number of  $d_i = 10^4$  expected number of photons per projection path  $i$ . The phantom was divided into cells of  $64 \times 64$  pixels.

A major problem was in the implementation of the coefficient matrix  $[l_{ij}]$  with a size of  $64^2 \times 64^2 \cdot [l_{ij}]$  is a collection of lengths of the path along the projection path  $i$  inside pixel  $j$ . This is a highly sparse matrix, as each path  $i$  passes through (Fig. 23.2.1) only a small fraction of the total number of pixels in the image, and each pixel has only a small fraction of the total number of paths passing through it. Furthermore, computation for each projection path  $i$  requires access to all corresponding  $[l_{ij}]$ . Even though only a fraction of elements of this matrix is nonzero, calculation of all of them on the fly for each iteration would still require an excessive amount of computation. To reduce this computational burden, the pathlengths are calculated only once and stored in a form that is easily accessible. These stored data are defined as the *pathlength matrix*.

A semispars matrix scheme is devised to handle the problem of computational burden. Two sets of path coefficients are stored. The first set is accessible in the

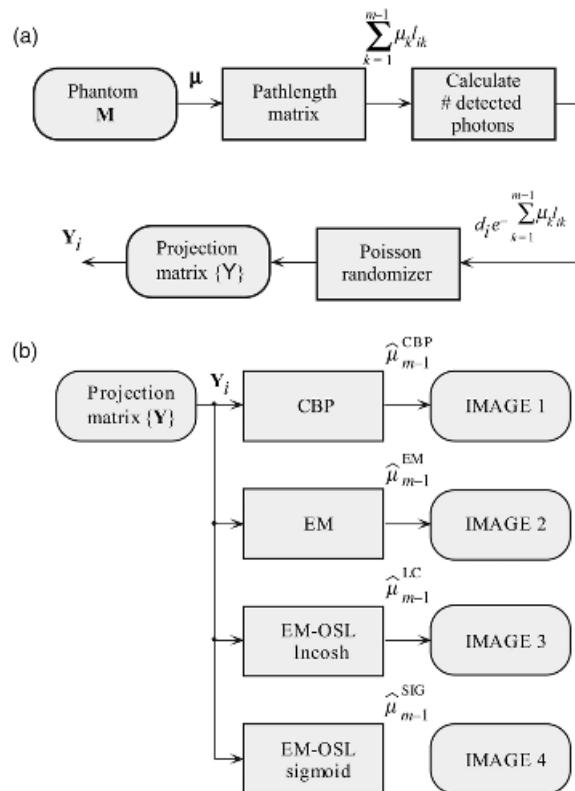


FIGURE 23.5.1

“forward” direction; that is, given the path  $i$ , all coefficients associated with it are stored. This is defined as the *forward pathlength matrix*. The second set is accessible in the “backward” direction; that is, given pixel  $j$ , all coefficients associated with it are stored. This is defined as the *backward pathlength matrix*.

The forward pathlength matrix is used to generate noise-free projection data. This matrix, in conjunction with a Poisson random-number generator routine, is used to

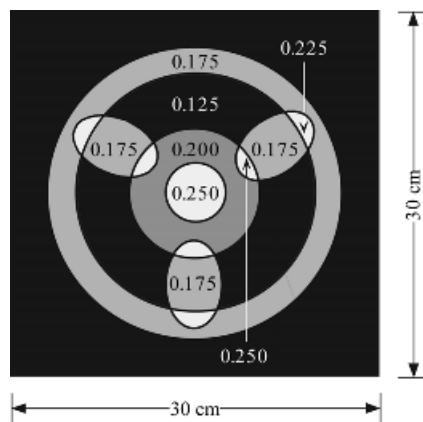


FIGURE 23.5.2

generate Poisson-randomized data. The Poisson random number is generated for a given intensity parameter and acquisition time. This random number is equal to a total number of events that have exponentially distributed interarrival times with mean equal to the given intensity parameter  $\gamma_{im} = d_i e^{-\sum_{k=1}^{m-1} \mu_k l_{ik}}$  occurring during the acquisition time. Exponential random variates are generated from a set of random variates uniformly distributed over [0,1] by probability transformation discussed in Chapter 15.

The data from the Poisson randomizer are the projection matrix  $\mathbf{Y}$ . From this projection matrix  $\mathbf{Y}$  four different types of reconstruction images are generated. The first of the reconstructed images used the conventional deterministic convolution backprojection algorithm. In the other three images using the EM algorithm, a combination of forward and backward pathlength matrices are used in reconstructing the image. Both  $N_{ik}$  and  $\hat{N}_{ik}$  defined in Eq. (23.3.3) are computed by using the forward pathlength matrix. The computed  $N_{ik}$  and  $\hat{N}_{ik}$  are then in turn utilized to determine the updated estimated attenuation coefficient  $\mu_k^{(n+1)}$  in Eqs. (23.3.7) and (23.3.11) using the backward pathlength matrix. These three images are reconstructed using (1) the conventional EM algorithm proposed by Lange [12], (2) the modified EM with Incosh potential function proposed by Green [9], and (3) the modified EM with the proposed sigmoidal potential function. Performances of these algorithms and their reconstructions are then compared. The criteria used for comparison are

- Convergence and convergence rates
- Image quality (primarily visual tempered with analytical criteria)
- Complexity (actual execution time)

Noise-free projection data were generated over 64 equally spaced projection angles between 0 and 180° using parallel beamscanner geometry. Each projection consisted of 64 equally spaced rays over the field of view of 30 cm in length. Poisson-randomized projection data were generated using the previous noise-free data as the mean value. The Poisson random-number generator algorithm as described previously is used for this purpose.

The CBP method was implemented on an Apollo Domain series 3500 system. All EM-based algorithms were implemented on the Alliant FX-2800 series system, which has both concurrent and vector processing capability. Each iteration took about 18 s. This may be slow but is not deemed a serious drawback because the speed of iteration can be considerably enhanced with multiprocessor machines. The grayscale images presented herein are windowed between 0.150 and 0.275 cm<sup>-1</sup> for better visual contrast.

## 23.6 RESULTS AND CONCLUSIONS

Figure 23.6.1 shows (a) the original phantom image, (b) reconstruction with the convolution backprojection (CBP), (c) reconstruction with the conventional EM, (d) reconstruction with EM-OSL using Incosh potential function, and (e) reconstruction with EM-OSL using sigmoidal potential function. Cross-sectional histograms of reconstructed images sectioned vertically through the center with the cross-sectional histogram of the phantom superimposed are shown in Fig. 23.6.2. In Fig. 23.6.3, the progression of RMS errors between the reconstructed images and the phantom as the iteration proceeds are shown for CBP, EM, Incosh, and sigmoid realizations.

The conventional convolution backprojection (CBP) reconstruction shown in Fig. 23.6.1b is achieved in a single step. It is the quickest in execution time because of

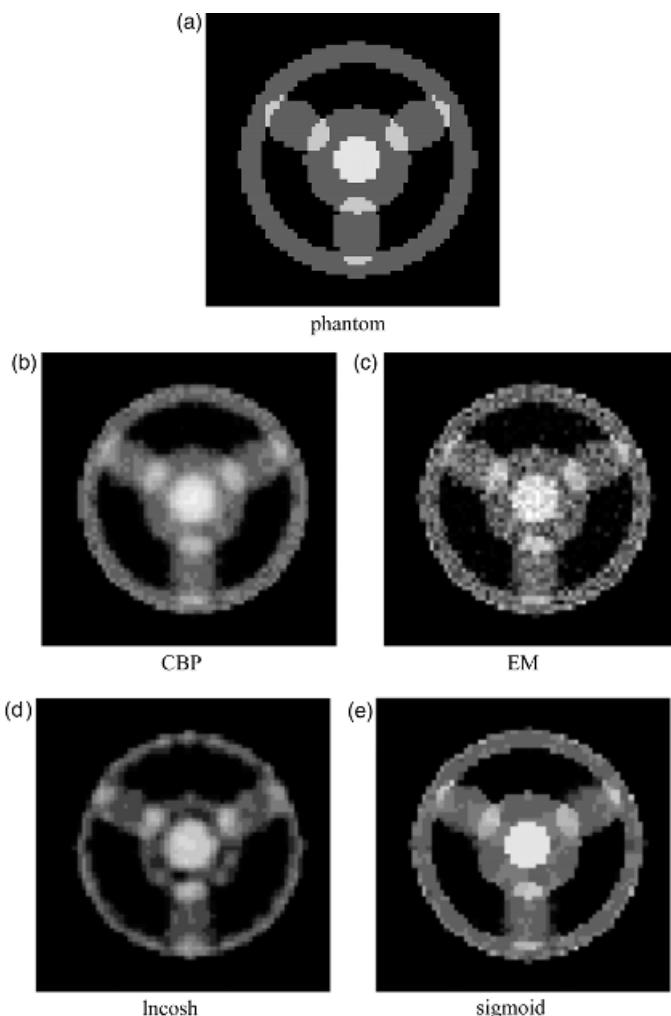


FIGURE 23.6.1

its deterministic nature. However, it is the poorest reconstruction among all the four. This poor performance is caused by factors such as intrinsic statistical variation of data, low photon counts, low angular and spatial sampling rates, and shielding ring effects. The cross-sectional histogram of Fig. 23.6.2a shows well-rounded edges, and the RMS error shown in Fig. 23.6.3 is quite high.

By using a probabilistic model, better results are obtained for the EM algorithms as shown in Figs. 23.6.1c–23.6.1e. However, the straight EM algorithm in Fig. 23.6.1c has a grainy and speckled appearance. This graininess is simply a manifestation of low correlation between parameter estimates for neighboring pixels. The cross-sectional histogram in Fig. 23.6.2b shows severely jagged edges, and the RMS error shown in Fig. 23.6.3 remains fairly constant after about 60 iterations. There is also the problem that this algorithm tends to diverge [13] after a certain number of iterations, and stopping rules are to be devised for the optimum number of iterations.

To smooth the speckles in the EM procedure, the algorithm was modified using Gibbs' prior with a Incosh potential function to penalize large differences between

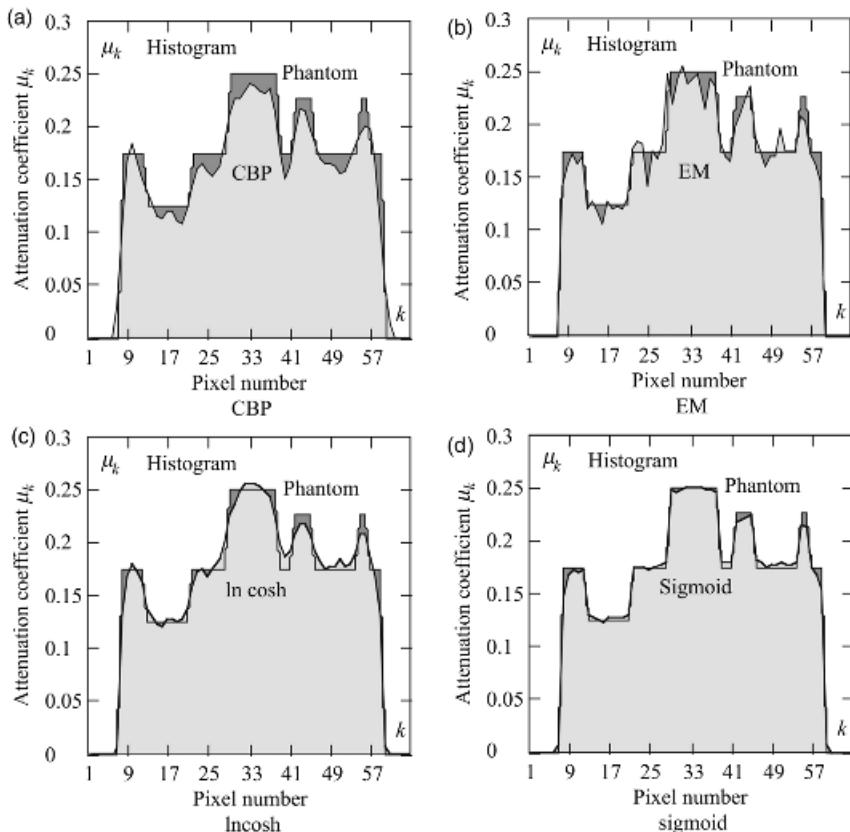


FIGURE 23.6.2

adjacent pixels. The Incosh reconstruction of Fig. 23.6.1d shows a smoother result, but the cross-sectional histogram of Fig. 23.6.2c indicates smoothed edges. The RMS error curve of Fig. 23.6.3 is also better than CBP and EM reconstructions but tends to saturate at a slightly higher error level. The difficulty stems from the fact that the Incosh potential function tends to oversmooth the image, resulting in the loss of sharp-edged features.

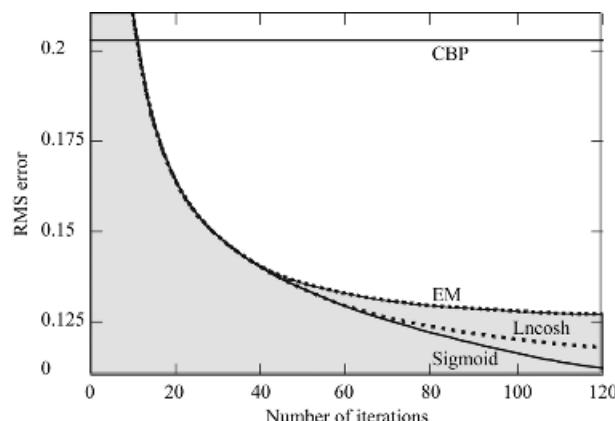


FIGURE 23.6.3

To overcome the oversmoothing difficulty, a sigmoidal potential function that preserves sharp edges while retaining the smoothing effect is proposed. Convergence curves for increasing iterations were plotted using the loglikelihood function and RMS error for different values of  $\beta$  and  $\xi$ ;  $\beta$  was varied between 0 and 0.0004, and  $\xi$  was varied between 0 and 7000. From these curves the optimum value of  $\beta$  was found to be 0.0002 and the optimum value of  $\xi$  was 5000. With these values, the reconstructed image of Fig. 23.6.1e is superior in quality to the rest. The image is visually very close to the phantom. The cross-sectional histogram of Fig. 23.6.2d also indicates a better edge performance than the Incosh potential function. The RMS error curve monotonically decreases as the iteration proceeds and at the 120th iteration is almost negligible. Thus, the modified EM algorithm with the sigmoidal potential function clearly performs better with regard to accuracy and smoothness of reconstruction.

## 23.7 DISCUSSION OF RESULTS

In comparing reconstructions in medical imaging, visual accuracy is the most important. Analytical criteria such as root-mean-square (RMS) error is secondary. The RMS error curves of Fig. 23.6.3 indicate major differences between the CBP reconstruction and EM reconstructions. The visual inspection of the reconstructed images also show discernible differences. On the other hand, the error curves for EM and Incosh reconstructions are not that far apart; but visually the differences are quite discernible. The sigmoidal reconstruction is clearly more appealing both visually and analytically.

The error curves also show that during the first 30 iterations all the EM algorithms converge almost linearly at the same rate. But after that, their convergence becomes sublinear and their graphs begin to fan out separately. Among all the four convergence rates presented, the sigmoidal algorithm again performs the best.

The speed of each iteration is dependent on the number of processors in the computer and can be considerably enhanced with more processors. The structure of this algorithm is more conducive to a multiprocessor machine.

The convergence rates and the quality of reconstruction indicate the decided superiority of this modified EM algorithm using sigmoidal potential functions.

## 23.8 REFERENCES FOR CHAPTER 23

1. Rockmore, A. J., and A. Macovski, A maximum likelihood approach to emission image reconstruction from projections, *IEEE Trans. Nucl. Sci.* **NS-23**, 1428–1432 (1976).
2. Rockmore, A. J., and A. Macovski, A maximum likelihood approach to transmission image reconstruction from projections, *IEEE Trans. Nucl. Sci.* **NS-24**, 1929–1935 (1977).
3. Dempster, A. P., N. M. Laird, and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Statist. Soc. B* **39**, 1–38 (1977).
4. Shepp, L. A., and Y. Vardi, Maximum likelihood reconstruction for emission tomography, *IEEE Trans. Med. Imag.* **MI-1**, 113–121 (1982).
5. Lange, K., and R. Carson, EM reconstruction algorithms for emission and transmission tomography, *J. Comput. Assist. Tomogr.* **8**(2), 306–316 (1984).
6. Geman, S., and D. E. McClure, Bayesian image analysis: An application to single photon emission tomography, *Proc. Am. Statist. Assoc. Statist. Comput.* 12–18 (1985).
7. Geman, S., and D. E. McClure, Statistical methods for tomographic image reconstruction, *Bull. Int. Statist. Inst.* **LII-4**, 5–21 (1987).

8. Hebert, T., and R. Leahy, A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs' priors, *IEEE Trans. Med. Imag.* **8**(2), 194–202 (1989).
9. Green, P. J., Bayesian reconstructions from emission tomography data using a modified EM algorithm, *IEEE Trans. Med. Imag.* **9**(1), 84–93 (1990).
10. Geman, S., and D. Geman, Stochastic relaxation, Gibbs' distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Machine Intell.* **PAMI-6**, 721–741 (1984).
11. Besag, J., Spatial interaction and the statistical analysis of lattice system, *J. Roy. Statist. Soc. B* **36**, 192–236 (1974).
12. Lange, K., Convergence of EM image reconstruction algorithms with Gibbs' smoothing, *IEEE Trans. Med. Imag.* **9**(4), 439–446 (1990).
13. Rajeevan, N., *A Stochastic Estimation Approach to Emission Tomography*, Ph.D. thesis, Electrical Engineering Dept., Indian Institute of Science, Bangalore, India, 1991.
14. Kamonpet, P., and V. Krishnan, *Modified EM Algorithm for Transmission Tomography*, Internal Memo T94-1, Center for Advanced computation, Univ. Massachusetts Lowell, 1994.

# A Fourier Transform Tables

The governing equations are as follows:

$$\text{Fourier transform: } X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt$$

$$\text{Inverse Fourier transform: } x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)e^{j\omega t} d\omega$$

Tables give, in the following order, the time function  $x(t)$ , the plot of the time function, the plot of the Fourier transform, and the frequency function  $X(\omega)$ .

## APPENDIX A Fourier transform tables

#	Time Function $x(t)$	$x(t)$	$ X(\omega) $	Frequency Function $X(\omega)$
1	$\frac{1}{\tau} p_{\tau/2}(t) = \left[ u\left(t + \frac{\tau}{2}\right) - u\left(t - \frac{\tau}{2}\right) \right]$			$\frac{\sin\left(\frac{\tau\omega}{2}\right)}{\left(\frac{\tau\omega}{2}\right)} = \frac{\sin(\pi\tau f)}{\pi\tau f} = \text{Sa}(\pi\tau f)$
2	$\frac{1}{\tau} \frac{\sin\left(\frac{\pi t}{\tau}\right)}{\left(\frac{\pi t}{\tau}\right)} = \frac{1}{\tau} \text{Sa}\left(\frac{\pi t}{\tau}\right)$			$p_{\pi/\tau}(\omega) = \left[ u\left(\omega + \frac{\pi}{\tau}\right) - u\left(\omega - \frac{\pi}{\tau}\right) \right]$
3	$\frac{1}{\tau} q_{\tau}(t) = \begin{cases} \frac{1}{\tau} - \frac{t}{\tau^2}, &  t  \leq \tau \\ 0, &  t  > \tau \end{cases}$			$\frac{\sin^2\left(\frac{\pi\omega}{2}\right)}{\left(\frac{\pi\omega}{2}\right)^2} = \text{Sa}^2(\pi\omega)$
4	$\frac{1}{\tau} \frac{\sin^2\left(\frac{\pi t}{\tau}\right)}{\left(\frac{\pi t}{\tau}\right)^2} = \frac{1}{\tau} \text{Sa}^2\left(\frac{\pi t}{\tau}\right)$			$q_{2\pi/\tau}(\omega) = \begin{cases} 1 - \frac{ \omega }{2\pi/\tau}, &  \omega  \leq \frac{2\pi}{\tau} \\ 0, &  \omega  > \frac{2\pi}{\tau} \end{cases}$
5	$\frac{1}{\tau\sqrt{2\pi}} e^{-(1/2)(t/\tau)^2}$			$e^{-(1/2)(\omega/\tau)^2}$
6	$\alpha e^{-\alpha t } u(t)$			$\frac{\alpha}{\alpha + j\omega}$
7	$\frac{\alpha}{2} e^{-\alpha t }$			$\frac{\alpha^2}{\alpha^2 + \omega^2}$
8	$\frac{\alpha/\pi}{1 - \alpha^2/\pi^2}$			$e^{- \omega/\alpha }$
9	$\alpha^2 t e^{-\alpha t } u(t)$			$\frac{\alpha^2}{(\alpha + j\omega)^2}$
10	$e^{-\alpha t} \sin(\beta t) u(t)$			$\frac{\beta}{(\alpha^2 + \beta^2) - \omega^2 + 2j\alpha\omega}$
11	$e^{-\alpha t} \cos(\beta t) u(t)$			$\frac{\alpha - j\omega}{(\alpha^2 + \beta^2) - \omega^2 + 2j\alpha\omega}$

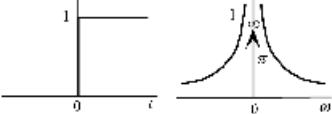
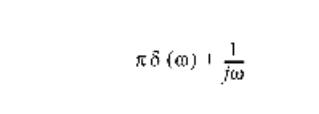
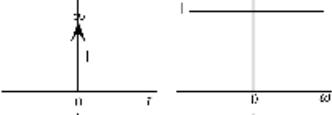
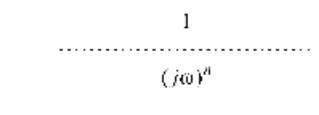
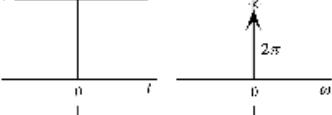
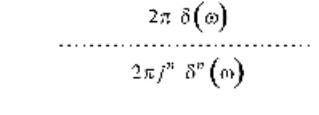
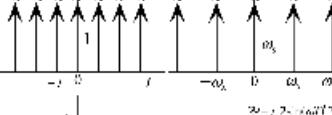
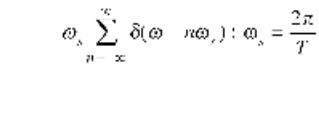
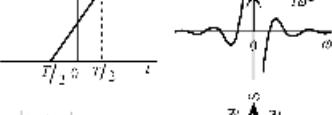
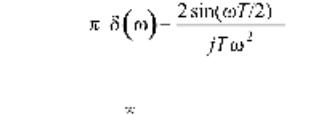
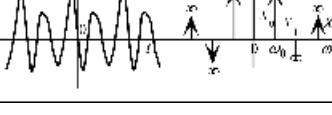
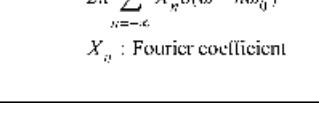
(Continued)

APPENDIX A *Continued*

#	Time Function $x(t)$	$x(t)$	$ X(\omega) $	Frequency Function $X(\omega)$
12	$\sin(\omega_0 t)$			$j\pi[\delta(\omega + \omega_0) - \delta(\omega - \omega_0)]$
13	$\cos(\omega_0 t)$			$\pi[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)]$
14	$\sin(\omega_0 t)u(t)$			$\frac{j\pi}{2}[\delta(\omega - \omega_0) - \delta(\omega + \omega_0)]$ $\frac{j\omega}{\omega_0^2 - \omega^2}$
15	$\cos(\omega_0 t)u(t)$			$\frac{\pi}{2}[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)]$ $\frac{j\omega}{\omega_0^2 - \omega^2}$
16	$\sin^2(\omega_0 t)u(t)$			$\frac{2\omega_0^2}{j\omega(4\omega_0^2 - \omega^2)}$
17	$\cos^2(\omega_0 t)u(t)$			$\frac{2\omega_0^2 - \omega^2}{j\omega(4\omega_0^2 - \omega^2)}$
18	$t \sin(\omega_0 t)u(t)$			$\frac{2j\omega\omega_0}{(\omega_0^2 - \omega^2)^2}$
19	$t \cos(\omega_0 t)u(t)$			$-\frac{(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2}$
20	$\begin{cases} \frac{2}{\pi\tau} \sqrt{1 - \left(\frac{t}{\tau}\right)^2}, &  t  \leq \tau \\ 0, &  t  > \tau \end{cases}$			$\frac{2J_1(\omega\tau)}{\omega\tau}$ : $J_1$ = Bessel function of order 1
21	$\text{sgn}(t) = \frac{t}{ t }$			$\frac{2}{j\omega}$
22	$\frac{1}{jt}$			$-\pi \text{sgn}(\omega) = -\pi \frac{\omega}{ \omega }$

(Continued)

APPENDIX A *Continued*

#	Time Function $x(t)$	$x(t)$	$ X(\omega) $	Frequency Function $X(\omega)$
23	$u(t) = \frac{1}{2} + \frac{1}{2} \operatorname{sgn}(t)$			$\pi \delta(\omega) + \frac{1}{j\omega}$
24	$\delta(t)$			$1$
25	$\delta^{(n)}(t) = \frac{d^n \delta(t)}{dt^n}$			$2\pi \delta(\omega)$
26	$\sum_{n=-\infty}^{\infty} \delta(t - nT)$			$\omega_0 \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_0) ; \omega_0 = \frac{2\pi}{T}$
27	$\frac{1}{T} \left[ \left( t - \frac{T}{2} \right) u \left( T + \frac{T}{2} \right) - \left( t - \frac{T}{2} \right) u \left( T - \frac{T}{2} \right) \right]$			$\pi \delta(\omega) - \frac{2 \sin(\omega T/2)}{jT\omega^2}$
28	$x(t) = x(t + nT) = \sum_{n=-\infty}^{\infty} X_n e^{jn\omega_0 t}$ $X_n = \frac{1}{T} \int_0^T x(t) e^{-jn\omega_0 t} dt$			$2\pi \sum_{n=-\infty}^{\infty} X_n \delta(\omega - n\omega_0)$ $X_n$ : Fourier coefficient

# Cumulative Gaussian Tables

The governing equations are as follows:

$$\text{Gaussian density: } f(x) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)x^2}$$

$$\text{Gaussian distribution function: } \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-(1/2)\xi^2} d\xi = p$$

Tables give the values of the probability  $p = \Phi(x)$  for values of  $x$  between  $-5.00$  and  $+5.00$ . Figure B.1 shows the standard Gaussian density and distribution with  $x$  along the horizontal axis (abscissa) and  $f(x)$ ,  $\Phi(x) = p$  along the vertical axis (ordinate). Representative point is shown for  $x = 0.5744$  and  $\Phi(0.5744) = 0.7$ .

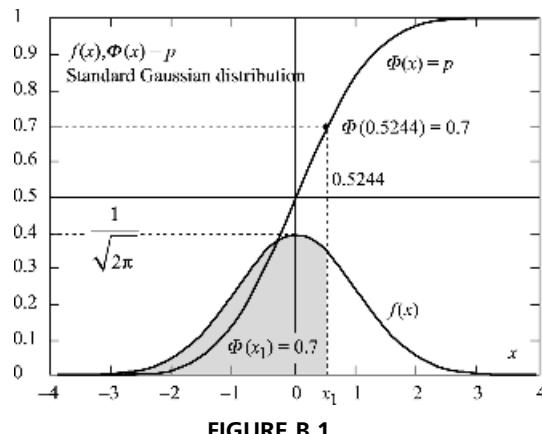


FIGURE B.1

## APPENDIX B Cumulative Gaussian tables

$x = -5.00, -2.51$									
$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$
-5	0.000000287	-4.5	0.000003398	-4	0.000031671	-3.5	0.000232629	-3	0.001349898
-4.99	0.000000302	-4.49	0.000003561	-3.99	0.000033037	-3.49	0.00024151	-2.99	0.001394887
-4.98	0.000000318	-4.48	0.000003732	-3.98	0.000034458	-3.48	0.000250707	-2.98	0.001441242
-4.97	0.000000335	-4.47	0.000003911	-3.97	0.000035936	-3.47	0.000260229	-2.97	0.001488999
-4.96	0.000000352	-4.46	0.000004098	-3.96	0.000037475	-3.46	0.000270088	-2.96	0.001538195
-4.95	0.000000371	-4.45	0.000004294	-3.95	0.000039076	-3.45	0.000280293	-2.95	0.00158887
-4.94	0.000000391	-4.44	0.000004498	-3.94	0.000040741	-3.44	0.000290857	-2.94	0.001641061
-4.93	0.000000411	-4.43	0.000004712	-3.93	0.000042473	-3.43	0.000301791	-2.93	0.00169481
-4.92	0.000000433	-4.42	0.000004935	-3.92	0.000044274	-3.42	0.000313106	-2.92	0.001750157
-4.91	0.000000455	-4.41	0.000005169	-3.91	0.000046148	-3.41	0.000324814	-2.91	0.001807144
-4.9	0.000000479	-4.4	0.000005413	-3.9	0.000048096	-3.4	0.000336929	-2.9	0.001865813
-4.89	0.000000504	-4.39	0.000005668	-3.89	0.000050122	-3.39	0.000349463	-2.89	0.001926209
-4.88	0.00000053	-4.38	0.000005934	-3.88	0.000052228	-3.38	0.000362429	-2.88	0.001988376
-4.87	0.000000558	-4.37	0.000006212	-3.87	0.000054418	-3.37	0.000375841	-2.87	0.002052359
-4.86	0.000000587	-4.36	0.000006503	-3.86	0.000056694	-3.36	0.000389712	-2.86	0.002118205
-4.85	0.000000617	-4.35	0.000006807	-3.85	0.000059059	-3.35	0.000404058	-2.85	0.002185961
-4.84	0.000000649	-4.34	0.000007124	-3.84	0.000061517	-3.34	0.000418892	-2.84	0.002255677
-4.83	0.000000683	-4.33	0.000007455	-3.83	0.000064072	-3.33	0.00043423	-2.83	0.0023274
-4.82	0.000000718	-4.32	0.000007801	-3.82	0.000066726	-3.32	0.000450087	-2.82	0.002401182
-4.81	0.000000755	-4.31	0.000008163	-3.81	0.000069483	-3.31	0.00046648	-2.81	0.002477075
-4.8	0.000000793	-4.3	0.00000854	-3.8	0.000072348	-3.3	0.000483424	-2.8	0.00255513
-4.79	0.000000834	-4.29	0.000008934	-3.79	0.000075324	-3.29	0.000500937	-2.79	0.002635402
-4.78	0.000000876	-4.28	0.000009345	-3.78	0.000078414	-3.28	0.000519035	-2.78	0.002717945
-4.77	0.000000921	-4.27	0.000009774	-3.77	0.000081624	-3.27	0.000537737	-2.77	0.002802815
-4.76	0.000000968	-4.26	0.000010221	-3.76	0.000084957	-3.26	0.000557061	-2.76	0.002890068
-4.75	0.000001017	-4.25	0.000010689	-3.75	0.000088417	-3.25	0.000577025	-2.75	0.002979763
-4.74	0.000001069	-4.24	0.000011176	-3.74	0.00009201	-3.24	0.000597648	-2.74	0.003071959
-4.73	0.000001123	-4.23	0.000011685	-3.73	0.00009574	-3.23	0.000618951	-2.73	0.003166716
-4.72	0.000001179	-4.22	0.000012215	-3.72	0.000099611	-3.22	0.000640953	-2.72	0.003264096
-4.71	0.000001239	-4.21	0.000012769	-3.71	0.00010363	-3.21	0.000663675	-2.71	0.00336416
-4.7	0.000001301	-4.2	0.000013346	-3.7	0.0001078	-3.2	0.000687138	-2.7	0.003466974
-4.69	0.000001366	-4.19	0.000013948	-3.69	0.000112127	-3.19	0.000711364	-2.69	0.003572601
-4.68	0.000001434	-4.18	0.000014575	-3.68	0.000116617	-3.18	0.000736375	-2.68	0.003681108
-4.67	0.000001506	-4.17	0.000015253	-3.67	0.000121275	-3.17	0.000762195	-2.67	0.003792562
-4.66	0.000001581	-4.16	0.000015912	-3.66	0.000126108	-3.16	0.000788846	-2.66	0.003907033
-4.65	0.00000166	-4.15	0.000016624	-3.65	0.00013112	-3.15	0.000816352	-2.65	0.004024589
-4.64	0.000001742	-4.14	0.000017365	-3.64	0.000136319	-3.14	0.000844739	-2.64	0.004145301
-4.63	0.000001828	-4.13	0.000018138	-3.63	0.000141711	-3.13	0.000874032	-2.63	0.004269243
-4.62	0.000001919	-4.12	0.000018944	-3.62	0.000147302	-3.12	0.000904255	-2.62	0.004396488
-4.61	0.000002013	-4.11	0.000019783	-3.61	0.000153099	-3.11	0.000935437	-2.61	0.004527111
-4.6	0.000002112	-4.1	0.000020658	-3.6	0.000159109	-3.1	0.000967603	-2.6	0.004661188
-4.59	0.000002216	-4.09	0.000021569	-3.59	0.000165339	-3.09	0.001000782	-2.59	0.004798797
-4.58	0.000002325	-4.08	0.000022518	-3.58	0.000171797	-3.08	0.001035003	-2.58	0.004940016
-4.57	0.000002439	-4.07	0.000023507	-3.57	0.000178491	-3.07	0.001070294	-2.57	0.005084926
-4.56	0.000002558	-4.06	0.000024536	-3.56	0.000185427	-3.06	0.001106685	-2.56	0.005233608
-4.55	0.000002682	-4.05	0.000025609	-3.55	0.000192616	-3.05	0.001144207	-2.55	0.005386146
-4.54	0.000002813	-4.04	0.000026726	-3.54	0.000200064	-3.04	0.001182891	-2.54	0.005542623
-4.53	0.000002949	-4.03	0.000027888	-3.53	0.00020778	-3.03	0.001222769	-2.53	0.005703126
-4.52	0.000003092	-4.02	0.000029099	-3.52	0.000215773	-3.02	0.001263873	-2.52	0.005867742
-4.51	0.000003241	-4.01	0.000030359	-3.51	0.000224053	-3.01	0.001306238	-2.51	0.006036558

(Continued)

APPENDIX B *Continued* $x = -2.50, -0.01$ 

$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$
-2.5	0.006209665	-2	0.022750132	-1.5	0.066807201	-1	0.158655254	-0.5	0.308537539
-2.49	0.006387155	-1.99	0.023295468	-1.49	0.068112118	-0.99	0.16108706	-0.49	0.312066949
-2.48	0.006569119	-1.98	0.023851764	-1.48	0.069436623	-0.98	0.163543059	-0.48	0.315613697
-2.47	0.006755653	-1.97	0.024419185	-1.47	0.070780877	-0.97	0.166023246	-0.47	0.319177509
-2.46	0.006946851	-1.96	0.024997895	-1.46	0.072145037	-0.96	0.168527607	-0.46	0.32275811
-2.45	0.007142811	-1.95	0.02558806	-1.45	0.07352926	-0.95	0.171056126	-0.45	0.32635522
-2.44	0.007343631	-1.94	0.026189845	-1.44	0.0749337	-0.94	0.17360878	-0.44	0.329968554
-2.43	0.007549411	-1.93	0.026803419	-1.43	0.07635851	-0.93	0.176185542	-0.43	0.333597821
-2.42	0.007760254	-1.92	0.02742895	-1.42	0.077803841	-0.92	0.17878638	-0.42	0.337242727
-2.41	0.00797626	-1.91	0.028066607	-1.41	0.079269841	-0.91	0.181411255	-0.41	0.340902974
-2.4	0.008197536	-1.9	0.02871656	-1.4	0.080756659	-0.9	0.184060125	-0.4	0.344578258
-2.39	0.008424186	-1.89	0.02937898	-1.39	0.082264439	-0.89	0.186732943	-0.39	0.348268273
-2.38	0.008656319	-1.88	0.030054039	-1.38	0.083793322	-0.88	0.189429655	-0.38	0.351972708
-2.37	0.008894043	-1.87	0.030741909	-1.37	0.085343451	-0.87	0.192150202	-0.37	0.355691245
-2.36	0.009137468	-1.86	0.031442763	-1.36	0.086914962	-0.86	0.194894521	-0.36	0.359423567
-2.35	0.009386706	-1.85	0.032156775	-1.35	0.088507991	-0.85	0.197662543	-0.35	0.363169349
-2.34	0.00964187	-1.84	0.032884119	-1.34	0.090122672	-0.84	0.200454193	-0.34	0.366928264
-2.33	0.009903076	-1.83	0.033624969	-1.33	0.091759136	-0.83	0.203269392	-0.33	0.370699981
-2.32	0.010170439	-1.82	0.034379502	-1.32	0.093417509	-0.82	0.206108054	-0.32	0.374484165
-2.31	0.010444077	-1.81	0.035147894	-1.31	0.095097918	-0.81	0.208970088	-0.31	0.378280478
-2.3	0.01072411	-1.8	0.035930319	-1.3	0.096800485	-0.8	0.211855399	-0.3	0.382088578
-2.29	0.011010658	-1.79	0.036726956	-1.29	0.098525329	-0.79	0.214763884	-0.29	0.385908119
-2.28	0.011303844	-1.78	0.03753798	-1.28	0.100272568	-0.78	0.217695438	-0.28	0.389738752
-2.27	0.011603792	-1.77	0.03836357	-1.27	0.102042315	-0.77	0.220649946	-0.27	0.393580127
-2.26	0.011910625	-1.76	0.039203903	-1.26	0.103834681	-0.76	0.223627292	-0.26	0.397431887
-2.25	0.012224473	-1.75	0.040059157	-1.25	0.105649774	-0.75	0.226627352	-0.25	0.401293674
-2.24	0.012545461	-1.74	0.040929509	-1.24	0.107487697	-0.74	0.229649997	-0.24	0.405165128
-2.23	0.012873721	-1.73	0.041815138	-1.23	0.109348552	-0.73	0.232695092	-0.23	0.409045885
-2.22	0.013209384	-1.72	0.042716221	-1.22	0.111232437	-0.72	0.235762498	-0.22	0.412935577
-2.21	0.013552581	-1.71	0.043632937	-1.21	0.1131319446	-0.71	0.238852068	-0.21	0.416833837
-2.2	0.013903448	-1.7	0.044565463	-1.2	0.11506967	-0.7	0.241963652	-0.2	0.420740291
-2.19	0.014262118	-1.69	0.045513977	-1.19	0.117023196	-0.69	0.245097094	-0.19	0.424654565
-2.18	0.014628731	-1.68	0.046478658	-1.18	0.119000107	-0.68	0.24825223	-0.18	0.428576284
-2.17	0.015003423	-1.67	0.047459682	-1.17	0.121000484	-0.67	0.251428895	-0.17	0.432505068
-2.16	0.015386335	-1.66	0.048457226	-1.16	0.123024403	-0.66	0.254626915	-0.16	0.436440537
-2.15	0.015777607	-1.65	0.049471468	-1.15	0.125071936	-0.65	0.257846111	-0.15	0.440382308
-2.14	0.016177383	-1.64	0.050502583	-1.14	0.127143151	-0.64	0.2610863	-0.14	0.444329995
-2.13	0.016585807	-1.63	0.051550748	-1.13	0.129238112	-0.63	0.264347292	-0.13	0.448283213
-2.12	0.017003023	-1.62	0.052616138	-1.12	0.131356881	-0.62	0.267628893	-0.12	0.452241574
-2.11	0.017429178	-1.61	0.053698928	-1.11	0.133499513	-0.61	0.270930904	-0.11	0.456204687
-2.1	0.017864421	-1.6	0.054799292	-1.1	0.135666061	-0.6	0.274253118	-0.1	0.460172163
-2.09	0.0183089	-1.59	0.055917403	-1.09	0.137856572	-0.59	0.277595325	-0.09	0.464143607
-2.08	0.018762766	-1.58	0.057053433	-1.08	0.140071109	-0.58	0.280957309	-0.08	0.468118628
-2.07	0.019226172	-1.57	0.058207556	-1.07	0.142309654	-0.57	0.284338849	-0.07	0.47209683
-2.06	0.01969927	-1.56	0.059379941	-1.06	0.1445723	-0.56	0.287739719	-0.06	0.476077817
-2.05	0.020182215	-1.55	0.060570758	-1.05	0.146859056	-0.55	0.291159687	-0.05	0.480061194
-2.04	0.020675163	-1.54	0.061780177	-1.04	0.14916995	-0.54	0.294598516	-0.04	0.484046563
-2.03	0.02117827	-1.53	0.063008364	-1.03	0.151505003	-0.53	0.298055965	-0.03	0.488033527
-2.02	0.021691694	-1.52	0.064255488	-1.02	0.15386423	-0.52	0.301531788	-0.02	0.492021686
-2.01	0.022215594	-1.51	0.065521712	-1.01	0.156247645	-0.51	0.305025731	-0.01	0.496010644

(Continued)

APPENDIX B *Continued*

$x = 0, 2.49$									
$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$
0	0.5	0.5	0.691462461	1	0.841344746	1.5	0.933192799	2	0.977249868
0.01	0.503989356	0.51	0.694974269	1.01	0.843752355	1.51	0.934478288	2.01	0.977784406
0.02	0.507978314	0.52	0.698468212	1.02	0.84613577	1.52	0.935744512	2.02	0.978308306
0.03	0.511966473	0.53	0.701944035	1.03	0.848494997	1.53	0.936991636	2.03	0.97882173
0.04	0.515953437	0.54	0.705401484	1.04	0.85083005	1.54	0.938219823	2.04	0.979324837
0.05	0.519938806	0.55	0.708840313	1.05	0.853140944	1.55	0.939429242	2.05	0.979817785
0.06	0.523922183	0.56	0.712260281	1.06	0.8554277	1.56	0.940620059	2.06	0.98030073
0.07	0.52790317	0.57	0.715661151	1.07	0.857690346	1.57	0.941792444	2.07	0.980773828
0.08	0.531881372	0.58	0.719042691	1.08	0.85992891	1.58	0.942946567	2.08	0.981237234
0.09	0.535856393	0.59	0.722404675	1.09	0.862143428	1.59	0.944082597	2.09	0.9816911
0.1	0.539827837	0.6	0.725746882	1.1	0.864333939	1.6	0.945200708	2.1	0.982135579
0.11	0.543795313	0.61	0.729069096	1.11	0.866500487	1.61	0.946301072	2.11	0.982570822
0.12	0.547758426	0.62	0.732371107	1.12	0.868643119	1.62	0.947383862	2.12	0.982996977
0.13	0.551716787	0.63	0.735652708	1.13	0.870761888	1.63	0.948449252	2.13	0.983414193
0.14	0.555670005	0.64	0.7389137	1.14	0.872856849	1.64	0.949497417	2.14	0.983822617
0.15	0.559617692	0.65	0.742153889	1.15	0.874928064	1.65	0.950528532	2.15	0.984222393
0.16	0.563559463	0.66	0.745373085	1.16	0.876975597	1.66	0.951542774	2.16	0.984613665
0.17	0.567494932	0.67	0.748571105	1.17	0.878999516	1.67	0.952540318	2.17	0.984996577
0.18	0.571423716	0.68	0.75174777	1.18	0.880999893	1.68	0.953521342	2.18	0.985371269
0.19	0.575345435	0.69	0.754902906	1.19	0.882976804	1.69	0.954486023	2.19	0.985737882
0.2	0.579259709	0.7	0.758036348	1.2	0.88493033	1.7	0.955434537	2.2	0.986096552
0.21	0.583166163	0.71	0.761147932	1.21	0.886860554	1.71	0.956367063	2.21	0.986447419
0.22	0.587064423	0.72	0.764237502	1.22	0.888767563	1.72	0.957283779	2.22	0.986790616
0.23	0.590954115	0.73	0.767304908	1.23	0.890651448	1.73	0.958184862	2.23	0.987126279
0.24	0.594834872	0.74	0.770350003	1.24	0.892512303	1.74	0.959070491	2.24	0.987454539
0.25	0.598706326	0.75	0.773372648	1.25	0.894350226	1.75	0.959940843	2.25	0.987775527
0.26	0.602568113	0.76	0.776372708	1.26	0.896165319	1.76	0.960796097	2.26	0.988089375
0.27	0.606419873	0.77	0.779350054	1.27	0.897957685	1.77	0.96163643	2.27	0.988396208
0.28	0.610261248	0.78	0.782304562	1.28	0.899727432	1.78	0.96246202	2.28	0.988696156
0.29	0.614091881	0.79	0.785236116	1.29	0.901474671	1.79	0.963273044	2.29	0.988989342
0.3	0.617911422	0.8	0.788144601	1.3	0.903199515	1.8	0.964069681	2.3	0.98927589
0.31	0.621719522	0.81	0.791029912	1.31	0.904902082	1.81	0.964852106	2.31	0.989555923
0.32	0.625515835	0.82	0.793891946	1.32	0.906582491	1.82	0.965620498	2.32	0.989829561
0.33	0.629300019	0.83	0.796730608	1.33	0.908240864	1.83	0.966375031	2.33	0.990096924
0.34	0.633071736	0.84	0.799545807	1.34	0.909877328	1.84	0.967115881	2.34	0.99035813
0.35	0.636830651	0.85	0.802337457	1.35	0.911492009	1.85	0.967843225	2.35	0.990613294
0.36	0.640576433	0.86	0.805105479	1.36	0.913085038	1.86	0.968557237	2.36	0.990862532
0.37	0.644308755	0.87	0.807849798	1.37	0.914656549	1.87	0.969258091	2.37	0.991105957
0.38	0.648027292	0.88	0.810570345	1.38	0.916206678	1.88	0.969945961	2.38	0.991343681
0.39	0.651731727	0.89	0.813267057	1.39	0.917735561	1.89	0.97062102	2.39	0.991575814
0.4	0.655421742	0.9	0.815939875	1.4	0.919243341	1.9	0.97128344	2.4	0.991802464
0.41	0.659097026	0.91	0.818588745	1.41	0.920730159	1.91	0.971933393	2.41	0.99202374
0.42	0.662757273	0.92	0.82121362	1.42	0.922196159	1.92	0.97257105	2.42	0.992239746
0.43	0.666402179	0.93	0.823814458	1.43	0.92364149	1.93	0.973196581	2.43	0.992450589
0.44	0.670031446	0.94	0.82639122	1.44	0.9250663	1.94	0.973810155	2.44	0.992656369
0.45	0.67364478	0.95	0.828943874	1.45	0.92647074	1.95	0.97441194	2.45	0.992857189
0.46	0.67724189	0.96	0.831472393	1.46	0.927854963	1.96	0.975002105	2.46	0.993053149
0.47	0.680822491	0.97	0.833976754	1.47	0.929219123	1.97	0.975580815	2.47	0.993244347
0.48	0.684386303	0.98	0.836456941	1.48	0.930563377	1.98	0.976148236	2.48	0.993430881
0.49	0.687933051	0.99	0.83891294	1.49	0.931887882	1.99	0.976704532	2.49	0.993612845

(Continued)

APPENDIX B *Continued*

$x = 2.50, 4.99$									
$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$
2.5	0.993790335	3	0.998650102	3.5	0.999767371	4	0.999968329	4.5	0.999996602
2.51	0.993963442	3.01	0.998693762	3.51	0.999775947	4.01	0.999969641	4.51	0.999996759
2.52	0.994132258	3.02	0.998736127	3.52	0.999784227	4.02	0.999970901	4.52	0.999996908
2.53	0.994296874	3.03	0.998777231	3.53	0.99979222	4.03	0.999972112	4.53	0.999997051
2.54	0.994457377	3.04	0.998817109	3.54	0.999799936	4.04	0.999973274	4.54	0.999997187
2.55	0.994613854	3.05	0.998855793	3.55	0.999807384	4.05	0.999974391	4.55	0.999997318
2.56	0.994766392	3.06	0.998893315	3.56	0.999814573	4.06	0.999975464	4.56	0.999997442
2.57	0.994915074	3.07	0.998929706	3.57	0.999821509	4.07	0.999976493	4.57	0.999997561
2.58	0.995059984	3.08	0.998964997	3.58	0.999828203	4.08	0.999977482	4.58	0.999997675
2.59	0.995201203	3.09	0.998999218	3.59	0.999834661	4.09	0.999978431	4.59	0.999997784
2.6	0.995338812	3.1	0.999032397	3.6	0.999840891	4.1	0.999979342	4.6	0.999997888
2.61	0.995472889	3.11	0.999064563	3.61	0.999846901	4.11	0.999980217	4.61	0.999997987
2.62	0.995603512	3.12	0.999095745	3.62	0.999852698	4.12	0.999981056	4.62	0.999998081
2.63	0.995730757	3.13	0.999125968	3.63	0.999858289	4.13	0.999981862	4.63	0.999998172
2.64	0.995854699	3.14	0.999155261	3.64	0.999863681	4.14	0.999982635	4.64	0.999998258
2.65	0.995975411	3.15	0.999183648	3.65	0.999868888	4.15	0.999983376	4.65	0.99999834
2.66	0.996092967	3.16	0.999211154	3.66	0.999873892	4.16	0.999984088	4.66	0.999998419
2.67	0.996207438	3.17	0.999237805	3.67	0.999878725	4.17	0.99998477	4.67	0.999998494
2.68	0.996318892	3.18	0.999263625	3.68	0.999883383	4.18	0.999985425	4.68	0.999998566
2.69	0.996427399	3.19	0.999288636	3.69	0.999887873	4.19	0.999986052	4.69	0.999998634
2.7	0.996533026	3.2	0.999312862	3.7	0.9998922	4.2	0.999986654	4.7	0.999998699
2.71	0.99663584	3.21	0.999336325	3.71	0.99989637	4.21	0.999987231	4.71	0.999998761
2.72	0.996735904	3.22	0.999359047	3.72	0.999900389	4.22	0.999987785	4.72	0.999998821
2.73	0.996833284	3.23	0.999381049	3.73	0.99990426	4.23	0.999988315	4.73	0.999998877
2.74	0.996928041	3.24	0.999402352	3.74	0.99990799	4.24	0.999988824	4.74	0.999998931
2.75	0.997020237	3.25	0.999422975	3.75	0.999911583	4.25	0.999989311	4.75	0.999998983
2.76	0.997109932	3.26	0.999442939	3.76	0.999915043	4.26	0.999989779	4.76	0.999999032
2.77	0.997197185	3.27	0.999462263	3.77	0.999918376	4.27	0.999990226	4.77	0.999999079
2.78	0.997282055	3.28	0.999480965	3.78	0.999921586	4.28	0.999990655	4.78	0.999999124
2.79	0.997364598	3.29	0.999499063	3.79	0.999924676	4.29	0.999991066	4.79	0.999999166
2.8	0.997444847	3.3	0.999516576	3.8	0.999927652	4.3	0.99999146	4.8	0.999999207
2.81	0.997522925	3.31	0.99953352	3.81	0.999930517	4.31	0.999991837	4.81	0.999999245
2.82	0.997598818	3.32	0.999549913	3.82	0.999933274	4.32	0.999992199	4.82	0.999999282
2.83	0.9976726	3.33	0.99956577	3.83	0.999935928	4.33	0.999992545	4.83	0.999999317
2.84	0.997744323	3.34	0.999581108	3.84	0.999938483	4.34	0.999992876	4.84	0.999999351
2.85	0.997814039	3.35	0.999595942	3.85	0.999940941	4.35	0.999993193	4.85	0.999999383
2.86	0.997881795	3.36	0.999610288	3.86	0.999943306	4.36	0.999993497	4.86	0.999999413
2.87	0.997947641	3.37	0.999624159	3.87	0.999945582	4.37	0.999993788	4.87	0.999999442
2.88	0.998011624	3.38	0.999637571	3.88	0.999947772	4.38	0.999994066	4.88	0.999999447
2.89	0.998073791	3.39	0.999650537	3.89	0.999949878	4.39	0.999994332	4.89	0.999999496
2.9	0.998134187	3.4	0.999663071	3.9	0.999951904	4.4	0.999994587	4.9	0.999999521
2.91	0.998192856	3.41	0.999675186	3.91	0.999953852	4.41	0.999994831	4.91	0.999999545
2.92	0.998249843	3.42	0.999686894	3.92	0.999955726	4.42	0.999995065	4.92	0.999999567
2.93	0.99830519	3.43	0.999698209	3.93	0.999957527	4.43	0.999995288	4.93	0.999999589
2.94	0.998358939	3.44	0.999709143	3.94	0.999959259	4.44	0.999995502	4.94	0.999999609
2.95	0.99841113	3.45	0.999719707	3.95	0.999960924	4.45	0.999995706	4.95	0.999999629
2.96	0.998461805	3.46	0.999729912	3.96	0.999962525	4.46	0.999995902	4.96	0.999999648
2.97	0.998511001	3.47	0.999739771	3.97	0.999964064	4.47	0.999996089	4.97	0.999999665
2.98	0.998558758	3.48	0.999749293	3.98	0.999965542	4.48	0.999996268	4.98	0.999999682
2.99	0.998605113	3.49	0.99975849	3.99	0.999966963	4.49	0.999996439	4.99	0.999999698

# Inverse Cumulative Gaussian Tables

The governing equations are as follows:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-(1/2)\xi^2} d\xi = p \quad \text{and} \quad x = \Phi^{-1}(p)$$

Tables give the values of  $x$  in  $x = \Phi^{-1}(p)$  for values of  $p$  between (0.01, 0.99) grouped in such a manner that adjacent column values of  $p$  add to 1. Figure C.1 shows  $f(x)$ ,  $p$  along the horizontal axis and  $x$ ,  $\Phi^{-1}(p)$  along the vertical axis. Since  $f(x)$  is an even function,  $\Phi^{-1}(p)$  has the symmetry  $\Phi^{-1}(p) = -\Phi^{-1}(1-p)$ . Representative points are shown for  $p = 0.1$  and  $0.9$  with  $\Phi^{-1}(0.1) = -1.2816$  and  $\Phi^{-1}(0.9) = 1.2816$ .

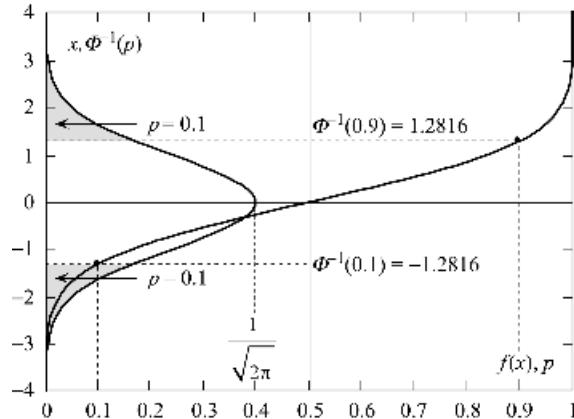


FIGURE C.1

## APPENDIX C Inverse cumulative Gaussian tables

$p = 0.01, 0.99$			
$p$	$x$	$p$	$x$
0.01	-2.326347874	0.99	2.326347874
0.02	-2.053748911	0.98	2.053748911
0.03	-1.880793608	0.97	1.880793608
0.04	-1.750686071	0.96	1.750686071
0.05	-1.644853627	0.95	1.644853627
0.06	-1.554773595	0.94	1.554773595
0.07	-1.475791028	0.93	1.475791028
0.08	-1.40507156	0.92	1.40507156
0.09	-1.340755034	0.91	1.340755034
0.1	-1.281551566	0.9	1.281551566
0.11	-1.22652812	0.89	1.22652812
0.12	-1.174986792	0.88	1.174986792
0.13	-1.126391129	0.87	1.126391129
0.14	-1.080319341	0.86	1.080319341
0.15	-1.036433389	0.85	1.036433389
0.16	-0.994457883	0.84	0.994457883
0.17	-0.954165253	0.83	0.954165253
0.18	-0.915365088	0.82	0.915365088
0.19	-0.877896295	0.81	0.877896295
0.2	-0.841621234	0.8	0.841621234
0.21	-0.806421247	0.79	0.806421247
0.22	-0.772193214	0.78	0.772193214
0.23	-0.738846849	0.77	0.738846849
0.24	-0.706302563	0.76	0.706302563
0.25	-0.67448975	0.75	0.67448975
0.26	-0.643345405	0.74	0.643345405
0.27	-0.612812991	0.73	0.612812991
0.28	-0.582841507	0.72	0.582841507
0.29	-0.55338472	0.71	0.55338472
0.3	-0.524400513	0.7	0.524400513
0.31	-0.495850347	0.69	0.495850347
0.32	-0.467698799	0.68	0.467698799
0.33	-0.439913166	0.67	0.439913166
0.34	-0.412463129	0.66	0.412463129
0.35	-0.385320466	0.65	0.385320466
0.36	-0.358458793	0.64	0.358458793
0.37	-0.331853346	0.63	0.331853346
0.38	-0.305480788	0.62	0.305480788
0.39	-0.279319034	0.61	0.279319034
0.4	-0.253347103	0.6	0.253347103
0.41	-0.227544977	0.59	0.227544977
0.42	-0.201893479	0.58	0.201893479
0.43	-0.176374165	0.57	0.176374165
0.44	-0.150969215	0.56	0.150969215
0.45	-0.125661347	0.55	0.125661347
0.46	-0.100433721	0.54	0.100433721
0.47	-0.075269862	0.53	0.075269862
0.48	-0.050153583	0.52	0.050153583
0.49	-0.025068908	0.51	0.025068908
0.5	0	0.5	0

# Inverse Chi-Square Tables

The governing equations are as follows:

$$\text{Chi-square density: } f(x) = \frac{x^{(v/2)-1} e^{-(x/2)}}{2^{v/2} \Gamma\left(\frac{v}{2}\right)} u(x)$$

$$\text{Chi-square distribution function: } F(x) = \int_0^x \frac{\xi^{(v/2)-1} e^{-(\xi/2)}}{2^{v/2} \Gamma\left(\frac{v}{2}\right)} d\xi = p$$

Tables give the values of  $x$  in  $x = F^{-1}(p)$  for values of  $p$  between (0.005, 0.995) grouped in such a manner that adjacent column values of  $p$  add to 1. Figure D.1 shows  $f(x)$ ,  $p$  along the horizontal axis and  $x$ ,  $F^{-1}(p)$  along the vertical axis. Representative points are shown for number of degrees of freedom  $v = 4$ , and  $p = 0.1$  and  $0.9$  with  $F^{-1}(0.1) = 1.06362$  and  $F^{-1}(0.9) = 7.7794$ . The asymmetry of the distributions is to be noted and  $F^{-1}(p) \neq -F^{-1}(1-p)$ .

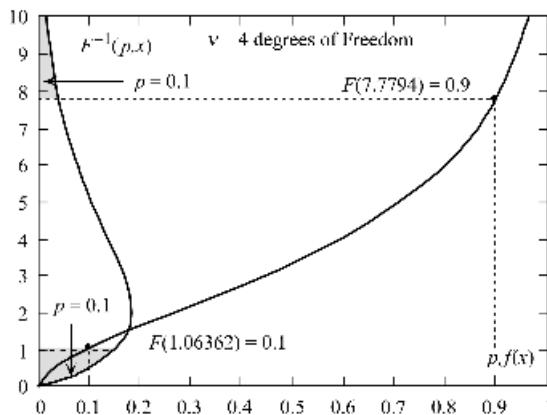


FIGURE D.1

## APPENDIX D Inverse chi-square tables

$\nu = 1-50; p = (0.005, 0.995), (0.01, 0.99), (0.025, 0.975)$							
$\nu$	$p = 0.005$	$p = 0.995$	$p = 0.01$	$p = 0.99$	$p = 0.025$	$p = 0.975$	$\nu$
1	0.00003927	7.879438577	0.000157088	6.634896601	0.000982069	5.023886187	1
2	0.010025084	10.596634733	0.020100672	9.210340372	0.050635616	7.377758908	2
3	0.071721775	12.838156467	0.114831802	11.34486673	0.215795283	9.348403604	3
4	0.206989093	14.860259001	0.297109481	13.276704136	0.484418557	11.143286782	4
5	0.411741904	16.749602344	0.554298077	15.086272469	0.831211613	12.832501994	5
6	0.675726777	18.547584179	0.87209033	16.81189383	1.237344246	14.449375335	6
7	0.989255683	20.277739875	1.239042306	18.475306907	1.689869181	16.012764275	7
8	1.344413087	21.954954991	1.646497373	20.09023503	2.179730747	17.534546139	8
9	1.734932905	23.589350781	2.087900736	21.665994333	2.7003895	19.022767799	9
10	2.155856481	25.188179572	2.55821216	23.209184564	3.24697278	20.483057648	10
11	2.603221891	26.756848916	3.053484107	24.724924519	3.815748252	21.919961571	11
12	3.073823638	28.299518822	3.570568971	26.216934848	4.403788507	23.336598106	12
13	3.56503458	29.819471224	4.106915472	27.688225996	5.008750512	24.735553943	13
14	4.074674957	31.319349623	4.660425063	29.141220162	5.628726103	26.118907954	14
15	4.600915572	32.801320646	5.229348884	30.577900808	6.262137795	27.488360751	15
16	5.142205443	34.267186538	5.81221247	31.999916564	6.907664353	28.8453246	16
17	5.697217101	35.718465659	6.407759778	33.408655452	7.56418645	30.190987574	17
18	6.264804685	37.156451457	7.014910901	34.8052992	8.230746195	31.526360444	18
19	6.843971445	38.582256555	7.632729648	36.190863807	8.906516482	32.852311662	19
20	7.433844263	39.996846313	8.260398333	37.566230383	9.590777392	34.169593932	20
21	8.03365342	41.401064771	8.897197942	38.932168983	10.282897783	35.478864731	21
22	8.642716401	42.795654999	9.542492339	40.289357281	10.98230734	36.780702372	22
23	9.260424776	44.18127525	10.195715556	41.638395386	11.688551922	38.075618741	23
24	9.886233502	45.558511937	10.856361476	42.979817739	12.401150217	39.364069513	24
25	10.519652112	46.92789016	11.523975372	44.314102758	13.119720025	40.646462438	25
26	11.160237406	48.289882332	12.198146924	45.641680735	13.843904982	41.923164113	26
27	11.807587351	49.644915299	12.878504393	46.962940359	14.573382731	43.194505574	27
28	12.461335948	50.993376268	13.564709755	48.278234136	15.307860553	44.460786948	28
29	13.121148888	52.335617786	14.256454576	49.587882943	16.047071695	45.722281346	29
30	13.78671986	53.67196193	14.953456528	50.892179866	16.790772266	46.979238156	30
31	14.457767386	55.00270388	15.655456402	52.191393454	17.538738581	48.231885827	31
32	15.134032105	56.32811496	16.362215548	53.48577051	18.290764907	49.480434253	32
33	15.815274424	57.648445256	17.073513672	54.775538476	19.046661503	50.725076818	33
34	16.501272476	58.963925876	17.789146924	56.060907496	19.806252939	51.965992158	34
35	17.191820342	60.274770905	18.508926227	57.342072207	20.569376631	53.20334569	35
36	17.886726503	61.581179115	19.232675832	58.619213294	21.335881561	54.437290942	36
37	18.585812465	62.883335454	19.960232036	59.892498851	22.105627161	55.667970718	37
38	19.288911559	64.181412357	20.691442062	61.16208558	22.878482329	56.895518116	38
39	19.995867875	65.475570903	21.426163065	62.428119839	23.654324558	58.120057429	39
40	20.706535317	66.765961833	22.164261253	63.690738578	24.433039171	59.341704937	40
41	21.420776761	68.052726455	22.905611106	64.950070163	25.214518638	60.560569618	41
42	22.138463303	69.335997457	23.650094678	66.206235111	25.998661968	61.776753768	42
43	22.859473591	70.615899618	24.397600972	67.459346747	26.785374166	62.990353565	43
44	23.583693213	71.892550459	25.148025383	68.70951179	27.574565744	64.201459567	44
45	24.311014165	73.166060818	25.901269193	69.956830881	28.366152292	65.410157164	45
46	25.041334352	74.436535372	26.65723912	71.201399058	29.160054074	66.61652698	46
47	25.77455715	75.704073105	27.415846908	72.443306179	29.956195682	67.820645235	47
48	26.510591006	76.968767732	28.177008953	73.682637316	30.754505709	69.022584082	48
49	27.24934907	78.230708087	28.940645973	74.919473096	31.554916463	70.222411896	49
50	27.990748866	79.489978467	29.706684365	76.153890028	32.357364362	71.42019355	50

(Continued)

APPENDIX D *Continued*

$v = 1-50; p = (0.05, 0.95), (0.1, 0.9), (0.2, 0.8)$							
v	$p = 0.05$	$p = 0.95$	$p = 0.1$	$p = 0.9$	$p = 0.2$	$p = 0.8$	v
1	0.00393214	3.841458821	0.015790774	2.705543454	0.064184755	1.642374415	1
2	0.102586589	5.991464547	0.210721031	4.605170186	0.446287103	3.218875825	2
3	0.351846318	7.814727903	0.584374374	6.251388631	1.005174013	4.641627676	3
4	0.710723021	9.487729037	1.063623217	7.77944034	1.648776618	5.988616694	4
5	1.145476226	11.070497694	1.610307987	9.2363569	2.342534306	7.289276127	5
6	1.635382894	12.591587244	2.204130656	10.644640676	3.070088405	8.55805972	6
7	2.167349909	14.067140449	2.833106918	12.017036624	3.822321908	9.8032499	7
8	2.732636793	15.507313056	3.489539126	13.361566137	4.593573612	11.03009143	8
9	3.325112843	16.918977605	4.168159008	14.683656573	5.380053212	12.24214547	9
10	3.940299136	18.307038053	4.865182052	15.987179172	6.179079256	13.441957575	10
11	4.574813079	19.675059322	5.57778479	17.275008518	6.988673512	14.631420509	11
12	5.226029488	21.026010276	6.30379606	18.549347787	7.807327679	15.811986222	12
13	5.891864338	22.361986155	7.04150458	19.811895377	8.633860835	16.984797018	13
14	6.570631384	23.684754534	7.78953361	21.064117144	9.467327987	18.150770562	14
15	7.260943928	24.995760467	8.546756242	22.307107631	10.306959007	19.310657111	15
16	7.961645572	26.296203303	9.312236354	23.541810867	11.152116471	20.465079294	16
17	8.671760205	27.587091472	10.085186335	24.7690203	12.002265725	21.614560534	17
18	9.390455081	28.869282498	10.864936117	25.989410406	12.856953096	22.759537649	18
19	10.117013064	30.143512836	11.650910032	27.203560236	13.715789706	23.900410245	19
20	10.850811394	31.41042053	12.44260921	28.411971308	14.578439217	25.037499634	20
21	11.591305209	32.670562694	13.239597975	29.615081395	15.444608404	26.171094726	21
22	12.338014579	33.92442919	14.041493189	30.813275318	16.314039795	27.30144947	22
23	13.090514188	35.172453474	14.847955799	32.006893497	17.186505854	28.428788503	23
24	13.848425027	36.415021289	15.658684053	33.196238807	18.061804323	29.553311674	24
25	14.611407639	37.65247771	16.473407999	34.381582128	18.939754458	30.67519771	25
26	15.379156583	38.885132903	17.29188499	35.563166883	19.820193955	31.794607209	26
27	16.15139585	40.11326688	18.113895967	36.741212787	20.702976421	32.911685118	27
28	16.927875044	41.337133446	18.939242372	37.91591895	21.587969273	34.026562783	28
29	17.708366183	42.556963516	19.767743559	39.087466492	22.475051978	35.139359671	29
30	18.492660982	43.772967897	20.599234615	40.256020733	23.364114574	36.250184822	30
31	19.280568559	44.985339664	21.4335645	41.421733062	24.255056418	37.35913808	31
32	20.071913465	46.194256176	22.270594477	42.584742523	25.147785116	38.466311141	32
33	20.866533991	47.399880813	23.110196744	43.745177181	26.042215603	39.571788454	33
34	21.664280713	48.60236447	23.952253271	44.9031553	26.93826936	40.675648	34
35	22.465015221	49.801846855	24.796654784	46.05878636	27.835873722	41.777961963	35
36	23.268609019	50.998457616	25.64329988	47.212171943	28.734961295	42.87877318	36
37	24.074942557	52.192317324	26.492094258	48.363406511	29.635469424	43.978216337	37
38	24.883904383	53.383538345	27.342950042	49.512578084	30.537339749	45.076277046	38
39	25.6953904	54.572225595	28.195785182	50.659768839	31.440517798	46.173033613	39
40	26.509303197	55.758477217	29.050522931	51.805055638	32.344952636	47.2685367	40
41	27.32555147	56.942385176	29.907091372	52.948510498	33.250596552	48.362833776	41
42	28.144049497	58.124035791	30.76542301	54.090201009	34.157404779	49.455969385	42
43	28.96471667	59.303510211	31.625454395	55.230190704	35.065335167	50.547985397	43
44	29.787477081	60.480884832	32.487125793	56.368539392	35.974348309	51.638921227	44
45	30.612259146	61.656231685	33.350380889	57.505303458	36.884406771	52.72881403	45
46	31.438995267	62.829618774	34.215166515	58.640536131	37.795475328	53.81769888	46
47	32.26762153	64.001110383	35.081432522	59.774287724	38.70752069	54.905608928	47
48	33.098077429	65.170767358	35.94913116	60.906605854	39.620511358	55.992575548	48
49	33.930305926	66.338647357	36.818217433	62.037535643	40.534417489	57.078628466	49
50	34.764252015	67.50480508	37.688648582	63.167119892	41.449210777	58.163795881	50

(Continued)

APPENDIX D *Continued*

$v = 1-50; p = (0.3, 0.7), (0.4, 0.6), 0.5$						
$v$	$p = 0.3$	$p = 0.7$	$p = 0.4$	$p = 0.6$	$p = 0.5$	$v$
1	0.148471862	1.074194171	0.274995898	0.708326301	0.454936423	1
2	0.713349888	2.407945609	1.021651248	1.832581464	1.386294361	2
3	1.423652243	3.664870783	1.8691168403	2.9461166073	2.365973884	3
4	2.194698421	4.878432967	2.752842684	4.044626491	3.35669398	4
5	2.999908133	6.064429984	3.655499623	5.131867074	4.351460191	5
6	3.827551588	7.231135332	4.570153808	6.210757195	5.348120627	6
7	4.671330449	8.383430829	5.49323486	7.283207633	6.345811196	7
8	5.527422085	9.524458193	6.42264556	8.350525468	7.344121498	8
9	6.393305964	10.656372007	7.357034502	9.413640094	8.342832692	9
10	7.267218166	11.780722627	8.295471761	10.473236231	9.341817766	10
11	8.147867778	12.898668202	9.237285424	11.529833841	10.340998074	11
12	9.034276588	14.011100168	10.181971379	12.583837967	11.340322377	12
13	9.925682415	15.11872165	11.12913994	13.635570994	12.339755883	13
14	10.821477722	16.222098613	12.078482479	14.685294256	13.339274149	14
15	11.721168973	17.321694498	13.029749599	15.733222952	14.338859511	15
16	12.624348764	18.417894392	13.982736339	16.77953671	15.338498885	16
17	13.53067614	19.511022353	14.937271803	17.824387263	16.338182377	17
18	14.439862342	20.601354114	15.893211722	18.867904121	17.337902369	18
19	15.351660263	21.689126583	16.850432972	19.910198856	18.337652897	19
20	16.265856485	22.774545074	17.808829473	20.951368378	19.337429229	20
21	17.182265184	23.857788896	18.76830905	21.991497491	20.337227564	21
22	18.100723373	24.939012326	19.728791007	23.030660899	21.337044808	22
23	19.021087158	26.018362133	20.690204216	24.068924809	22.336878423	23
24	19.943228742	27.095958619	21.6524856	25.106348219	23.336726306	24
25	20.867034014	28.171912889	22.615578908	26.142983969	24.336586698	25
26	21.792400577	29.246324805	23.579433726	27.178877884	25.336458117	26
27	22.719236133	30.319284516	24.544004658	28.214076541	26.336339309	27
28	23.647457138	31.390873728	25.509250654	29.248616986	27.336229199	28
29	24.576987684	32.46116677	26.47513445	30.282534702	28.336126867	29
30	25.507758554	33.530231496	27.441622092	31.315862101	29.336030553	30
31	26.439706425	34.598130043	28.408682542	32.348628886	30.33594159	31
32	27.372773189	35.664919486	29.376287336	33.380862353	31.335858304	32
33	28.306905369	36.730652396	30.344410296	34.412587658	32.335780169	33
34	29.242053617	37.795377322	31.313026722	35.443828044	33.335706722	34
35	30.178172286	38.859139207	32.282115467	36.474605037	34.335637554	35
36	31.115219045	39.921979754	33.25165522	37.504938619	35.335572301	36
37	32.053154559	40.989377744	34.221626767	38.534847379	36.33551064	37
38	32.991941893	42.045049313	35.192012219	39.564348645	37.335452283	38
39	33.93154751	43.105348196	36.162794893	40.593458597	38.335396972	39
40	34.871939106	44.16486595	37.133959198	41.622192378	39.335344473	40
41	35.813086755	45.223632142	38.105490539	42.650564178	40.335294578	41
42	36.754962385	46.281674518	39.077375226	43.678587316	41.335247098	42
43	37.697539616	47.339019162	40.049600404	44.706274319	42.335201862	43
44	38.640793628	48.395690628	41.022153978	45.733636979	43.335158714	44
45	39.58470103	49.451712062	41.995024553	46.760686418	44.335117514	45
46	40.529239748	50.507105318	42.96820138	47.787433134	45.335078131	46
47	41.474388925	51.561891048	43.941674302	48.813887054	46.33504045	47
48	42.420128828	52.616088801	44.915433716	49.840057576	47.335004361	48
49	43.366440764	53.669717099	45.889470523	50.865953603	48.334969765	49
50	44.313307007	54.722793511	46.863776099	51.891583585	49.334936573	50

(Continued)

APPENDIX D *Continued*

$\nu = 51-100; p = (0.005, 0.095), (0.01, 0.99), (0.025, 0.975)$							
$\nu$	$p = 0.005$	$p = 0.995$	$p = 0.01$	$p = 0.99$	$p = 0.025$	$p = 0.975$	$\nu$
51	28.734711993	80.746658954	30.475049216	77.385960787	33.161787043	72.615990662	51
52	29.481163844	82.000825703	31.245674187	78.615754476	33.968127113	73.809861815	52
53	30.230033356	83.252551205	32.01849447	79.843336874	34.776329306	75.001862766	53
54	30.981252782	84.501904533	32.793448296	81.068770649	35.586340961	76.192046632	54
55	31.734757475	85.748951559	33.570476733	82.292115562	36.39811185	77.380464062	55
56	32.490485693	86.993755161	34.349523512	83.513428655	37.211594026	78.567163394	56
57	33.248378424	88.23637541	35.13053486	84.732764419	38.026741678	79.752190799	57
58	34.008379217	89.476869741	35.913459346	85.950174949	38.843511005	80.93559042	58
59	34.770434027	90.715293114	36.698247743	87.165710093	39.66186009	82.117404486	59
60	35.5344933	91.951697435	37.4848529	88.379417585	40.48174879	83.297673434	60
61	36.300502901	93.186134541	38.273229616	89.591343164	41.303138632	84.47643601	61
62	37.068417486	94.418652128	39.063334534	90.801530695	42.125992722	85.653729366	62
63	37.838191332	95.649296617	39.855126035	92.010022268	42.950275647	86.829589151	63
64	38.609780517	96.878112584	40.648564143	93.216858304	43.775953402	88.004049597	64
65	39.383142817	98.105142865	41.443610435	94.422077642	44.602993308	89.177143594	65
66	40.15823762	99.33042865	42.240227958	95.625717624	45.431363941	90.348902761	66
67	40.935025839	100.554009565	43.038381154	96.827814178	46.261035068	91.519357519	67
68	41.713469833	101.775923762	43.838035781	98.028401888	47.091977581	92.688537149	68
69	42.493533335	102.996207988	44.639158855	99.227514066	47.924163443	93.856469852	69
70	43.27518138	104.214897658	45.441718577	100.425182814	48.757565633	95.023182807	70
71	44.058380242	105.432026925	46.245684281	101.621439089	49.592158091	96.188702216	71
72	44.843097376	106.647628736	47.051026376	102.816312756	50.427915678	97.353053358	72
73	45.629301357	107.861734898	47.857716293	104.009832639	51.264814124	98.516260633	73
74	46.416961827	109.074376125	48.665726437	105.202026577	52.102829993	99.6783476	74
75	47.206049448	110.285582098	49.475030143	106.392921468	52.941940637	100.839337024	75
76	47.996535854	111.495381505	50.285601629	107.58254331	53.782124164	101.999250906	76
77	48.788393603	112.703802091	51.09741596	108.770917246	54.623359403	103.158110523	77
78	49.58159614	113.910870703	51.910449004	109.958067602	55.465625868	104.315936459	78
79	50.376117755	115.116613326	52.724677404	111.144017925	56.30890373	105.472748638	79
80	51.171933545	116.321055121	53.540078535	112.328791013	57.153173787	106.62856635	80
81	51.969019383	117.524220465	54.356630479	113.512408954	57.998417438	107.783408283	81
82	52.76735188	118.726132984	55.174311991	114.694893153	58.844616655	108.937292543	82
83	53.566908355	119.926815581	55.993102472	115.87626436	59.691753959	110.090236688	83
84	54.367666807	121.126290472	56.812981941	117.056542701	60.539812398	111.242257743	84
85	55.169605886	122.324579209	57.633931009	118.235747703	61.388775523	112.393372224	85
86	55.972704867	123.521702713	58.455930856	119.413898318	62.238627371	113.543596164	86
87	56.776943622	124.717681296	59.278963207	120.591012945	63.089352441	114.692945124	87
88	57.5823026	125.912534686	60.10301031	121.767109456	63.940935677	115.841434222	88
89	58.3887628	127.106282051	60.928054917	122.942205213	64.793362453	116.989078142	89
90	59.196305755	128.298942021	61.754080261	124.116317093	65.646618551	118.135891156	90
91	60.004913504	129.490532709	62.581070042	125.2894615	66.50069015	119.281887138	91
92	60.814568582	130.68107173	63.409008403	126.46165439	67.355563806	120.427079583	92
93	61.625253991	131.87057622	64.23787992	127.632911283	68.211226444	121.571481613	93
94	62.43695319	133.059062856	65.067669579	128.803247283	69.067665336	122.715106003	94
95	63.249650077	134.246547869	65.898362767	129.972677093	69.924868096	123.857965182	95
96	64.063328968	135.433047066	66.729945253	131.141215025	70.782822662	125.000071256	96
97	64.877974588	136.618575842	67.562403175	132.308875022	71.641517287	126.141436011	97
98	65.693572054	137.803149192	68.39572303	133.475670665	72.500940526	127.282070931	98
99	66.51010686	138.986781733	69.229891655	134.64161519	73.361081226	128.421987207	99
100	67.327564863	140.169487711	70.06489622	135.806721497	74.221928516	129.561195745	100

(Continued)

APPENDIX D *Continued*

$v = 51-100; p = (0.05, 0.95), (0.1, 0.9), (0.2, 0.8)$							
$v$	$p = 0.05$	$p = 0.95$	$p = 0.1$	$p = 0.9$	$p = 0.2$	$p = 0.8$	$v$
51	35.599864291	68.669292475	38.560384001	64.295399247	42.364864348	59.248104569	51
52	36.437093611	69.832158932	39.433385087	65.422412349	43.281352656	60.331579986	52
53	37.275893194	70.993451453	40.307615103	66.54819597	44.198651397	61.414246354	53
54	38.116218476	72.15321481	41.183039056	67.672785133	45.11673742	62.496126741	54
55	38.958026959	73.311491694	42.059623582	68.796213232	46.035588661	63.577243145	55
56	39.80127808	74.468322845	42.937336842	69.918512133	46.955184064	64.657616553	56
57	40.645933095	75.623747173	43.816148427	71.03971227	47.875503522	65.737267009	57
58	41.491954958	76.777801876	44.69602927	72.159842735	48.796527819	66.816213674	58
59	42.339308228	77.93052254	45.576951561	73.278931356	49.718238573	67.894474875	59
60	43.187958967	79.081943236	46.45888678	74.397004778	50.640618187	68.972068158	60
61	44.037874654	80.232096609	47.341815113	75.514088528	51.563649805	70.049010333	61
62	44.889024106	81.38101396	48.225706407	76.630207082	52.487317265	71.125317519	62
63	45.741377397	82.528725323	49.110539094	77.745383921	53.411605064	72.201005177	63
64	46.594905793	83.675259533	49.996290646	78.859641592	54.336498318	73.276088154	64
65	47.449581685	84.820644296	50.882939414	79.973001754	55.261982729	74.350580711	65
66	48.305378528	85.964906247	51.770464592	81.08548523	56.188044553	75.424496558	66
67	49.162270784	87.108071007	52.65884616	82.197112051	57.11467057	76.497848883	67
68	50.02023387	88.25016324	53.548064854	83.307901494	58.041848059	77.57065038	68
69	50.879244111	89.391206695	54.438102118	84.417872126	58.969564766	78.642913271	69
70	51.739278687	90.531224262	55.328940071	85.527041837	59.897808889	79.714649334	70
71	52.600315599	91.670238007	56.220561478	86.635427875	60.826569047	80.785869924	71
72	53.462333623	92.808269218	57.112949708	87.743046878	61.755834264	81.856585994	72
73	54.325312271	93.945338439	58.006088714	88.849914902	62.685593946	82.926808115	73
74	55.189231761	95.081465509	58.899963	89.956047452	63.615837867	83.996546493	74
75	56.05407298	96.216669595	59.794557594	91.061459508	64.546556146	85.065810988	75
76	56.919817453	97.350969222	60.689858029	92.166165545	65.477739236	86.134611128	76
77	57.786447313	98.484382304	61.585850311	93.270179563	66.409377904	87.20295613	77
78	58.653945277	99.61692617	62.482520907	94.373515102	67.341463219	88.270854905	78
79	59.522294613	100.748617592	63.379856716	95.47618527	68.27398654	89.33831608	79
80	60.391479123	101.879472812	64.277845055	96.578202754	69.206939498	90.405348008	80
81	61.261483114	103.009507559	65.17647364	97.679579848	70.14031399	91.471958778	81
82	62.13229138	104.138737076	66.075730564	98.78032846	71.074102163	92.538156228	82
83	63.003889178	105.267176143	66.975604289	99.880460136	72.008296406	93.603947959	83
84	63.87626221	106.394839088	67.876083622	100.979986072	72.9422889336	94.669341337	84
85	64.749396606	107.521739815	68.777157707	102.078917131	73.877873794	95.734343511	85
86	65.623278902	108.647891817	69.678816008	103.177263852	74.813242831	96.79896142	86
87	66.497896029	109.773308192	70.581048294	104.27503647	75.748989702	97.863201796	87
88	67.373235291	110.898001663	71.483844633	105.372244921	76.685107857	98.927071179	88
89	68.249284356	112.021984589	72.387195374	106.468898862	77.621590933	99.990575925	89
90	69.126031238	113.14526898	73.291091138	107.565007673	78.558432745	101.053722205	90
91	70.003464285	114.267866513	74.195522809	108.660580476	79.495627284	102.116516023	91
92	70.881572165	115.389788542	75.100481519	109.755626139	80.433168703	103.178963214	92
93	71.760343854	116.511046112	76.005958646	110.850153291	81.371051319	104.241069456	93
94	72.639768627	117.631649971	76.911945798	111.944170325	82.309269599	105.302840272	94
95	73.519836042	118.75161058	77.818434808	113.037685412	83.24781816	106.364281037	95
96	74.400535934	119.870938123	78.725417723	114.130706506	84.18669176	107.425396987	96
97	75.281858403	120.989642518	79.632886801	115.223241354	85.125885295	108.486193218	97
98	76.163793804	122.107733429	80.540834496	116.315297503	86.065393794	109.546674696	98
99	77.046332738	123.22522027	81.44925346	117.406882305	87.005212413	110.606846258	99
100	77.929466046	124.342112218	82.358136527	118.498002927	87.945336428	111.666712621	100

(Continued)

APPENDIX D *Continued*

$\nu = 51-100; p = (0.3, 0.7), (0.4, 0.6), 0.5$						
$\nu$	$p = 0.3$	$p = 0.7$	$p = 0.4$	$p = 0.6$	$p = 0.5$	$\nu$
51	45.260710728	55.775334723	47.838342257	52.916955546	50.334904701	51
52	46.208635932	56.827356596	48.813161218	53.942077115	51.33487407	52
53	47.157067403	57.878874222	49.788225583	54.966955553	52.334844611	53
54	48.105990651	58.92990198	50.763528307	55.991597778	53.334816257	54
55	49.055391864	59.980453574	51.739062677	57.016010384	54.334788947	55
56	50.005257867	61.030542085	52.714822287	58.040199666	55.334762624	56
57	50.955576076	62.080180003	53.690801024	59.064171637	56.334737237	57
58	51.906334467	63.129379269	54.666993045	60.087932043	57.334712735	58
59	52.85752154	64.178151304	55.643392763	61.111486384	58.334689074	59
60	53.809126286	65.226507044	56.61999483	62.134839925	59.334666211	60
61	54.761138161	66.274456964	57.596794126	63.157997709	60.334644106	61
62	55.713547056	67.322011111	58.57378574	64.180964574	61.334622722	62
63	56.666343275	68.369179117	59.550964966	65.203745161	62.334602024	63
64	57.619517508	69.415970239	60.528327283	66.226343926	63.334581981	64
65	58.573060815	70.462393365	61.50586835	67.24876515	64.334562561	65
66	59.526964601	71.508457038	62.483583997	68.27101295	65.334543735	66
67	60.481220601	72.554169478	63.461470211	69.293091285	66.334525478	67
68	61.43582086	73.599538596	64.439523131	70.315003967	67.334507763	68
69	62.390757719	74.64457201	65.417739042	71.336754666	68.334490567	69
70	63.346023798	75.689277062	66.396114361	72.358346919	69.334473867	70
71	64.301611984	76.733660827	67.374645639	73.379784137	70.334457642	71
72	65.257515415	77.77730133	68.35329545	74.40106961	71.334441872	72
73	66.213727471	78.821491568	69.332162869	75.422206512	72.334426538	73
74	67.17024176	79.864951494	70.31114251	76.443197909	73.334411623	74
75	68.127052106	80.908116056	71.290265472	77.464046764	74.334397109	75
76	69.084152542	81.950991193	72.269528863	78.48475594	75.334382981	76
77	70.0415373	82.993582648	73.248929883	79.505328204	76.334369223	77
78	70.999200799	84.035895978	74.228465828	80.525766236	77.334355821	78
79	71.957137638	85.077936559	75.208134078	81.546072627	78.334342762	79
80	72.91534259	86.119709596	76.187932098	82.566249886	79.334330032	80
81	73.873810591	87.161220133	77.167857435	83.586300446	80.334317619	81
82	74.832536734	88.202473054	78.147907708	84.606226661	81.334305511	82
83	75.791516265	89.243473097	79.128080614	85.626030816	82.334293697	83
84	76.750744571	90.284224854	80.108373916	86.645715125	83.334282167	84
85	77.710217182	91.32473278	81.088785446	87.665281736	84.334270911	85
86	78.669929754	92.365001199	82.069313102	88.684732735	85.334259919	86
87	79.629878076	93.40503431	83.04995484	89.704070146	86.334249181	87
88	80.590058056	94.444836188	84.030708678	90.723295935	87.33423869	88
89	81.550465719	95.484410794	85.01157269	91.742412011	88.334228436	89
90	82.511097202	96.523761975	85.992545004	92.761420231	89.334218411	90
91	83.471948753	97.562893474	86.973623801	93.780322399	90.334208609	91
92	84.43301672	98.601808927	87.954807313	94.79912027	91.334199022	92
93	85.394297553	99.640511874	88.936093818	95.81781555	92.334189642	93
94	86.355787797	100.679005757	89.917481644	96.8364099	93.334180463	94
95	87.31748409	101.717293927	90.898969161	97.854904937	94.334171479	95
96	88.27938316	102.755379646	91.880554784	98.873302234	95.334162684	96
97	89.241481819	103.793266092	92.862236969	99.891603324	96.334154071	97
98	90.203776963	104.83095636	93.844014212	100.9098097	97.334145635	98
99	91.166265566	105.868453465	94.825885047	101.927922816	98.334137371	99
100	92.12894468	106.905760346	95.807848047	102.945944092	99.334129274	100

# Inverse Student-*t* Tables

The governing equations are as follows:

$$\text{Student-}t \text{ density: } f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{t^2}{v}\right)^{-[(v+1)/2]}$$

$$\text{Student-}t \text{ distribution function: } F(t) = \int_{-\infty}^t \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{\tau^2}{v}\right)^{-[(v+1)/2]} d\tau$$

Since  $f(x)$  is an even function,  $F^{-1}(p)$  has the symmetry  $F^{-1}(p) = -F^{-1}(-p)$ . The absolute values of  $t$  for  $p$  between (0.005, 0.45) are the same as the values for  $p$  between (0.55, 0.999) with  $t = F^{-1}(0.5) = 0$  for all  $v$ . Hence the tables give the values of  $t = F^{-1}(p)$  for  $p$  only between (0.55, 0.999). Figure E.1 shows  $f(t)$ ,  $p$  along the horizontal axis and  $t$ ,  $F^{-1}(p)$  along the vertical axis. Representative points are shown for number of degrees of freedom  $v = 5$ , and  $p = 0.1$  and  $0.9$  with  $F^{-1}(0.1) = -1.4759$  and  $F^{-1}(0.9) = 1.4759$ .

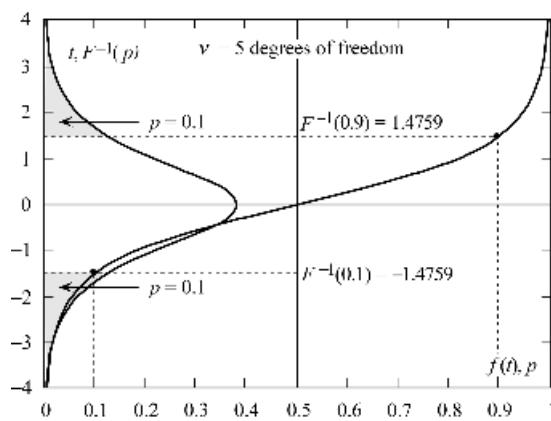


FIGURE E.1

## APPENDIX E Inverse student-t tables

v = 1–50; p = 0.9–0.999								
v	p = 0.9	p = 0.925	p = 0.95	p = 0.975	p = 0.99	p = 0.995	p = 0.999	v
1	3.077683537	4.16529977	6.313751515	12.706204736	31.820515954	63.656741163	318.308838985	1
2	1.885618083	2.281930588	2.91998558	4.30265273	6.964556734	9.924843201	22.32712477	2
3	1.637744354	1.924319657	2.353363435	3.182446305	4.540702859	5.84090931	10.214531852	3
4	1.533206274	1.778192164	2.131846786	2.776445105	3.746947388	4.604094871	7.17318222	4
5	1.475884049	1.699362566	2.015048373	2.570581836	3.364929999	4.032142984	5.893429531	5
6	1.439755747	1.650173154	1.943180281	2.446911851	3.142668403	3.707428021	5.207626239	6
7	1.414923928	1.616591737	1.894578605	2.364624252	2.997951567	3.499483297	4.785289629	7
8	1.396815131	1.59222144	1.859548038	2.306004135	2.896459448	3.355387331	4.500790934	8
9	1.383028738	1.573735785	1.833112933	2.262157163	2.821437925	3.249835542	4.296805663	9
10	1.372183641	1.559235933	1.812461123	2.228138852	2.763769458	3.169272673	4.143700494	10
11	1.363430318	1.547559766	1.795884819	2.20098516	2.718079184	3.105806516	4.024701038	11
12	1.356217334	1.537956495	1.782287556	2.17881283	2.680997993	3.054539589	3.929633265	12
13	1.350171289	1.529919606	1.770933396	2.160368656	2.650308838	3.012275839	3.851982391	13
14	1.345030374	1.523095061	1.761310136	2.144786688	2.624494068	2.976842734	3.787390238	14
15	1.340605608	1.517227969	1.753050356	2.131449546	2.602480295	2.946712883	3.732834425	15
16	1.336757167	1.51213017	1.745883676	2.119905299	2.583487185	2.920781622	3.686154793	16
17	1.33337939	1.507659754	1.739606726	2.109815578	2.566933984	2.89823052	3.64576738	17
18	1.330390944	1.503707672	1.734063607	2.10092204	2.55237963	2.878440473	3.610484885	18
19	1.327728209	1.500188756	1.729132812	2.093024054	2.539483191	2.860934606	3.579400149	19
20	1.325340707	1.497035518	1.724718243	2.085963447	2.527977003	2.84533971	3.551808343	20
21	1.323187874	1.494193795	1.720742903	2.079613845	2.517648016	2.831359558	3.527153669	21
22	1.321236742	1.491619612	1.717144374	2.073873068	2.508324553	2.818756061	3.504992031	22
23	1.31946024	1.489276897	1.713871528	2.06865761	2.499866739	2.807335684	3.484964375	23
24	1.317835934	1.487135783	1.71088208	2.063898562	2.492159473	2.796939505	3.466777298	24
25	1.316345073	1.485171326	1.708140761	2.059538553	2.485107175	2.787435814	3.450188727	25
26	1.314971864	1.483362535	1.70561792	2.055529439	2.478629824	2.778714533	3.434997182	26
27	1.313702913	1.481691617	1.703288446	2.051830516	2.472659912	2.770682957	3.421033621	27
28	1.312526782	1.48014339	1.701130934	2.048407142	2.467140098	2.763262455	3.408155178	28
29	1.311433647	1.478704821	1.699127027	2.045229642	2.46202136	2.756385904	3.396240288	29
30	1.310415025	1.477364662	1.697260887	2.042272456	2.457261542	2.749995654	3.385184867	30
31	1.309463549	1.47611315	1.695518783	2.039513446	2.452824193	2.744041919	3.37489928	31
32	1.308572793	1.474941772	1.693888748	2.036933343	2.448677634	2.738481482	3.365305926	32
33	1.307737124	1.473843072	1.692360309	2.034515297	2.4447942	2.733276642	3.356337279	33
34	1.306951587	1.47281049	1.690924255	2.032244509	2.441149628	2.728394367	3.347934313	34
35	1.306211802	1.471838233	1.689572458	2.030107928	2.437722547	2.723805589	3.340045202	35
36	1.305513886	1.470921166	1.688297714	2.028094001	2.434494061	2.71948463	3.332624257	36
37	1.304854381	1.470054719	1.68709362	2.026192463	2.4314474	2.715408722	3.325631045	37
38	1.304230204	1.469234815	1.68595446	2.024394164	2.428567631	2.711557602	3.319029655	38
39	1.303638589	1.468457801	1.684785122	2.02269092	2.42584141	2.707913184	3.312788083	39
40	1.303077053	1.467720399	1.683851013	2.02107539	2.423256779	2.704459267	3.306877714	40
41	1.302543359	1.467019655	1.682878002	2.01954097	2.420802992	2.701181304	3.301272889	41
42	1.302035487	1.466352901	1.681952357	2.018081703	2.41847036	2.698066186	3.295950529	42
43	1.301551608	1.465717725	1.681070703	2.016692199	2.416250129	2.695102079	3.290889821	43
44	1.30109006	1.465111933	1.680229977	2.015367574	2.414134368	2.692278266	3.286071946	44
45	1.300649332	1.464533534	1.679427393	2.014103389	2.412115876	2.689585019	3.281479848	45
46	1.300228048	1.463980712	1.678660414	2.012895599	2.410188096	2.687013492	3.277098029	46
47	1.299824947	1.463451805	1.677926722	2.011740514	2.40834505	2.684555618	3.272912378	47
48	1.299438879	1.462945295	1.677224196	2.010634758	2.406581273	2.682204027	3.268910018	48
49	1.299068785	1.46245979	1.676550893	2.009575237	2.40489176	2.679951974	3.265079173	49
50	1.298713694	1.461994009	1.675905025	2.008559112	2.403271917	2.677793271	3.261409056	50

(Continued)

APPENDIX E *Continued*

$v = 1-50; p = 0.9-0.999$								
$v$	$p = 0.9$	$p = 0.925$	$p = 0.95$	$p = 0.975$	$p = 0.99$	$p = 0.995$	$p = 0.999$	$v$
1	3.077683537	4.16529977	6.313751515	12.706204736	31.820515954	63.656741163	318.308838985	1
2	1.885618083	2.281930588	2.91998558	4.30265273	6.964556734	9.924843201	22.32712477	2
3	1.637744354	1.924319657	2.353363435	3.182446305	4.540702859	5.84090931	10.214531852	3
4	1.533206274	1.778192164	2.131846786	2.776445105	3.746947388	4.604094871	7.17318222	4
5	1.475884049	1.699362566	2.015048373	2.570581836	3.364929999	4.032142984	5.893429531	5
6	1.439755747	1.650173154	1.943180281	2.446911851	3.142668403	3.707428021	5.207626239	6
7	1.414923928	1.616591737	1.894578605	2.364624252	2.997951567	3.499483297	4.785289629	7
8	1.39681531	1.59222144	1.859548038	2.306004135	2.896459448	3.355387331	4.500790934	8
9	1.383028738	1.573735785	1.833112933	2.262157163	2.821437925	3.249835542	4.296805663	9
10	1.372183641	1.559235933	1.812461123	2.228138852	2.763769458	3.169272673	4.143700494	10
11	1.363430318	1.547559766	1.795884819	2.20098516	2.718079184	3.105806516	4.024701038	11
12	1.356217334	1.537956495	1.782287556	2.17881283	2.680997993	3.054539589	3.929633265	12
13	1.350171289	1.529919606	1.770933396	2.160368656	2.650308838	3.012275839	3.851982391	13
14	1.345030374	1.523095061	1.761310136	2.144786688	2.624494068	2.976842734	3.787390238	14
15	1.340605608	1.517227969	1.753050356	2.131449546	2.602480295	2.946712883	3.732834425	15
16	1.336757167	1.51213017	1.745883676	2.119905299	2.583487185	2.920781622	3.686154793	16
17	1.33337939	1.507659754	1.739606726	2.109815578	2.566933984	2.89823052	3.64576738	17
18	1.330390944	1.503707672	1.734063607	2.10092204	2.55237963	2.878440473	3.610484885	18
19	1.327728209	1.500188756	1.729132812	2.093024054	2.539483191	2.860934606	3.579400149	19
20	1.325340707	1.497035518	1.724718243	2.085963447	2.527977003	2.84533971	3.551808343	20
21	1.323187874	1.494193795	1.720742903	2.079613845	2.517648016	2.831359558	3.527153669	21
22	1.321236742	1.491619612	1.717144374	2.073873068	2.508324553	2.818756061	3.504992031	22
23	1.31946024	1.489276897	1.713871528	2.06865761	2.499866739	2.807335684	3.484964375	23
24	1.317835934	1.487135783	1.71088208	2.063898562	2.492159473	2.796939505	3.466777298	24
25	1.316345073	1.485171326	1.708140761	2.059538553	2.485107175	2.787435814	3.450188727	25
26	1.314971864	1.483362535	1.705617192	2.055529439	2.478629824	2.778714533	3.434997182	26
27	1.313702913	1.481691617	1.703288446	2.051830516	2.472659912	2.770682957	3.421033621	27
28	1.312526782	1.48014339	1.701130934	2.048407142	2.467140098	2.763262455	3.408155178	28
29	1.311433647	1.478704821	1.699127027	2.045229642	2.46202136	2.756385904	3.396240288	29
30	1.310415025	1.477364662	1.697260887	2.042272456	2.457261542	2.749995654	3.385184867	30
31	1.309463549	1.47611315	1.695518783	2.039513446	2.452824193	2.744041919	3.37489928	31
32	1.308572793	1.474941772	1.693888748	2.036933343	2.448677634	2.738481482	3.365305926	32
33	1.307737124	1.473843072	1.692360309	2.034515297	2.4447942	2.733276642	3.356337279	33
34	1.306951587	1.47281049	1.690924255	2.032244509	2.441149628	2.728394367	3.347934313	34
35	1.306211802	1.471838233	1.689572458	2.030107928	2.437722547	2.723805589	3.340045202	35
36	1.305513886	1.470921166	1.688297714	2.028094001	2.434494061	2.71948463	3.332624257	36
37	1.304854381	1.470054719	1.68709362	2.026192463	2.4314474	2.715408722	3.325631045	37
38	1.304230204	1.469234815	1.68595446	2.024394164	2.428567631	2.711557602	3.319029655	38
39	1.303638589	1.468457801	1.684875122	2.02269092	2.42584141	2.707913184	3.312788083	39
40	1.303077053	1.467720399	1.683851013	2.02107539	2.423256779	2.704459267	3.306877114	40
41	1.302543359	1.467019655	1.682878002	2.01954097	2.420802992	2.701181304	3.301272889	41
42	1.302035487	1.466352901	1.681952357	2.018081703	2.41847036	2.698066186	3.295950529	42
43	1.301551608	1.465717725	1.681070703	2.016692199	2.416250129	2.695102079	3.290889821	43
44	1.30109006	1.465111933	1.680229977	2.015367574	2.414134368	2.692278266	3.286071946	44
45	1.300649332	1.464533534	1.679427393	2.014103389	2.412115876	2.689585019	3.281479848	45
46	1.300228048	1.463980712	1.678660414	2.012895599	2.410188096	2.687013492	3.277098029	46
47	1.299824947	1.463451805	1.677926722	2.011740514	2.40834505	2.684555618	3.272912378	47
48	1.299438879	1.462945295	1.677224196	2.010634758	2.406581273	2.682204027	3.268910018	48
49	1.299068785	1.46245979	1.676550893	2.009575237	2.40489176	2.679951974	3.265079173	49
50	1.298713694	1.461994009	1.675905025	2.008559112	2.403271917	2.677793271	3.261409056	50

# Cumulative Poisson Distribution

The governing equations are as Follows:

$$\text{Poisson mass function: } p(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \geq 0$$

$$\text{Poisson distribution function: } F(k; \lambda) = \sum_{m=0}^k e^{-\lambda} \frac{\lambda^m}{m!}, \quad k \geq 0$$

Tables give the values of the probability  $F(k; \lambda)$  for values of  $\lambda$  between (0.01, 10) and  $k$  between (0,20). Figure F.1 shows the Poisson mass function and distribution with  $k$  along the horizontal axis and  $p(k; \lambda)$ ,  $F(k; \lambda)$  along the vertical axis for  $\lambda = 10$ . The representative point is shown for  $k = 10$  and  $p(k; 10) = 0.1251$ ,  $F(k; 10) = 0.5830$ .

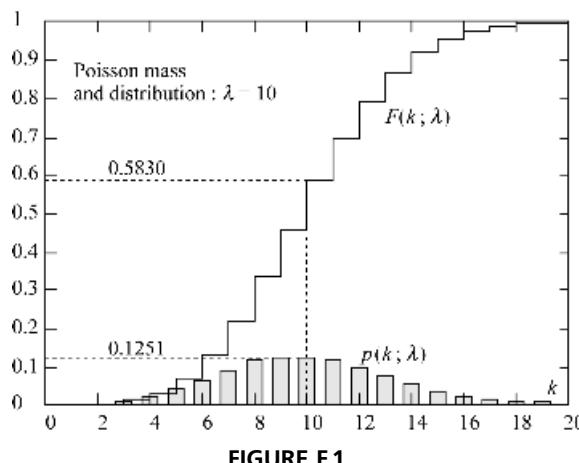


FIGURE F.1

## APPENDIX F Cumulative Poisson distribution

$\lambda = 0.01-4.0$										
$k$	$\lambda = 0.01$	$\lambda = 0.02$	$\lambda = 0.03$	$\lambda = 0.04$	$\lambda = 0.05$	$\lambda = 0.06$	$\lambda = 0.07$	$\lambda = 0.08$	$\lambda = 0.09$	$k$
0	0.99005	0.980199	0.970446	0.960789	0.951229	0.941765	0.932394	0.923116	0.913931	0
1	0.99995	0.999803	0.999559	0.999221	0.998791	0.99827	0.997661	0.996966	0.996185	1
2	1	0.999999	0.999996	0.99999	0.99998	0.999966	0.999946	0.99992	0.999886	2
3	1	1	1	1	1	0.999999	0.999999	0.999998	0.999997	3
$k$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 0.7$	$\lambda = 0.8$	$\lambda = 0.9$	$\lambda = 1.0$
0	0.904837	0.818731	0.740818	0.67032	0.606531	0.548812	0.496585	0.449329	0.40657	0.367879
1	0.995321	0.982477	0.963064	0.938448	0.909796	0.878099	0.844195	0.808792	0.772482	0.735759
2	0.999845	0.998852	0.996401	0.992074	0.985612	0.976885	0.965858	0.952577	0.937143	0.919699
3	0.999996	0.999943	0.999734	0.999224	0.998248	0.996642	0.994247	0.99092	0.986541	0.981012
4	1	0.999998	0.999984	0.999939	0.999828	0.999606	0.999214	0.998589	0.997656	0.99634
5	1	1	0.999999	0.999996	0.999986	0.999961	0.99991	0.999816	0.999657	0.999406
6	1	1	1	1	0.999999	0.999997	0.999991	0.999979	0.999957	0.999917
7	1	1	1	1	1	0.999999	0.999998	0.999995	0.99999	7
8	1	1	1	1	1	1	1	1	0.999999	8
$k$	$\lambda = 1.1$	$\lambda = 1.2$	$\lambda = 1.3$	$\lambda = 1.4$	$\lambda = 1.5$	$\lambda = 1.6$	$\lambda = 1.7$	$\lambda = 1.8$	$\lambda = 1.9$	$\lambda = 2.0$
0	0.332871	0.301194	0.272532	0.246597	0.22313	0.201897	0.182684	0.165299	0.149569	0.135335
1	0.699029	0.662627	0.626823	0.591833	0.557825	0.524931	0.493246	0.462837	0.433749	0.406006
2	0.900416	0.879487	0.857112	0.833498	0.808847	0.783358	0.757223	0.730621	0.70372	0.676676
3	0.974258	0.966231	0.956905	0.946275	0.934358	0.921187	0.906811	0.891292	0.874702	0.857123
4	0.994565	0.992254	0.989337	0.985747	0.981424	0.976318	0.970385	0.963593	0.955919	0.947347
5	0.999032	0.9985	0.997769	0.996799	0.995544	0.99396	0.992001	0.989622	0.986781	0.983436
6	0.999851	0.999749	0.999596	0.999378	0.999074	0.998664	0.998125	0.997431	0.996554	0.995466
7	0.99998	0.999963	0.999936	0.999893	0.99983	0.99974	0.999612	0.999438	0.999207	0.998903
8	0.999998	0.999995	0.999991	0.999984	0.999972	0.999955	0.999928	0.99989	0.999837	0.999763
9	1	0.999999	0.999999	0.999998	0.999996	0.999993	0.999988	0.999981	0.99997	0.999954
10	1	1	1	1	0.999999	0.999999	0.999998	0.999997	0.999995	0.999992
11	1	1	1	1	1	1	1	0.999999	0.999999	11
$k$	$\lambda = 2.2$	$\lambda = 2.4$	$\lambda = 2.6$	$\lambda = 2.8$	$\lambda = 3.0$	$\lambda = 3.2$	$\lambda = 3.4$	$\lambda = 3.6$	$\lambda = 3.8$	$\lambda = 4.0$
0	0.110803	0.090718	0.074274	0.06081	0.049787	0.040762	0.033373	0.027324	0.022371	0.018316
1	0.35457	0.308441	0.267385	0.231078	0.199148	0.171201	0.146842	0.125689	0.10738	0.091578
2	0.622714	0.569709	0.51843	0.469454	0.42319	0.379904	0.33974	0.302747	0.268897	0.238103
3	0.819352	0.778723	0.736002	0.691937	0.647232	0.60252	0.558357	0.515216	0.473485	0.43347
4	0.927504	0.904131	0.877423	0.847676	0.815263	0.780613	0.744182	0.706438	0.667844	0.628837
5	0.97509	0.964327	0.950963	0.93489	0.916082	0.894592	0.870542	0.844119	0.815556	0.78513
6	0.992539	0.988406	0.98283	0.975589	0.966491	0.955381	0.942147	0.926727	0.909108	0.889326
7	0.998022	0.996661	0.994666	0.991869	0.988095	0.98317	0.976926	0.969211	0.959893	0.948866
8	0.99953	0.999138	0.998513	0.997567	0.996197	0.994286	0.991707	0.988329	0.984016	0.978637
9	0.999899	0.999798	0.999624	0.99934	0.998898	0.998238	0.997291	0.995976	0.994201	0.991868
10	0.99998	0.999957	0.999913	0.999836	0.999708	0.999503	0.99919	0.998729	0.998071	0.99716
11	0.999996	0.999992	0.999982	0.999963	0.999929	0.999871	0.999777	0.99963	0.999408	0.999085
12	0.999999	0.999998	0.999996	0.999992	0.999984	0.999969	0.999943	0.9999	0.999832	0.999726
13	1	1	0.999999	0.999998	0.999997	0.999993	0.999986	0.999975	0.999955	0.999924
14	1	1	1	1	0.999999	0.999999	0.999997	0.999994	0.999989	0.99998
15	1	1	1	1	1	0.999999	0.999999	0.999997	0.999995	0.999995
16	1	1	1	1	1	1	1	0.999999	0.999999	16

(Continued)

APPENDIX F *Continued*

$\lambda = 4.2-8.0$											
$k$	$\lambda = 4.2$	$\lambda = 4.4$	$\lambda = 4.6$	$\lambda = 4.8$	$\lambda = 5.0$	$\lambda = 5.2$	$\lambda = 5.4$	$\lambda = 5.6$	$\lambda = 5.8$	$\lambda = 6.0$	$k$
0	0.014996	0.012277	0.010052	0.00823	0.006738	0.005517	0.004517	0.003698	0.003028	0.002479	0
1	0.077977	0.066298	0.05629	0.047733	0.040428	0.034203	0.028906	0.024406	0.020587	0.017351	1
2	0.210238	0.185142	0.162639	0.142539	0.124652	0.108787	0.094758	0.082388	0.071511	0.061969	2
3	0.395403	0.359448	0.325706	0.29423	0.265026	0.238065	0.213291	0.190622	0.169963	0.151204	3
4	0.589827	0.551184	0.513234	0.476259	0.440493	0.406128	0.373311	0.34215	0.312718	0.285057	4
5	0.753143	0.719912	0.68576	0.651006	0.615961	0.580913	0.546132	0.511861	0.478315	0.44568	5
6	0.867464	0.843645	0.818029	0.790805	0.762183	0.732393	0.701671	0.670258	0.638391	0.606303	6
7	0.936057	0.921421	0.904949	0.886666	0.866628	0.844922	0.821659	0.796975	0.771026	0.74398	7
8	0.972068	0.964197	0.954928	0.944183	0.931906	0.918065	0.90265	0.885678	0.867186	0.847237	8
9	0.988873	0.98511	0.980473	0.974859	0.968172	0.960326	0.951245	0.94087	0.929156	0.916076	9
10	0.995931	0.994312	0.992223	0.989583	0.986305	0.982301	0.977486	0.971778	0.965099	0.957379	10
11	0.998626	0.997992	0.997137	0.996008	0.994547	0.99269	0.990368	0.987513	0.98405	0.979908	11
12	0.999569	0.999342	0.999021	0.998578	0.997981	0.997191	0.996165	0.994856	0.99321	0.991173	12
13	0.999874	0.999799	0.999688	0.999527	0.999302	0.998992	0.998573	0.998019	0.997297	0.996372	13
14	0.999966	0.999942	0.999907	0.999853	0.999774	0.999661	0.999502	0.999284	0.99899	0.9986	14
15	0.999991	0.999984	0.999974	0.999957	0.999931	0.999892	0.999836	0.999756	0.999644	0.999491	15
16	0.999998	0.999996	0.999993	0.999988	0.99998	0.999968	0.999949	0.999922	0.999882	0.999823	16
17	1	0.999999	0.999998	0.999997	0.999995	0.999991	0.999985	0.999976	0.999963	0.999943	17
18	1	1	1	0.999999	0.999999	0.999998	0.999996	0.999993	0.999989	0.999982	18
19	1	1	1	1	1	0.999999	0.999999	0.999998	0.999997	0.999995	19
20	1	1	1	1	1	1	1	1	0.999999	0.999999	20
$k$	$\lambda = 6.2$	$\lambda = 6.4$	$\lambda = 6.6$	$\lambda = 6.8$	$\lambda = 7.0$	$\lambda = 7.2$	$\lambda = 7.4$	$\lambda = 7.6$	$\lambda = 7.8$	$\lambda = 8.0$	$k$
0	0.002029	0.001662	0.00136	0.001114	0.000912	0.000747	0.000611	0.0005	0.00041	0.000335	0
1	0.014612	0.012296	0.010339	0.008687	0.007295	0.006122	0.005135	0.004304	0.003606	0.003019	1
2	0.053618	0.046324	0.039968	0.034438	0.029636	0.025474	0.021871	0.018757	0.01607	0.013754	2
3	0.134229	0.118919	0.105151	0.092806	0.081765	0.071917	0.063153	0.055371	0.048477	0.04238	3
4	0.259177	0.23507	0.212704	0.192031	0.172992	0.155516	0.139525	0.124939	0.11167	0.099632	4
5	0.414113	0.383744	0.354673	0.326977	0.300708	0.275897	0.252557	0.230681	0.210251	0.191236	5
6	0.574213	0.542329	0.510839	0.479916	0.449711	0.420356	0.391962	0.364621	0.338407	0.313374	6
7	0.716016	0.687321	0.658082	0.628486	0.598714	0.568941	0.539333	0.510042	0.481209	0.452961	7
8	0.825914	0.803315	0.779557	0.75477	0.729091	0.702668	0.675651	0.648192	0.620441	0.592547	8
9	0.901621	0.885799	0.868639	0.850184	0.830496	0.80965	0.787735	0.764851	0.741109	0.716624	9
10	0.948559	0.938589	0.927433	0.915066	0.901479	0.886677	0.870677	0.853513	0.83523	0.815886	10
11	0.975015	0.969303	0.962709	0.955175	0.94665	0.937094	0.926474	0.91477	0.90197	0.888076	11
12	0.988684	0.985684	0.982111	0.977903	0.973	0.967345	0.960883	0.953566	0.945351	0.936203	12
13	0.995203	0.993749	0.991962	0.989792	0.987189	0.984099	0.980469	0.976247	0.97138	0.965819	13
14	0.99809	0.997435	0.996605	0.995566	0.994283	0.992715	0.990822	0.988559	0.985882	0.982743	14
15	0.999284	0.999008	0.998648	0.998184	0.997593	0.996851	0.995929	0.994798	0.993423	0.991769	15
16	0.999746	0.999638	0.999491	0.999297	0.999042	0.998712	0.998291	0.997761	0.997099	0.996282	16
17	0.999915	0.999874	0.999818	0.999742	0.999638	0.9995	0.99932	0.999085	0.998785	0.998406	17
18	0.999973	0.999959	0.999938	0.99991	0.99987	0.999816	0.999742	0.999645	0.999516	0.99935	18
19	0.999992	0.999987	0.99998	0.99997	0.999956	0.999935	0.999907	0.999868	0.999816	0.999747	19
20	0.999998	0.999996	0.999994	0.99999	0.999986	0.999978	0.999968	0.999954	0.999933	0.999906	20
21	0.999999	0.999999	0.999998	0.999997	0.999995	0.999993	0.999989	0.999984	0.999977	0.999967	21
22	1	1	0.999999	0.999999	0.999999	0.999998	0.999997	0.999995	0.999992	0.999989	22
23	1	1	1	1	1	1	1	1	0.999998	0.999996	23
24	1	1	1	1	1	1	1	1	0.999999	0.999999	24

(Continued)

APPENDIX F *Continued*

$\lambda = 8.2\text{--}10.0$											
$k$	$\lambda = 8.2$	$\lambda = 8.4$	$\lambda = 8.6$	$\lambda = 8.8$	$\lambda = 9.0$	$\lambda = 9.2$	$\lambda = 9.4$	$\lambda = 9.6$	$\lambda = 9.8$	$\lambda = 10.0$	$k$
0	0.000275	0.000225	0.000184	0.000151	0.000123	0.000101	0.000083	0.000068	0.000055	0.000045	0
1	0.002527	0.002114	0.001767	0.001477	0.001234	0.001031	0.00086	0.000718	0.000599	0.000499	1
2	0.011761	0.010047	0.008576	0.007314	0.006232	0.005307	0.004515	0.003839	0.003262	0.002769	2
3	0.037	0.03226	0.028093	0.024434	0.021226	0.01842	0.015967	0.013826	0.01196	0.010336	3
4	0.08874	0.078908	0.070054	0.062098	0.054964	0.04858	0.042878	0.037795	0.033271	0.029253	4
5	0.173594	0.157277	0.142228	0.128387	0.115691	0.104074	0.093471	0.083815	0.075041	0.067086	5
6	0.289562	0.266993	0.245676	0.22561	0.206781	0.189165	0.172733	0.157447	0.143265	0.130141	6
7	0.425409	0.398652	0.372771	0.347834	0.323897	0.301	0.279171	0.258428	0.238779	0.220221	7
8	0.564653	0.536894	0.509397	0.482281	0.455653	0.429609	0.404235	0.379606	0.355783	0.33282	8
9	0.691519	0.66592	0.639951	0.61374	0.587408	0.561076	0.534858	0.508862	0.483188	0.45793	9
10	0.79555	0.774301	0.752228	0.729423	0.705988	0.682026	0.657644	0.632948	0.608045	0.58304	10
11	0.8731	0.857066	0.840008	0.82197	0.803008	0.783185	0.76257	0.741241	0.719281	0.696776	11
12	0.926093	0.915001	0.902916	0.889838	0.875773	0.860739	0.844762	0.827876	0.810124	0.791556	12
13	0.959519	0.952436	0.944533	0.935779	0.926149	0.915624	0.904193	0.891852	0.878605	0.864464	13
14	0.979097	0.974897	0.970098	0.964657	0.958534	0.951691	0.944097	0.935721	0.926542	0.916542	14
15	0.989799	0.987475	0.984755	0.981598	0.977964	0.973812	0.969103	0.963798	0.957861	0.95126	15
16	0.995285	0.994078	0.992633	0.990916	0.988894	0.986532	0.983794	0.980643	0.977044	0.972958	16
17	0.99793	0.997341	0.996618	0.995739	0.99468	0.993416	0.991917	0.990156	0.988102	0.985722	17
18	0.999136	0.998864	0.998522	0.998097	0.997574	0.996934	0.99616	0.99523	0.994123	0.992813	18
19	0.999656	0.999537	0.999384	0.999189	0.998944	0.998638	0.998258	0.997793	0.997228	0.996546	19
20	0.999869	0.99982	0.999755	0.99967	0.999561	0.999421	0.999245	0.999024	0.99875	0.998412	20
21	0.999952	0.999933	0.999906	0.999871	0.999825	0.999765	0.999686	0.999586	0.99946	0.9993	21
22	0.999983	0.999976	0.999966	0.999952	0.999933	0.999908	0.999875	0.999832	0.999776	0.999704	22
23	0.999994	0.999992	0.999988	0.999983	0.999975	0.999966	0.999952	0.999934	0.999911	0.99988	23
24	0.999998	0.999997	0.999996	0.999994	0.999991	0.999988	0.999982	0.999975	0.999966	0.999953	24
25	0.999999	0.999999	0.999999	0.999998	0.999997	0.999996	0.999994	0.999991	0.999987	0.999982	25
26	1	1	1	0.999999	0.999999	0.999999	0.999998	0.999997	0.999996	0.999994	26
27	1	1	1	1	1	1	0.999999	0.999999	0.999998	0.999998	27
28	1	1	1	1	1	1	1	0.999999	0.999999	0.999998	28

# Cumulative Binomial Distribution

The governing equations are as follows:

$$\text{Binomial mass function: } b(k; n, p) = \frac{n!}{k!(n-k)!} \times p^k (1-p)^{n-k}, \quad 0 \leq k \leq n$$

$$\text{Binomial distribution function: } F(k; n, p) = \sum_{m=0}^k \frac{n!}{m!(n-m)!} \times p^m (1-p)^{n-m}, \quad 0 \leq k \leq n$$

Tables give the values of the probability  $F(k; n, p)$  for values of  $p$  between (0.05, 0.9) and  $n$  between (2, 20). Figure G.1 shows the binomial mass function and distribution with  $k$  along the horizontal axis and  $b(k; n, p)$ ,  $F(k; n, p)$  along the vertical axis for  $p = 0.3$  and  $n = 10$ . The representative point is shown for  $k = 3$  and  $b(k; 10, 0.3) = 0.2668$ ,  $F(k; 10, 0.3) = 0.6496$ . The range of  $k$  is only upto  $n$ .

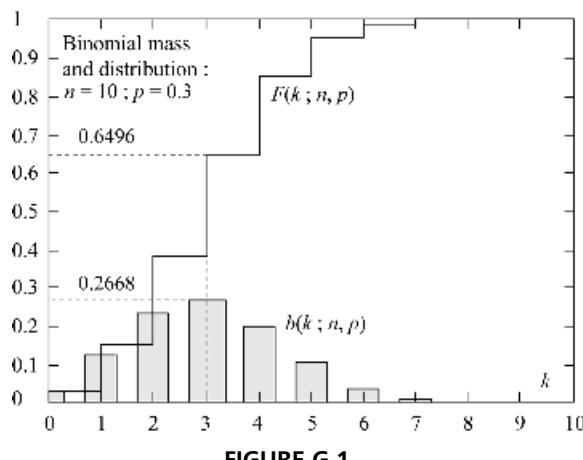


FIGURE G.1

## APPENDIX G Cumulative binomial distribution

$n = 2, 8$											
$k$	$p = 0.05$	$p = 0.1$	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$	$p = 0.6$	$p = 0.7$	$p = 0.8$	$p = 0.9$	
$n = 2$											
0	0.9025	0.81	0.64	0.49	0.36	0.25	0.16	0.09	0.04	0.01	
1	0.9975	0.99	0.96	0.91	0.84	0.75	0.64	0.51	0.36	0.19	
2	1	1	1	1	1	1	1	1	1	1	
$n = 3$											
0	0.857375	0.729	0.512	0.343	0.216	0.125	0.064	0.027	0.008	0.001	
1	0.99275	0.972	0.896	0.784	0.648	0.5	0.352	0.216	0.104	0.028	
2	0.999875	0.999	0.992	0.973	0.936	0.875	0.784	0.657	0.488	0.271	
3	1	1	1	1	1	1	1	1	1	1	
$n = 4$											
0	0.814506	0.6561	0.4096	0.2401	0.1296	0.0625	0.0256	0.0081	0.0016	0.0001	
1	0.985981	0.9477	0.8192	0.6517	0.4752	0.3125	0.1792	0.0837	0.0272	0.0037	
2	0.999519	0.9963	0.9728	0.9163	0.8208	0.6875	0.5248	0.3483	0.1808	0.0523	
3	0.999994	0.9999	0.9984	0.9919	0.9744	0.9375	0.8704	0.7599	0.5904	0.3439	
4	1	1	1	1	1	1	1	1	1	1	
$n = 5$											
0	0.773781	0.59049	0.32768	0.16807	0.07776	0.03125	0.01024	0.00243	0.00032	0.00001	
1	0.977407	0.91854	0.73728	0.52822	0.33696	0.1875	0.08704	0.03078	0.00672	0.00046	
2	0.998842	0.99144	0.94208	0.83692	0.68256	0.5	0.31744	0.16308	0.05792	0.00856	
3	0.99997	0.99954	0.99328	0.96922	0.91296	0.8125	0.66304	0.47178	0.26272	0.08146	
4	1	0.99999	0.99968	0.99757	0.98976	0.96875	0.92224	0.83193	0.67232	0.40951	
5	1	1	1	1	1	1	1	1	1	1	
$n = 6$											
0	0.735092	0.531441	0.262144	0.117649	0.046656	0.015625	0.004096	0.000729	0.000064	0.000001	
1	0.967226	0.885735	0.65536	0.420175	0.23328	0.109375	0.04096	0.010935	0.0016	0.000055	
2	0.99777	0.98415	0.90112	0.74431	0.54432	0.34375	0.1792	0.07047	0.01696	0.00127	
3	0.999914	0.99873	0.98304	0.92953	0.8208	0.65625	0.45568	0.25569	0.09888	0.01585	
4	0.999998	0.999945	0.9984	0.989065	0.95904	0.890625	0.76672	0.579825	0.34464	0.114265	
5	1	0.999999	0.999936	0.999271	0.995904	0.984375	0.953344	0.882351	0.737856	0.468559	
6	1	1	1	1	1	1	1	1	1	1	
$n = 7$											
0	0.698337	0.478297	0.209715	0.082354	0.027994	0.007812	0.001638	0.000219	0.000013	0	
1	0.955619	0.850306	0.576717	0.329417	0.15863	0.0625	0.018842	0.003791	0.000371	0.000006	
2	0.996243	0.974309	0.851968	0.64707	0.419904	0.226563	0.096256	0.028796	0.004672	0.000177	
3	0.999806	0.997272	0.966656	0.873964	0.710208	0.5	0.289792	0.126036	0.033344	0.002728	
4	0.999994	0.999823	0.995328	0.971205	0.903744	0.773437	0.580096	0.35293	0.148032	0.025691	
5	1	0.999994	0.999629	0.996209	0.981158	0.9375	0.84137	0.670583	0.423283	0.149694	
6	1	1	0.999987	0.999781	0.998362	0.992188	0.972006	0.917646	0.790285	0.521703	
7	1	1	1	1	1	1	1	1	1	1	
$n = 8$											
0	0.66342	0.430467	0.167772	0.057648	0.016796	0.003906	0.000655	0.000066	0.000003	0	
1	0.942755	0.813105	0.503316	0.255298	0.106376	0.035156	0.00852	0.00129	0.000084	0.000001	
2	0.994212	0.961908	0.796918	0.551774	0.315395	0.144531	0.049807	0.011292	0.001231	0.000023	
3	0.999628	0.994976	0.943718	0.805896	0.594086	0.363281	0.17367	0.057968	0.010406	0.000432	
4	0.999985	0.999568	0.989594	0.942032	0.82633	0.636719	0.405914	0.194104	0.056282	0.005024	
5	1	0.999977	0.998769	0.988708	0.950193	0.855469	0.684605	0.448226	0.203082	0.038092	
6	1	0.999999	0.999916	0.99871	0.99148	0.964844	0.893624	0.744702	0.496684	0.186895	
7	1	1	0.999997	0.999934	0.999345	0.996094	0.983204	0.942352	0.832228	0.569533	
8	1	1	1	1	1	1	1	1	1	1	

(Continued)

APPENDIX G *Continued*

<i>n</i> = 9,12											
<i>k</i>	<i>p</i> = 0.05	<i>p</i> = 0.1	<i>p</i> = 0.2	<i>p</i> = 0.3	<i>p</i> = 0.4	<i>p</i> = 0.5	<i>p</i> = 0.6	<i>p</i> = 0.7	<i>p</i> = 0.8	<i>p</i> = 0.9	
<i>n</i> = 9											
0	0.630249	0.38742	0.134218	0.040354	0.010078	0.001953	0.000262	0.00002	0.000001	0	
1	0.928789	0.774841	0.436208	0.196003	0.070544	0.019531	0.003801	0.000433	0.000019	0	
2	0.991639	0.947028	0.738198	0.462831	0.231787	0.089844	0.025035	0.004291	0.000314	0.000003	
3	0.999357	0.991669	0.914358	0.729659	0.48261	0.253906	0.099353	0.025295	0.003066	0.000064	
4	0.999967	0.999109	0.980419	0.901191	0.733432	0.5	0.266568	0.098809	0.019581	0.000891	
5	0.999999	0.999936	0.996934	0.974705	0.900647	0.746094	0.51739	0.270341	0.085642	0.008331	
6	1	0.999997	0.999686	0.995709	0.974965	0.910156	0.768213	0.537169	0.261802	0.052972	
7	1	1	0.999981	0.999567	0.996199	0.980469	0.929456	0.803997	0.563792	0.225159	
8	1	1	0.999999	0.99998	0.999738	0.998047	0.989922	0.959646	0.865782	0.61258	
9	1	1	1	1	1	1	1	1	1	1	
<i>n</i> = 10											
0	0.598737	0.348678	0.107374	0.028248	0.006047	0.000977	0.000105	0.000006	0	0	
1	0.913862	0.736099	0.37581	0.149308	0.046357	0.010742	0.001678	0.000144	0.000004	0	
2	0.988496	0.929809	0.6778	0.382783	0.16729	0.054688	0.012295	0.00159	0.000078	0	
3	0.998972	0.987205	0.879126	0.649611	0.382281	0.171875	0.054762	0.010592	0.000864	0.000009	
4	0.999936	0.998365	0.967207	0.849732	0.633103	0.376953	0.166239	0.047349	0.006369	0.000147	
5	0.999997	0.999853	0.993631	0.952651	0.833761	0.623047	0.366897	0.150268	0.032793	0.001635	
6	1	0.999991	0.999136	0.989408	0.945238	0.828125	0.617719	0.350389	0.120874	0.012795	
7	1	1	0.999922	0.99841	0.987705	0.945312	0.83271	0.617217	0.3222	0.070191	
8	1	1	0.999996	0.999856	0.998322	0.989258	0.953643	0.850692	0.62419	0.263901	
9	1	1	1	0.999994	0.999895	0.999023	0.993953	0.971752	0.892626	0.651322	
10	1	1	1	1	1	1	1	1	1	1	
<i>n</i> = 11											
0	0.5688	0.313811	0.085899	0.019773	0.003628	0.000488	0.000042	0.000002	0	0	
1	0.898105	0.697357	0.322123	0.11299	0.030233	0.005859	0.000734	0.000047	0.000001	0	
2	0.984765	0.910438	0.617402	0.31274	0.118917	0.032715	0.005924	0.000578	0.000019	0	
3	0.998448	0.981465	0.838861	0.569562	0.296284	0.113281	0.029281	0.004291	0.000235	0.000001	
4	0.999888	0.997249	0.94959	0.789695	0.532774	0.274414	0.099353	0.021619	0.001965	0.000023	
5	0.999994	0.999704	0.988346	0.921775	0.753498	0.5	0.246502	0.078225	0.011654	0.000296	
6	1	0.999977	0.998035	0.978381	0.900647	0.725586	0.467226	0.210305	0.05041	0.002751	
7	1	0.999999	0.999765	0.995709	0.970719	0.886719	0.703716	0.430438	0.161139	0.018535	
8	1	1	0.999981	0.999422	0.994076	0.967285	0.881083	0.68726	0.382598	0.089562	
9	1	1	0.999999	0.999953	0.999266	0.994141	0.969767	0.88701	0.677877	0.302643	
10	1	1	1	0.999998	0.999958	0.999512	0.996372	0.980227	0.914101	0.686189	
11	1	1	1	1	1	1	1	1	1	1	
<i>n</i> = 12											
0	0.54036	0.28243	0.068719	0.013841	0.002177	0.000244	0.000017	0.000001	0	0	
1	0.88164	0.659002	0.274878	0.085025	0.019591	0.003174	0.000319	0.000015	0	0	
2	0.980432	0.88913	0.558346	0.252815	0.083443	0.019287	0.00281	0.000206	0.000005	0	
3	0.997764	0.974363	0.794569	0.492516	0.225337	0.072998	0.015267	0.001692	0.000062	0	
4	0.999816	0.995671	0.927445	0.723655	0.438178	0.193848	0.05731	0.009489	0.000581	0.000003	
5	0.999989	0.999459	0.980595	0.882151	0.665209	0.387207	0.158212	0.038601	0.003903	0.00005	
6	1	0.99995	0.996097	0.961399	0.841788	0.612793	0.334791	0.117849	0.019405	0.000541	
7	1	0.999997	0.999419	0.990511	0.94269	0.806152	0.561822	0.276345	0.072555	0.004329	
8	1	1	0.999938	0.998308	0.984733	0.927002	0.774663	0.507484	0.205431	0.025637	
9	1	1	0.999995	0.999794	0.99719	0.980713	0.916557	0.747185	0.441654	0.11087	
10	1	1	1	0.999985	0.999681	0.996826	0.980409	0.914975	0.725122	0.340998	
11	1	1	1	0.999999	0.999983	0.999756	0.997823	0.986159	0.931281	0.71757	
12	1	1	1	1	1	1	1	1	1	1	

(Continued)

APPENDIX G *Continued*

<i>n</i> = 13,15											
<i>k</i>	<i>p</i> = 0.05	<i>p</i> = 0.1	<i>p</i> = 0.2	<i>p</i> = 0.3	<i>p</i> = 0.4	<i>p</i> = 0.5	<i>p</i> = 0.6	<i>p</i> = 0.7	<i>p</i> = 0.8	<i>p</i> = 0.9	
<i>n</i> = 13											
0	0.513342	0.254187	0.054976	0.009689	0.001306	0.000122	0.000007	0	0	0	0
1	0.864576	0.621345	0.233646	0.06367	0.012625	0.001709	0.000138	0.000005	0	0	0
2	0.975492	0.866117	0.501652	0.202478	0.057902	0.01123	0.001315	0.000073	0.000001	0	0
3	0.996897	0.965839	0.747324	0.420606	0.16858	0.046143	0.007793	0.000652	0.000016	0	0
4	0.999713	0.99354	0.900869	0.654314	0.353042	0.133423	0.032084	0.004031	0.000166	0	0
5	0.99998	0.99908	0.969965	0.834603	0.574396	0.290527	0.097671	0.018223	0.001246	0.000008	0
6	0.999999	0.999901	0.992996	0.937625	0.771156	0.5	0.228844	0.062375	0.007004	0.000099	0
7	1	0.999992	0.998754	0.981777	0.902329	0.709473	0.425604	0.165397	0.030035	0.00092	0
8	1	1	0.999834	0.995969	0.967916	0.866577	0.646958	0.345686	0.099131	0.00646	0
9	1	1	0.999984	0.999348	0.992207	0.953857	0.83142	0.579394	0.252676	0.034161	0
10	1	1	0.999999	0.999927	0.998685	0.98877	0.942098	0.797522	0.498348	0.133883	0
11	1	1	1	0.999995	0.999862	0.998291	0.987375	0.93633	0.766354	0.378655	0
12	1	1	1	1	0.999993	0.999878	0.998694	0.990311	0.945024	0.745813	0
13	1	1	1	1	1	1	1	1	1	1	0
<i>n</i> = 14											
0	0.487675	0.228768	0.04398	0.006782	0.000784	0.000061	0.000003	0	0	0	0
1	0.847014	0.584629	0.197912	0.047476	0.008098	0.000916	0.000059	0.000002	0	0	0
2	0.969946	0.84164	0.448051	0.160836	0.039792	0.00647	0.000609	0.000025	0	0	0
3	0.995827	0.955867	0.69819	0.355167	0.124309	0.028687	0.003906	0.000246	0.000004	0	0
4	0.999573	0.99077	0.87016	0.584201	0.279257	0.089783	0.01751	0.001666	0.000046	0	0
5	0.999967	0.998526	0.956146	0.780516	0.485855	0.211975	0.058319	0.008289	0.000382	0.000001	0
6	0.999998	0.999819	0.98839	0.906718	0.692452	0.395264	0.15014	0.031469	0.002397	0.000017	0
7	1	0.999983	0.997603	0.968531	0.849986	0.604736	0.307548	0.093282	0.01161	0.000181	0
8	1	0.999999	0.999618	0.991711	0.941681	0.788025	0.514145	0.219484	0.043854	0.001474	0
9	1	1	0.999954	0.998334	0.98249	0.910217	0.720743	0.415799	0.12984	0.00923	0
10	1	1	0.999996	0.999754	0.996094	0.971313	0.875691	0.644833	0.30181	0.044133	0
11	1	1	1	0.999975	0.999391	0.99353	0.960208	0.839164	0.551949	0.15836	0
12	1	1	1	0.999998	0.999941	0.999084	0.991902	0.952524	0.802088	0.415371	0
13	1	1	1	1	0.999997	0.999939	0.999216	0.993218	0.95602	0.771232	0
14	1	1	1	1	1	1	1	1	1	1	0
<i>n</i> = 15											
0	0.463291	0.205891	0.035184	0.004748	0.00047	0.000031	0.000001	0	0	0	0
1	0.829047	0.549043	0.167126	0.035268	0.005172	0.000488	0.000025	0.000001	0	0	0
2	0.9638	0.815939	0.398023	0.126828	0.027114	0.003693	0.000279	0.000009	0	0	0
3	0.994533	0.944444	0.648162	0.296868	0.090502	0.017578	0.001928	0.000092	0.000001	0	0
4	0.999385	0.98728	0.835766	0.515491	0.217278	0.059235	0.009348	0.000672	0.000012	0	0
5	0.999947	0.99775	0.938949	0.721621	0.403216	0.150879	0.033833	0.003653	0.000113	0	0
6	0.999996	0.999689	0.981941	0.868857	0.609813	0.303619	0.095047	0.015243	0.000785	0.000003	0
7	1	0.999966	0.99576	0.949987	0.786897	0.5	0.213103	0.050013	0.00424	0.000034	0
8	1	0.999997	0.999215	0.984757	0.904953	0.696381	0.390187	0.131143	0.018059	0.000311	0
9	1	1	0.999887	0.996347	0.966167	0.849121	0.596784	0.278379	0.061051	0.00225	0
10	1	1	0.999988	0.999328	0.990652	0.940765	0.782722	0.484509	0.164234	0.01272	0
11	1	1	0.999999	0.999908	0.998072	0.982422	0.909498	0.703132	0.351838	0.055556	0
12	1	1	1	0.999991	0.999721	0.996307	0.972886	0.873172	0.601977	0.184061	0
13	1	1	1	0.999999	0.999975	0.999512	0.994828	0.964732	0.832874	0.450957	0
14	1	1	1	1	0.999999	0.999969	0.99953	0.995252	0.964816	0.794109	0
15	1	1	1	1	1	1	1	1	1	1	0

(Continued)

APPENDIX G *Continued*

$n = 16, 18$											
$k$	$p = 0.05$	$p = 0.1$	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$	$p = 0.6$	$p = 0.7$	$p = 0.8$	$p = 0.9$	
$n = 16$											
0	0.440127	0.185302	0.028147	0.003323	0.000282	0.000015	0	0	0	0	
1	0.81076	0.514728	0.140737	0.026112	0.003291	0.000259	0.000011	0	0	0	
2	0.957062	0.789249	0.351844	0.09936	0.018337	0.00209	0.000127	0.000003	0	0	
3	0.992996	0.931594	0.598134	0.245856	0.065147	0.010635	0.000938	0.000034	0	0	
4	0.999143	0.982996	0.798245	0.449904	0.166567	0.038406	0.004896	0.000266	0.000003	0	
5	0.999919	0.996703	0.918312	0.659782	0.32884	0.105057	0.019142	0.001566	0.000033	0	
6	0.999994	0.999495	0.973343	0.824687	0.527174	0.227249	0.058319	0.00713	0.000248	0	
7	1	0.999939	0.992996	0.925648	0.716063	0.40181	0.14227	0.025674	0.001476	0.000006	
8	1	0.999994	0.998524	0.974326	0.85773	0.59819	0.283937	0.074352	0.007004	0.000061	
9	1	1	0.999752	0.99287	0.941681	0.772751	0.472826	0.175313	0.026657	0.000505	
10	1	1	0.999967	0.998434	0.980858	0.894943	0.67116	0.340218	0.081688	0.003297	
11	1	1	0.999997	0.999734	0.995104	0.961594	0.833433	0.550096	0.201755	0.017004	
12	1	1	1	0.999966	0.999062	0.989365	0.934853	0.754144	0.401866	0.068406	
13	1	1	1	0.999997	0.999873	0.99791	0.981663	0.90064	0.648156	0.210751	
14	1	1	1	1	0.999989	0.999741	0.996709	0.973888	0.859263	0.485272	
15	1	1	1	1	1	0.999985	0.999718	0.996677	0.971853	0.814698	
16	1	1	1	1	1	1	1	1	1	1	
$n = 17$											
0	0.41812	0.166772	0.022518	0.002326	0.000169	0.000008	0	0	0	0	
1	0.792228	0.481785	0.118219	0.019275	0.002088	0.000137	0.000005	0	0	0	
2	0.949747	0.761797	0.309622	0.077385	0.012319	0.001175	0.000057	0.000001	0	0	
3	0.991199	0.917359	0.548876	0.201907	0.046423	0.006363	0.000451	0.000012	0	0	
4	0.998835	0.977856	0.758223	0.38869	0.125999	0.024521	0.002521	0.000103	0.000001	0	
5	0.99988	0.995333	0.894299	0.596819	0.263931	0.071732	0.010594	0.000656	0.000009	0	
6	0.99999	0.999216	0.962337	0.775215	0.447841	0.166153	0.034813	0.003235	0.000076	0	
7	0.999999	0.999894	0.989066	0.89536	0.640508	0.314529	0.091899	0.012693	0.000493	0.000001	
8	1	0.999989	0.997419	0.959723	0.801064	0.5	0.198936	0.040277	0.002581	0.000011	
9	1	0.999999	0.999507	0.987307	0.908101	0.685471	0.359492	0.10464	0.010934	0.000106	
10	1	1	0.999924	0.996765	0.965187	0.833847	0.552159	0.224785	0.037663	0.000784	
11	1	1	0.999991	0.999344	0.989406	0.928268	0.736069	0.403181	0.105701	0.004667	
12	1	1	0.999999	0.999897	0.997479	0.975479	0.874001	0.61131	0.241777	0.022144	
13	1	1	1	0.999988	0.999549	0.993637	0.953577	0.798093	0.451124	0.082641	
14	1	1	1	0.999999	0.999943	0.998825	0.987681	0.922615	0.690378	0.238203	
15	1	1	1	1	0.999995	0.999863	0.997912	0.980725	0.881781	0.518215	
16	1	1	1	1	0.999992	0.999831	0.997674	0.977482	0.833228		
17	1	1	1	1	1	1	1	1	1		
$n = 18$											
0	0.397214	0.150095	0.018014	0.001628	0.000102	0.000004	0	0	0	0	
1	0.773523	0.450284	0.099079	0.01419	0.00132	0.000072	0.000002	0	0	0	
2	0.941871	0.733796	0.271342	0.059952	0.008226	0.000656	0.000026	0	0	0	
3	0.989127	0.901803	0.501025	0.16455	0.032781	0.003769	0.000215	0.000004	0	0	
4	0.998454	0.971806	0.716354	0.332655	0.094169	0.015442	0.001279	0.000039	0	0	
5	0.999828	0.993585	0.867084	0.53438	0.208758	0.048126	0.00575	0.000269	0.000003	0	
6	0.999985	0.998828	0.948729	0.721696	0.374277	0.118942	0.020282	0.00143	0.000022	0	
7	0.999999	0.999827	0.98372	0.859317	0.563441	0.240341	0.057647	0.006073	0.000159	0	
8	1	0.999979	0.995748	0.940414	0.736841	0.407265	0.134714	0.020968	0.000911	0.000002	
9	1	0.999998	0.999089	0.979032	0.865286	0.592735	0.263159	0.059586	0.004252	0.000021	
10	1	1	0.999841	0.993927	0.942353	0.759659	0.436559	0.140683	0.01628	0.000173	
11	1	1	0.999978	0.99857	0.979718	0.881058	0.625723	0.278304	0.051271	0.001172	
12	1	1	0.999997	0.999731	0.99425	0.951874	0.791242	0.46562	0.132916	0.006415	

(Continued)

## APPENDIX G *Continued*

# REFERENCES

1. Abramowitz, M., and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, 1972.
2. Akansu, A. N., and R. A. Haddad, *Multiresolution Signal Processing*, Academic Press, 1992.
3. Anderson, R. L., Distribution of the serial correlation coefficient, *Annals Math. Statist.*, **13**, 1–13, 1942.
4. Aunon, J. I., and V. Chandrasekhar, *Introduction to Probability and Random Processes*, McGraw-Hill, 1997.
5. Benford, F., The law of anomalous numbers, *Proc. Am. Phil. Soc.*, **78**, 551–572, 1938.
6. Bharucha-Reid, A. T., *Elements of the Theory of Markov Processes and Their Applications*, Dover Publications, 1988.
7. Black, F., and M. Scholes, The pricing of options and corporate liabilities, *J. Polit. Econ.*, 1973.
8. Börjesson, P. O., and C. E. W. Sundberg, Simple approximations of the error function  $Q(x)$  for communications applications, *IEEE Trans. Commun.*, **27**, 639–643, 1979.
9. Box, G. E. P., G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis*, Prentice-Hall, 1994.
10. Chatfield, C., *The Analysis of Time Series: An Introduction*, 2nd ed., Chapman & Hall, 1980.
11. Childers, D. G., *Probability and Random Processes*, Irwin, 1997.
12. Cooper, G. R., and C. D. McGillem, *Probabilistic Methods of Signal and System Analysis*, 3rd ed., Oxford Univ. Press, 1999.
13. Cox, D. R., and H. D. Miller, *The Theory of Stochastic Processes*, Chapman & Hall/CRC, 1977.
14. Davenport, W. B., and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, IEEE Press, 1987.
15. Davis, J., *Performance Models for Correlated Traffic Data*, Ph.D. thesis, Univ. Massachusetts—Lowell, 2000.
16. Doob, J. L., *Stochastic Processes*, Wiley, 1953.
17. Feller, W., *An Introduction to Probability Theory and Its Applications*, 2nd ed., Wiley, 1968, Vol. I.
18. Fogiel, M., *Handbook of Mathematical, Scientific, and Engineering Formulas, Tables, Functions, Graphs, Transforms*, Staff of Research and Education Assoc., 1988.
19. Fujisaki, M., G. Kallianpur, and H. Kunita, Stochastic differential equations for nonlinear filtering problem, *Osaka J. Math.*, **9**, 19–40, 1972.
20. Gantmacher, F., *The Theory of Matrices*, Chelsea Publishing, 1959.
21. Gardner, W. A., *Introduction to Random Processes*, 2nd ed., McGraw-Hill, 1990.
22. Grimmett G., and D. Stirzaker, *Probability and Random Processes*, 3rd ed., Oxford Univ. Press, 2001.
23. Gross, D., and C. M. Harris, *Queueing Theory*, Wiley, 1998.
24. Higgins, J. J., and S. Keller-McNulty, *Concepts in Probability and Stochastic Modeling*, Duxbury Press, 1995.
25. Jenkins, G. M., and D. G. Watts, *Spectral Analysis and Its Applications*, Holden-Day, 1968.
26. Kallianpur, G., *Stochastic Filtering Theory*, Springer-Verlag, 1980.
27. Kalman, R. E., and R. S. Bucy, New results in linear filtering and prediction theory, *Trans. ASME, J. Basic Eng., ser. D*, **83**, 95–108, 1961.
28. Kamonpet, P., and V. Krishnan, *Modified EM Algorithm for Transmission Tomography*, Internal Memo T94-1, Center for Advanced Computation, Univ. Massachusetts Lowell, 1994.

29. Kamonpet, P., *Probabilistic Image Reconstruction from Noisy and Incomplete Projections*, Ph.D. thesis, Univ. Massachusetts—Lowell, 1995.
30. Kleinrock, L., *Queueing Systems, Theory*, Wiley, 1975, Vol. II.
31. Kleinrock, L., *Queueing Systems, Computer Applications*, John Wiley, 1976, Vol. II.
32. Krishnan, V., *Estimation of Autocovariance Functions*, Technical Report M-215U, Dynamics Research Corporation, 1974.
33. Krishnan, V., *Confidence Limits and Equivalence of Statistical Tests*, Technical Report M-256U, Dynamics Research Corporation, 1976.
34. Krishnan, V., *Distribution of Spectral Estimators*, Technical Report E-9334U, Dynamics Research Corporation, 1984.
35. Krishnan, V., *Nonlinear Filtering and Smoothing*, Wiley, 1984.
36. Krishnan, V., *Linear Systems Properties*, CRC Press, 1998.
37. Krishnan, V., *Nonlinear Filtering and Smoothing*, Dover Publications, 2005.
38. Lange, K., Convergence of EM image reconstruction algorithms with Gibbs' smoothing, *IEEE Trans. Med. Imag.*, **9**, 439–446, 1990.
39. Leon-Garcia, A., *Probability and Random Processes for Electrical Engineering*, Addison-Wesley, 1993.
40. Lipster, R. S., and A. N. Shiryaev, *Statistics of Random Processes*, Springer-Verlag, 1977, Vols. I and II.
41. Meyer, P. A., A decomposition theorem for supermartingales, *Ill. J. Math.* **6**, 193–205, 1962.
42. Mood, A. M., and F. A. Graybill, *Introduction to the Theory of Statistics*, McGraw-Hill, 1963.
43. Naidu, P. S. *Modern Spectrum Analysis of Time Series*, CRC Press, 1996.
44. Ostle, B., *Statistics in Research*, Iowa State Univ. Press, 1963.
45. Papoulis, A., *Fourier Integral and Its Applications*, Mc-Graw Hill, 1962.
46. Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, 1991.
47. Peebles Jr., Peyton Z., *Probability, Random Variables and Random Signal Principles*, 4th ed., Mc-Graw-Hill, 2001.
48. Priestley, M. B., *Spectral Analysis and Time Series*, Academic Press, 1981.
49. Proakis, J. G., *Digital Communications*, McGraw-Hill, 1995.
50. Rice, S. O., Mathematical analysis of random noise, in *Selected Papers on Noise and Stochastic Processes*, N. Wax, ed., Dover Publications, 1945, 1954.
51. Rome, H. J., and V. Krishnan, Causal probability model for evaluating future transoceanic airplane separations, *IEEE Trans. Aerospace Electron. Syst.*, **26**, 804–817, 1990.
52. Ross, S. M., *A First Course in Probability*, 6th ed., Prentice-Hall, 2002.
53. Ross, S. M., *Stochastic Processes*, 2nd ed., John-Wiley, 1996.
54. Ross, S. M., *Introduction to Probability Models*, 7th ed., Academic Press, 2000.
55. Shanmugan, S., and A. M. Breiphol, *Random Signals, Detection, Estimation and Data Analysis*, Wiley, 1988.
56. Shilov, G. E., *An Introduction to the Theory of Linear Spaces*, Prentice-Hall, 1961.
57. Srinath, M. D., P. K. Rajasekaran, and R. Viswanathan, *Introduction to Statistical Signal Processing with Applications*, Prentice-Hall, 1996.
58. Stark, H., and J. W. Woods, *Probability and Random Processes with Applications to Signal Processing*, 3rd ed., Prentice-Hall, 2002.
59. Therrien, C. W., *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall, 1992.
60. Thomas, J. B., *An Introduction to Communication Theory and Systems*, Springer-Verlag, 1988.
61. Trivedi, K. S., *Probability and Statistics with Reliability, Queueing and Computer Science Applications*, Prentice-Hall, 1982.
62. Viniotis, Y., *Probability and Random Processes for Electrical Engineers*, McGraw-Hill, 1998.
63. Xu, L., *Probabilistic Reconstruction of Tomographic Images*, Ph.D. thesis, Univ. Massachusetts—Lowell, 1997.
64. Yates, R. D., and D. J. Goodman, *Probability and Stochastic Processes*, Wiley, 1999.

## Web References

- W1. <http://www.itl.nist.gov/div898/handbook/eda/section3/eda36.htm>.
- W2. [http://www.geocities.com/~mikemclaughlin/math\\_stat/Dists/Compendium.html](http://www.geocities.com/~mikemclaughlin/math_stat/Dists/Compendium.html).

- acceptance–rejecztion, 280
- algorithm, 281
- beta density, 282
- comparison function, 280
- adaptive coefficients, 307
- aperiodic random processes, 563
  - discrete Karhunen–Loeve (K–L) expansion, 571
  - Karhunen–Loeve (K–L) expansion, 565
    - eigenfunctions, 565
    - eigenvalues, 565
    - nonstationary process, 566
- AR(1) process, 522
- arcsine law, 192, 194
- attenuation coefficient, 668
- autoregressive process, 439
- bandpass signals, 606
  - analytic signal representation, 607
  - natural envelope, 606
    - lower side band, 607
    - upper side band, 606
  - complex envelope, 610
  - quadrature representation, 607
- Bayesian estimation, 384
  - discrete probability, 398
  - MAP Criterion, 395
    - ML criterion, 396
  - Bayes' theorem, 20, 669
  - Bayes' theorem for densities, 132
- Bernoulli
  - distribution, 37
  - trials, 37
    - independent identically distributed, 37
- Bernoulli process, 505
  - modified, 510
- Bessel function of zero order
  - modified, 107, 217, 621
- bilinear form, 302
- binomial coefficient, 29
  - Pascal's identity, 29
    - triangle, 30
  - properties, 29
  - sum, 39
  - Vandermonde's identity, 30
- binomial process, 506, 509
- birthdays problem, 27, 53
- bivariate Gaussian distribution, 144, 154
- block matrices, 308
  - properties, 309
- Bose Einstein statistics, 32
- bound, 242
  - Chebyshev, 242
    - alternate forms, 244
    - geometric derivation, 243
  - Chernoff, 248
- Brownian motion (see Wiener process), 492, 523
- Catalan numbers, 31
- Cauchy
  - density, 232
  - distribution, 115, 221, 226
- Cauchy–Riemann conditions, 303
- Cayley Hamilton theorem, 294
- central limit theorem, 260
- characteristic functions, 155, 212, 213
  - examples, 157
    - continuous random variables, 159
    - discrete random variables, 157
  - existence, 155
  - joint characteristic functions, 157
  - moment generating properties, 156
- chi-squared test, 267
- combinations, 28
  - sampling with replacement, 32
  - sampling without replacement, 29
- comparison of Gaussian and Poisson
  - approximations, 93
  - conditional densities, 122

- conditional distributions, 122
  - properties, 122
- conditional distributions and densities, 122,
  - 313
  - clarification, 315
  - $P(A) = 0$ , 126
  - $P(A) \neq 0$ , 122
- confidence coefficient, 351, 364
- confidence interval, 351, 364
  - coefficients of simple linear regression, 371
  - unknown mean of population, 364
    - known variance, 364
    - unknown variance, 366
  - unknown variance with unknown mean, 369
- confidence region
  - simple linear regression line, 372
- congruential generator, 264
  - linear, 264
  - mixed, 264
  - multiplicative, 264
- continuous distributions, 79, 95
  - beta distribution, 114
  - Cauchy distribution, 115
  - chi distribution, 104
    - Maxwell density, 106
    - Nakagami density, 108
    - Rayleigh density, 105
    - Rice's density, 107
  - chi-square, 102
  - double Laplace distribution, 118
  - Erlang distribution, 97
  - exponential distribution, 80
    - hazard rate, 82
      - instantaneous, 82
    - Poisson arrival process, 81
  - gamma, 99
    - reproductive property, 100
  - Gaussian distribution, 84
  - Gibbs distribution, 118, 674
    - potential function, 118
  - Laplace distribution, 96
    - double Laplace distribution, 118
  - lognormal distribution, 112
  - Pareto distribution, 117
  - Snedecor  $F$ -distribution, 111
  - Student- $t$  density, 110
    - summary, 119
  - triangular, 95
  - uniform distribution, 79
  - Weibull, 101
- continuous Kalman filter, 660
  - assumptions, 660
- matrix Riccati equation, 661
- continuous random variables, 151
  - means and variances, 150
    - exponential distribution, 151
    - Gaussian distribution, 152
  - convergence, 256
    - almost sure, 257
    - Cauchy criterion, 257
    - in distribution, 258
    - in probability, 257
    - mean square, 258
    - pointwise, 256
    - properties, 258
    - quadratic mean, 258
  - convolution techniques, 279
    - Erlang distribution, 279
    - triangular distribution, 279
  - counting, 25
    - addition rule, 25
    - multiplication rule, 25
  - convex function, 254
  - covariance matrices, 320
  - diagonalization principles, 330
    - diagonalization to an identity matrix, 331
    - principles, 330
  - simultaneous diagonalization, 334
    - summary of procedures, 335
  - covariance of ACF estimators, 452
  - craps game, 16
  - cumulants, 167
    - Gaussian distribution, 169
    - Poisson distribution, 169
  - cumulant generating functions, 167
  - cyclostationary process, 560
  - density function, 65
  - differentiation, 301
    - complex variable, 303
      - Cauchy–Riemann conditions, 303
      - partial derivative, 304
      - regular point, 304
      - singular point, 304
      - total derivative, 303
    - complex vectors, 305
      - partial differentiation, 305
    - real matrices, 308
    - real vectors, 301
  - direct determination of density, 194
    - conditions, 196
    - steps, 196
  - direct determination of joint density, 227
  - discrete distributions
    - Bernoulli, 37
    - Benford, 55

- discrete distributions (*Continued*)
  - binomial, 38
    - approximation, 50
  - geometric, 42
  - hypergeometric, 46
  - logarithmic, 55
    - finite law, 55
      - first significant digit, 55
      - second significant digit, 58
    - infinite law, 60
  - multinomial, 41
  - negative binomial, 44
    - Pascal distribution, 44
  - Poisson, 48
    - approximation to binomial, 50
    - summary, 62
- discrete Kalman filter, 647
  - across observations, 649, 651
  - algorithm, 655
  - assumptions, 650
  - between observations, 649, 652
  - comments, 654
  - covariance propagation, 655
  - Kalman gain, 650
  - scalar case, 657
  - timing diagram, 649, 653
- discrete random variables, 150
  - means and variances, 150
    - binomial distribution, 150
    - Poisson distribution, 151
- discrete time processes, 458
  - Anderson's theorem, 470
  - estimation of autocorrelation functions, 467
  - estimation of covariance of NACF, 469
  - estimation of NACF, 468
    - large-lag approximations, 469
    - variance of autocovariance estimators, 464
    - white noise, 470
- discrete time stationary processes, 437
  - moments, 437
- distribution and density functions, 66
  - properties, 73
    - cumulative distribution function  $F_X(x)$ , 73
    - probability density function  $f_X(x)$ , 74
    - probability mass function, 74
- distribution function, 66
  - from density, 75
- distribution of  $Y = g(X)$ , 174
  - steps for determination, 175
- eigenvalues of matrices, 294
  - characteristic polynomial, 293
- eigenvectors of matrices, 296
  - normalization, 297
- emission tomography, 667
- ergodic processes, 439
  - correlation-ergodic, 441
  - mean-ergodic, 439
  - power ergodic, 442
  - time average, 439
- estimated regression coefficients, 358
  - covariance, 359
  - from data, 358
- estimated regression line, 360
  - variance, 360
- estimation
  - likelihood function, 389
  - likelihood ratio, 389
  - linear estimation of vector variables, 337
    - overdetermined system, 338
    - underdetermined system, 338
  - maximum a posteriori probability (MAP), 388
    - minimization of average probability of error, 389
  - maximum likelihood, ML, 391
    - properties, 391
  - parameters, 350
    - Bayesian estimation, 384
    - covariance, 355
    - higher order moments, 358
    - interval estimation, 351
    - mean, 351
      - unbiased, 351
      - minimum mean square, 353
    - point estimation, 351
    - variance, 354
      - standard error, 355
    - statistic, 351
  - estimation criteria, 340
    - consistent, 341
    - efficient, 341
      - Rao–Cramer bound, 341
  - log-likelihood, 341
  - maximum likelihood, 341
  - mean square error, 340
  - minimum variance, 340
  - unbiased, 340
- estimation of a random variable
  - by a constant, 342
  - by a function of another random variable, 342
- expectation maximization algorithm, 672
  - one step late, 672
  - RMS error, 680

- field, 8, 11
  - Boolean field, 9
  - Borel  $\sigma$ -field, 9
- functions of a single random variable, 173
  - higher order moments, 203
  - moments, 202
  - random variable  $g(X)$ , 173
    - conditions, 173
- function of two random variables, 206
  - image, 206
  - inverse image, 206
  - conditions, 206
  - steps for determining  $F_Z(z)$ , 207
- gambler's ruin, 518
- gamma function, 99
  - incomplete, 100
- Gaussian approximation to binomial distribution, 91
  - Demoivre–Laplace limit theorem, 91
- Gaussian distribution, 84
  - density function, 84
  - distribution function, 84
  - mean, 84
  - standard, 85
    - properties, 87
  - tails function  $Q(x)$ , 88
    - lower bound, 89
    - properties, 90
    - upper bound, 89
  - variance, 84
- Gaussian process, 521
  - properties, 521
- generating functions, 161
  - examples, 162
    - binomial distribution, 163
    - geometric distribution, 164
    - Poisson distribution, 163
  - moment generating properties, 162
- Gibbs' distribution, 118, 674
  - potential function, 674
    - Incosh, 674
    - sigmoid, 675
- Gibbs' prior, 667
- higher order moments, 153
  - cross moments, 154
- Hilbert transform, 607
  - table, 610
- histograms, 266
- hypothesis testing, 373
  - acceptance region, 374
- alternate hypothesis, 373
- binary, 375
- composite hypothesis, 374
- cost for errors, 401
- mean  $\mu$ —known variance  $\sigma^2$ , 378
- mean  $\mu$ —unknown variance  $\sigma^2$ , 379
- ML rule, 394
- number of samples  $n$  for given  $\alpha$  and  $\beta$ , 382
- null hypothesis, 373
- operating characteristic, 376
- power of the test, 376
- procedure for testing, 378
- rejection region, 374
- significance testing, 373, 374
- simple hypothesis, 374
- statistic, 373
- image, 64
- independence, 4, 18
  - functional, 4, 18
  - statistical, 4, 18
- independent increment process, 507
  - stationary independent increment, 507
- indicator function, 205, 243
  - properties, 205
- inequalities, 242
  - Cauchy–Schwartz, 251
  - frequency domain, 253
- Chebyshev, 242
  - geometric derivation, 243
  - generalized Chebyshev, 244
- Jensen, 254
- Markov, 246
- innovations process, 648
- interval estimation, 364
- inverse Hilbert transform, 607
- inverse image, 64
- inverse problem  $g(x)$  from  $f_X(x)$  and  $f_Y(y)$ , 200
- inverse transformation, 269
- discrete distributions, 270
- continuous distributions, 272
  - exponential distribution, 272
  - Rayleigh distribution, 274
  - Weibull distribution, 273
- tables, 278
- Jacobian determinant, 227, 230, 233
- Jacobian matrix, 227, 230, 233
- joint cumulants, 169
- joint distribution, 135
  - continuous distribution, 136
  - discrete distribution, 135

- joint distribution (*Continued*)
  - joint probability mass function, 135
  - marginal distribution, 135
  - properties, 137
  - region bounded by a function, 140
  - regions bounded by constants, 138
  - regions bounded by functions, 141
- joint distribution of two vectors, 312
- joint moments of two random vectors, 321
  - cross correlation matrix, 321
  - cross covariance matrix, 322
- joint Gaussian density, 213
- Karhunen–Loeve transform, 565
  - continuous, 565
  - discrete, 571
  - matrix formulation, 572
- Kolmogorov axioms, 11
  - Boolean additivity, 11
  - non negativity, 11
  - normalization, 11
  - sigma additivity, 11
- kurtosis, 168
- Lagrange multipliers, 351
- likelihood function, 341
- log likelihood function, 341
- limit theorems, 259
  - central limit theorem, 260
  - strong law of large numbers, 260
  - weak law of large numbers, 259
- linear filters, 592
  - carrier frequency, 592
  - cut-off frequency, 592
  - ideal band-pass, 592
  - ideal high-pass, 592
  - ideal low-pass, 592
  - matched filters, 596
- linear systems, 574
  - causal system, 574
  - convolution integral, 574
  - impulse response, 574
  - state transition matrix, 575
  - transfer function, 575
- linear regression, 343
  - evaluating coefficients, 343
  - multiple, 343
  - simple, 343, 360
    - estimated regression line, 361
    - variance of regression coefficients, 362
- linear transformations, 328
- marginal distributions, 142, 143
- Markov chains, 536
  - classification of states, 544
    - absorbing state, 544
    - aperiodic, 545
    - ergodic, 545
    - intercommunicating, 544
    - irreducible, 544
    - periodic, 545
    - recurrent, 545
    - reducible, 544
    - transient, 545
  - discrete, 536
  - first passage probability, 544
  - first return probability, 545
  - homogeneous, 536
  - mean first passage time, 545
  - mean recurrence time, 545
    - null recurrent, 545
    - positive recurrent, 545
  - $m$ -step transition probability, 536
    - steady state, 539
  - properties, 550
  - state occupancy probabilities, 538
  - stochastic matrix, 537
- Markov process, 527
  - AR process, 534
  - Chapman–Kolmogorov equation, 528
  - homogeneous, 529
  - Markov property, 527
  - Poisson process, 532
  - properties, 530
  - transition probability density, 528
  - Wiener process, 533
- martingale process, 551
  - continuous martingale, 552
  - discrete martingale, 552
  - likelihood ratio martingale, 555
  - Poisson martingale, 554
  - random walk, 553
  - Wiener martingale, 554
- Mass megabucks, 46
- matched filters, 596
  - input signal to noise ratio, 597
  - non-white Gaussian noise, 604
  - output signal to noise ratio, 597
  - properties, 599
  - rectangular pulse, 600
  - sinusoidal pulse, 600
- matrices, 284
  - addition, 285
  - adjoint, 291
  - characteristic equation, 293
  - characteristic polynomial, 293
  - definiteness, 295
  - negative definite, 295

- positive definite, 295
- definition, 284
- determinant, 286
  - properties, 287
- diagonalization, 298
- inverse, 290
  - properties, 292
- multiplication, 285
- orthogonal, 293
- principal minors, 295
- rank, 288
- similarity transformation, 300
- singular, 287
- square, 284
- symmetric, 299
  - properties, 299
- trace, 288
- transpose, 286
- means and variances table, 170
- means of estimated autocovariances, 449
- minimum mean square estimator
  - mean, 353
- mixed distributions, 118
- modal matrix, 297
- modified Bessel function  $I_0(z)$ , 107, 217, 621
- modulation, 434
- moment generating functions, 164
  - gamma distribution, 167
  - Gaussian distribution, 166
  - negative binomial, 165
- Monty Hall problem, 22
- multiple functions of random variables, 238
- narrowband bandpass process, 612
  - analytic signal representation, 615
  - complex envelope, 615
  - natural envelope representation, 618
  - quadrature representation, 614
    - properties, 615
  - with additive sinusoid, 619
- normal equations, 344
- operating characteristic, 376
- orthogonal increment processes, 508
- orthogonality principle, 346
  - random processes, 626
  - random variables, 625
- Paley–Wiener criterion, 593, 631
- partial differentiation, 304
  - complex vectors, 305
- real matrices, 308
- real vectors, 301
- periodic random processes, 557
  - stationary, 558
- permutations, 26
  - sampling with replacement, 26
  - sampling without replacement, 26
- Poisson process, 491, 492, 508
  - formal derivative, 503
  - generalized, 501
    - autocorrelation, 502
    - autocovariance, 503
    - mean, 502
  - interarrival time, 496
    - distribution, 496
  - waiting time, 496
    - distribution, 497
- Poisson white noise, 504
- Polya urn model, 534
- power spectral density, 472
- predictable process, 557
- probability, 10
  - conditional, 14
  - reduced sample space, 14, 17
- density, 13
- distribution function, 13
- measure, 11
- space, 10
- total, 20
- probability mass function, 38, 74
- pseudoinverse, 338
- quadrature components
  - properties, 615
- quadratic error surface, 307
- quadratic form, 302
- random binary wave, 429
- random number generator, 264
  - properties, 265
- random processes, 406
  - bandpass processes, 606
  - classification, 490
    - continuous state continuous time, 491
    - continuous state discrete time, 491
    - discrete state continuous time, 491
    - discrete state discrete-time process, 490
  - definition, 406
  - distribution and density, 408
  - estimation of parameters, 445
    - continuous time processes, 445
      - autocovariance, 447
      - covariance, 452
      - variance, 454

- random processes (*Continued*)
  - mean, 446
  - variance, 446
  - discrete time processes, 458
    - autocovariance, 459
    - estimation, 461
    - Lagrange multiplier, 459
    - mean, 458
  - higher order distributions, 414
  - independent, 416
  - means and variances, 408
  - normalized autocovariance, 416
  - normalized cross covariance, 416
  - orthogonal, 417
  - second order moments, 415
    - autocorrelation, 415
    - autocovariance, 415
    - cross correlation, 416, 435
    - cross covariance, 416
    - normalized autocovariance, 416
    - normalized cross-covariance, 416
  - second order process, 414
  - through linear systems, 578
    - cross spectral density, 580
    - output power spectral density, 580
    - threshold filter, 588
  - uncorrelated, 417- random telegraph wave, 432
- random variable, 64
  - conditional moments, 146
    - conditional expectation, 148
    - conditional variance, 149
  - continuous, 79
  - correlation coefficient, 150
  - covariance, 150
  - definition, 64
  - degenerate, 241
  - discrete, 37
  - independent, 147
  - distribution and density, 66
    - properties, 73
  - moments, 146
    - expectation, 146
      - properties, 147
    - joint, 148
    - variance, 149
      - properties, 149
  - orthogonal, 147
  - uncorrelated, 147
- random vectors, 311
  - Bayes theorem, 316
  - conditional density, 328
  - inner product, 320
  - independent, 323
  - joint density, 313, 326
- moments, 319
  - correlation matrix, 319
  - properties, 324
- covariance matrix, 320
  - properties, 324
- expectation vector, 319
- orthogonal, 323
- outer product, 320
- uncorrelated, 323
- random walk process, 490, 512
  - elementary, 509
  - symmetric, 510
  - two absorbing barriers, 514
  - unrestricted, 512
- Rayleigh density, 215, 621
- Rayleigh distribution, 215, 232, 619
- recursive estimation, 648
- regression coefficients, 343
- Rice's density, 217, 621
- Rice's distribution, 217, 621
- sample space, 10
  - continuous, 10
  - discrete, 10
- sample variance, 354
- set, 1
  - cardinality, 2
  - unions and intersections, 5
  - Cartesian product, 7
  - complement, 6
  - De Morgan's laws, 8
  - difference, 6
  - element, 1
  - empty set, 2
  - equal, 3
  - event, 9
  - finite, 1
  - inclusion–exclusion principle, 5
  - intersection, 4
  - mutually exclusive, 4
  - null, 11
  - partition, 7
  - power set, 3
  - relative complement, 6
  - subset, 2
  - symmetric difference, 7
  - table of set properties, 8
  - union, 3
  - universal, 2
  - Venn diagram, 2
- shaping filter, 580
- signal to noise ratio, 253, 597
- significance level, 351, 364, 374
- similarity transformations, 300
- skewness, 168

- solution of linear equations, 291
- special methods, 274
  - Box–Mueller transformation, 275
    - Gaussian distribution, 275
      - Rayleigh distribution with phase, 277
  - spectral density–continuous time, 472
    - bandlimited process, 475, 476
    - bandpass process, 477
    - cross spectral density, 473
    - power spectral density, 473
      - alternate form, 480
      - estimation, 482
      - periodogram, 480
      - properties, 479
      - white noise, 478
    - Wiener–Khinchine theorem, 472
  - spectral density–discrete time, 484
    - power spectral density, 484
      - alternate form, 486
      - discrete white noise, 488
  - spectral factorization, 633
    - additive factoring, 634
    - multiplicative factoring, 633
  - standard error, 355
  - stationary random processes, 420
    - autocorrelation, 421
    - autocovariance, 421
    - cross correlation, 422
    - cross covariance, 422
      - first order, 420
      - independent, 421
      - mean, 421
      - normalized autocovariance, 421
      - normalized cross covariance, 422
      - n*th order stationary, 420
      - orthogonal, 421
      - periodic, 423
      - properties of correlation
        - functions, 422
      - second order, 420
      - strict sense, 420
      - uncorrelated, 421
      - variance, 421
      - wide sense, 420
    - stochastic estimation algorithm, 671
      - expectation step, 671
      - maximization step, 672
    - strong law of large numbers, 260
    - submartingale, 552
      - decomposition, 557
    - supermartingale, 552
  - tables
    - cumulative binomial, 708
    - cumulative Gaussian, 687
  - cumulative Poisson, 704
  - Fourier transform, 683
  - inverse chi-square, 694
  - inverse Gaussian, 692
  - inverse Student-*t*, 701
  - tails probability, 242
  - total probability for densities, 131
  - transmission tomography
    - computer simulation, 676
    - pathlength matrix, 676
      - backward, 677
      - forward, 677
    - phantom description, 676
    - probability model, 668
    - reconstructed images, 679
      - cross sectional histograms, 680
      - RMS error, 680
  - two functions of two random variables, 222
  - conditions, 223
  - auxiliary random variable, 233
  - derivation of Student-*t*, 237
  - type I error, 374
    - false alarm, 374
    - producer's risk, 374
  - type II error, 374
    - consumer's risk, 374
    - missed alarm, 374
  - uncorrelated increment process, 508
  - uniform distribution, 264
  - variance of the estimated regression line, 360
  - vector Gaussian random variables, 323
  - weak law of large numbers, 259
  - weighted Poisson process, 501
  - white Gaussian noise, 521
  - whitening filter, 632
  - white noise process, 422
  - Wiener filtering, 627
    - causal, 631
    - filtering, 627, 638
    - noncausal, 627
    - prediction, 627, 641
    - smoothing, 627, 642
  - Wiener–Hopf integral equation, 632
    - filtering solution, 634
    - prediction-smoothing solution, 637
  - Wiener process, 523
    - conditions, 523
    - derivation, 524
    - formal derivative, 526
    - properties, 526